Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2566

A DNA sequence (GASx671) was identified in *S.pyogenes* <SEQ ID 7631> which encodes the amino acid sequence <SEQ ID 7632>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4294(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2567

A DNA sequence (GASx672R) was identified in *S.pyogenes* <SEQ ID 7633> which encodes the amino acid sequence <SEQ ID 7634>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have a cleavable N-term signal seq.
  INTEGRAL    Likelihood = -6.37   Transmembrane  106 - 122 ( 104 - 125)

----- Final Results -----
              bacterial membrane --- Certainty=0.3548(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2568

A DNA sequence (GASx673) was identified in *S.pyogenes* <SEQ ID 7635> which encodes the amino acid sequence <SEQ ID 7636>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4781(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB18697 GB:U38906 ORF22 [Bacteriophage r1t]
 Identities = 78/207 (37%), Positives = 123/207 (58%), Gaps = 2/207 (0%)

Query: 28  EIHRILGIDEVYKAPKRLTDILFDKDSREDIFRQFLKYETDVSYDWFMQYFEEEQADRKN 87
           + + +L +DE      R+ +++FDK  RE+ + + L    D+  D+F  YF    A
Sbjct: 7   QFYDMLNVDEHMNFTNRIQELVFDKKGREEFYSKILNIHHDMGVDFFRDYFMAHSAVSA- 65

Query: 88  KKQDFTPKSVSTLLSKIISGNQYYEVA-VGTGGILIQAWQEQRLNDSPFTYRPSKYWYHV 146
           K Q +TP  +  L + ++ G+   ++   GTG ++IQ WQ+ R+N   F Y PS YWY
Sbjct: 66  KGQHYTPDELGKLTALLVGGSGGADLTGAGTGTLIIQKWQDDRMNTDFFNYLPSNYWYQA 125

Query: 147 EELSDKAVPFLLFNMSIRGINGVVVHGDSLTRQVKNIYFLQNTKDDMLSFSDINVMPRTQ 206
            ELSD+A+ FL+   +IRG+NGVV+HGD+L   VK +YF+QN+ ++ + FS+INV+P ++
Sbjct: 126 LELSDEAISFLIHAFAIRGMNGVVIHGDALEMAVKQVYFIQNSANNPIGFSEINVIPHSK 185

Query: 207 DIEREFNVKEWIGDGIEHIENPLIEWI 233
           D      + EW    IEHIE+   +WI
Sbjct: 186 DAMEFLGIHEWTEQAIEHIESKFPDWI 212
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2569

A DNA sequence (GASx674) was identified in *S.pyogenes* <SEQ ID 7637> which encodes the amino acid sequence <SEQ ID 7638>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.00   Transmembrane  122 - 138 ( 122 - 138)

----- Final Results -----
               bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF63071 GB:AF158600 gp137 [Streptococcus thermophilus
          bacteriophage Sfi11]
 Identities = 66/135 (48%), Positives = 89/135 (65%), Gaps = 2/135 (1%)

Query: 5   PEIDIQKTKSNAKRKLREYPRWRRIANDVDTQKVTATYSFEPRQSHGVPSKPVERLALNR 64
           PEID + T    KRKLREYPRWR IA+D   QK+T  ++F PR   G  +KPVE +A+ R
Sbjct: 4   PEIDEKATLKRCKRKLREYPRWREIAHDSAEQKITQEFTFMPRG--GGVNKPVENIAVRR 61

Query: 65  VSAEQELDAIEQAVSMILEPERRRILYDKYLAPYKKADKVIYTELCMSESFYYDTLDIAL 124
           V A  EL+AIEQAV+ +  P+ RRIL +KYLA   K +  I   +  + + L+ ++
Sbjct: 62  VDALNELEAIEQAVNGLYRPDYRRILIEKYLAYPPKPNWQIAQSIGFERTAFQELLNNSI 121

Query: 125 LAFAELYREGVLLVE 139
           LAFAELYR+G L+VE
Sbjct: 122 LAFAELYRDGRLIVE 136
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2570

A DNA sequence (GASx675) was identified in *S.pyogenes* <SEQ ID 7639> which encodes the amino acid sequence <SEQ ID 7640>. Analysis of this protein sequence reveals the following:

```
Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1865(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2571

A DNA sequence (GASx676) was identified in *S.pyogenes* <SEQ ID 7641> which encodes the amino acid sequence <SEQ ID 7642>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4870(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB07254 GB:AP001519 unknown [Bacillus halodurans]
 Identities = 194/451 (43%), Positives = 262/451 (58%), Gaps = 69/451 (15%)

Query: 1   MEFVDKKLSEITPYKNNPRNNDEAVGPVAE----SIKEFGFKVPIVV-DKNGEIVNGHTR 55
           +  V+KK+ ++ P + NPR + +    P  E     SI+EFG    PIV  ++ G +V GH R
Sbjct: 3   IRIVNKKIDDLVPAEYNPRLDLQPGDPEYEKLKRSIEEFGLVEPIVFNERTGRVVGGHQR 62

Query: 56  YKAAQKLGLETVPVIVADDLSEEQIKAFRLADNKV-GEIAVWDLDLLNEELNDILDLDMS 114
            K  ++LG E VPV V D L +    KA  +A NK+ G+    + L  L EEL+  L +D++
Sbjct: 63  LKILRELGWEEVPVSVVD-LDDHHEKALNVALNKIEGDWDNFKLKELLEELDSGL-IDVT 120

Query: 115 AFGFDVLDNLDDL-----IEDEKDL--DDF----TGTVPDEPKSKLGDIYQLGSHKLMCG 163
              GFD  + ++DL     +EDE ++  DDF          +EP +K GD++ LG H L+ G
Sbjct: 121 LTGFDE-EEIEDLMTQFFVEDENEIKEDDFDPDEVAEEIEEPITKPGDLWHLGRHFLLVG 179

Query: 164 DSTNGADVKKLMNGELADLLLTDPPYNVAYEGKTKDSLTIKNDSMDNDSFRQFLVNAFSS 223
           DST   DVK+LM  E AD++ TDPPYNV YEG T   + IKND+M++  F QFL +AF +
Sbjct: 180 DSTKIEDVKRLMGNEKADMIFTDPPYNVDYEGAT--GMKIKNDNMEDSEFYQFLFDAFVA 237

Query: 224 ANEVMKPGAVFYIWHADSEGYNFRGACFDIGWTVRQCLIWNKNSMVLGRQDYHWKHEPCL 283
             +V K G   Y+ HADSEG  FR A  D G+ ++QCLIW KNS+VLGRQDYHW+HEP L
Sbjct: 238 MYQVTKEGGPIYVCHADSEGLTFRKAFQDSGFLLKQCLIWVKNSLVLGRQDYHWRHEPIL 297

Query: 284 YGWKDGAGHLWASDRKQTSVID---------------------------------------- 305
           YGWK GA H W   RKQ++VI+
Sbjct: 298 YGWKPGAAHKWYGGRKQSTVIEDPVDLAITPKVDHVLLTFNNGISSTVVKVPSYEIIHDG 357
```

```
Query: 306 ---------YEKPQRNGVHPTMKPVGLFDYQIKNNTKGSDIVLDLFGGSGTTLIACESNG 356
                    E+P+RN  HPTMKP+ L    I+N++K  + VLD FGGSG+TLIACE  G
Sbjct: 358 SDEGMTTWRIERPKRNADHPTMKPIALCARAIQNSSKPGERVLDPFGGSGSTLIACEQTG 417

Query: 357 RHARLMEYDPKYVDVIIKRWEELTGESVIQL 387
           R   +MEYDP Y +VII+RWEE TG++ ++L
Sbjct: 418 RICHMMEYDPVYAEVIIRRWEEWTGQNAVKL 448
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2572

A DNA sequence (GASx677) was identified in *S.pyogenes* <SEQ ID 7643> which encodes the amino acid sequence <SEQ ID 7644>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4744(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2573

A DNA sequence (GASx678) was identified in *S.pyogenes* <SEQ ID 7645> which encodes the amino acid sequence <SEQ ID 7646>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.27   Transmembrane   90 - 106 (  90 - 106)

----- Final Results -----
              bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2574

A DNA sequence (GASx679) was identified in *S.pyogenes* <SEQ ID 7647> which encodes the amino acid sequence <SEQ ID 7648>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3408(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA66734 GB:X98106 minor capsid protein [Bacteriophage phig1e]
 Identities = 213/494 (43%), Positives = 323/494 (65%), Gaps = 19/494 (3%)

Query: 1   MGVIQKIKNLVTRSKYVM-TTQSLTNITDHPKIAISKLEYDRITTNLKYYKSDWDSVLYL 59
           MG+IQ+IK+L  +       T SL+ ITD P+I+I   EY RI T+L YY      + Y
Sbjct: 1   MGLIQRIKDLFWKGAAATGVTGSLSKITDDPRISIDPDEYVRIQTDLDYYSDKLQYIHYQ 60

Query: 60  NTDGETKKRDLNHLPIARTAAKKIASLVFNEQAEIKV-DDDAANEFISETLKNDRFNKNF 118
            +DG  KKR  N + +A+TAA++IAS+VFNE+AEI V D++ A++F+++ L+++ F   F
Sbjct: 61  ASDGIKKKRLKNTINMAKTAARRIASVVFNEKAEIHVKDNNEADKFLNDVLEDNDFKNKF 120

Query: 119 ERYLESCLALGGLAMRPYVDGDKVRVAFVQAPVFLPLQSNTQDVSSAAVVIKSVKTINGK 178
           E  LE  +ALGG AMRPY+DG+ +++A+V+A  F PLQSNT D+S AA+  ++ +T + +
Sbjct: 121 EEALEKGVALGGFAMRPYIDGNHIKIAWVRADQFYPLQSNTNDISEAAIASRTQRTESNQ 180

Query: 179 EVYYTLIEFHEWQSSDDYVISNELYRSDDKAKVGSRVPLS--EVYKDLKDEAKVTDVTRP 236
             YYTL+EFH+WQ +  Y I+NELY+SD    VG++VPLS   VYK+L  +  ++ + RP
Sbjct: 181 TKYYTLLEFHQWQDNGSYQITNELYKSDSPDIVGNQVPLSTLPVYKELAPQVTISGLQRP 240

Query: 237 IFTYLKTPGMNNKDINSPLGLSIFDNAKTTIDFINTTYDEFMWEVKMGQRRVAVPESLTA 296
           +F Y KTPG NN +I SPLGL + DNAK  +D IN T+D+F+WE+++GQ+ +AV   +
Sbjct: 241 LFAYFKTPGANNINIESPLGLGVVDNAKHVLDDINDTHDQFIWEIRLGQKHIAVQPGMLR 300

Query: 297 LTVRTADGDVVPRPRFESDQNVYIRMGGRDLDSSAIQDLTTPIRADDYIKAINEGLSLFE 356
                D   +P F+++QNVY+ +   D +   ++D+TTPIR   Y  AI+  +  FE
Sbjct: 301 F-------DDEHKPTFDTEQNVYVGVLSDDNNGLGVKDMTTPIRTVQYKDAIDHFIKEFE 353

Query: 357 MQIGVSAGLFSFDGKSMKTATEIVSENSDTYQMRNSIVTLVEQSLKELVISIFEIAKAYD 416
           +QIG+S G FS+     +KTATE+VS NS TYQ R+S +T+VE+++ EL  SIFE+A A
Sbjct: 354 VQIGLSTGTFSYSNDGVKTATEVVSNNSMTYQTRSSYLTMVEKAIDELCQSIFELANAGA 413

Query: 417 LYQSEVP--SMDNISISL------DDGVFTDRDAELDYWIKVVNAGFGTREMAIQKVLNV 468
           L+    P  ++D+ S  L      DDGVF ++D +L+   KV+  G  +++  +Q+  +
Sbjct: 414 LFDDGKPLFTLDSASQPLDIECHFDDGVFVNKDKQLEEDAKVLAIGALSKQTFLQRNYGM 473

Query: 469 TEEKAQEIAAEINT 482
           T+E+A E  A+I +
Sbjct: 474 TDEQAAEELAKIQS 487
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2575

A DNA sequence (GASx680) was identified in *S.pyogenes* <SEQ ID 7649> which encodes the amino acid sequence <SEQ ID 7650>. Analysis of this protein sequence reveals the following:

```
Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1840(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB53790 GB:AJ242593 gp4 [Bacteriophage A118]
 Identities = 114/385 (29%), Positives = 187/385 (47%), Gaps = 23/385 (5%)

Query: 8   LNDEQLLLEASQLSDMYHQLTLDLFDQVIERIKARGSASLADNPYLWQANKLHDVGLLNA 67
            L   QL L    + D+Y  L  +LF  ++ R+K + + S ADN   WQ  KL+ V  L+
Sbjct: 3   LTPRQLDLFVQPIVDVYTGLENELFTLIVRRLKTKKNIS-ADNVLAWQIEKLNQVHALDQ 61

Query: 68  DNIKLIAKYSGIAEAQLRYIIKNEGFKIYKNTSEQLEEALGRESGV-------NSTIQDD 120
              I+ I+K SG++  +L  ++K+ G+   K        +    E+G           TI D
Sbjct: 62  QMIERISKASGVSAKKLFSVVKDAGYSDLKQVDNYFSKLA--EAGAVLPLVSDGQTIVDK 119

Query: 121 LSNYARQAIDDVHNLTNTTLPFSVIGAYQGIIQDAVAGVVTGLKTPDQAINQTVIKWFKK 180
            +     +  +    N T+        Y  II +    V+ GLKT  QA+ +TV K+ +
Sbjct: 120 VMRSYFKLAESNYKRINQTMLSQARQIYSDIIHETTQSVLAGLKTHRQALAETVTKFAEN 179

Query: 181 GFYGFTDKAGRKWRADSYARTVINTTTWRVFNEAKEAPAREFGIDTFYYSKKATAREMCA 240
            G     DKA ++W  ++Y RTV  TT   V+N  ++    E+G+D    S+  AR  C+
Sbjct: 180 GVPALVDKANKRWTPEAYVRTVTRTTVNSVYNSVEDERMNEYGVDLVRISQHVGARPTCS 239

Query: 241 PLQHQIV---TTGEAREEGGIKILALSD----YGHGEPDGCLGINCKHTKTPFVVGVNSK 293
             +Q +++   +  E R + G K +++      YG+G  DG  G NC+H +  F+ G+N
Sbjct: 240 IVQGKVICLLSVEETRSKYGNKYMSIYSPELRYGYG--DGIFGCNCRHHRFAFIEGINIA 297

Query: 294 PELPEHLKNITPAQAKANANAQAKQRAIERSIRKSKELLHVAKQLGDKELIRQYQSDVRS 353
            P+  E    I   + K       +QR +ER IR +K  L  A++LGD+ +  +++ + VR+
Sbjct: 298 PDESE---LIDEEENKRVYALSQQQRLMERDIRAAKRKLSAAEELGDELAVKKAKQAVRT 354

Query: 354 KQDALNYLINNNAFLHRNQAREKRY 378
            KQ  L   +  +  L R  +REK Y
Sbjct: 355 KQSKLRAFVKTHN-LTRQYSREKVY 378
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2576

A DNA sequence (GASx681) was identified in *S.pyogenes* <SEQ ID 7651> which encodes the amino acid sequence <SEQ ID 7652>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2756(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2577

A DNA sequence (GASx682) was identified in *S.pyogenes* <SEQ ID 7653> which encodes the amino acid sequence <SEQ ID 7654>:

```
TLDNQSVIKAIGDTVDYIKKNYKRKWGK
```

Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2578

A DNA sequence (GASx683) was identified in *S.pyogenes* <SEQ ID 7655> which encodes the amino acid sequence <SEQ ID 7656>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5288(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2579

A DNA sequence (GASx685) was identified in *S.pyogenes* <SEQ ID 7657> which encodes the amino acid sequence <SEQ ID 7658>:

```
GATEVGANRVVSGVYGEVLGVQIVRSRKCPKGTAYMVRKGALRIMLKRNTMVETDRDITKAINQIVANKHYGVYLYKAEKAVKITLKDAAK
K
```

Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1750(Affirmative) < succ>
```

```
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA59185 GB:X84706 major head protein [Bacteriophage B1]
 Identities = 138/270 (51%), Positives = 186/270 (68%), Gaps = 6/270 (2%)

Query: 1   MAVGTTKMAQMLDPEVLADMIDAEVGKAIRFAPLAEVDTTLEGQPGTTLTVPK-WDYIGD 59
           M+   T +A +++PEVLA ++  E+ KA+RFAPLA+VDTTL+GQPG TL  P  + YIGD
Sbjct: 1   MSKQKTTLADLVNPEVLATIVSYELNKALRFAPLAQVDTTLQGQPGNTLKFPDPFTYIGD 60

Query: 60  AEDVAEGEAIPMTQLGFKKTTMTIKKAGKGVEITDEAILSGYGDPVGQAAKQIVEAIDHK 119
           A DVAEG  I + ++G    ++TIKKA KG EITDEA LSGYGDP+G++ KQ+  ++ +K
Sbjct: 61  AADVAEGGEISLDKIGTTTKSVTIKKAAKGTEITDEAALSGYGDPIGESNKQLGLSLANK 120

Query: 120 VDADVLDALSKSTQTVEATATVDGVSKALDIFNDEDDAETVIVMNPADASTLRLDAAKEW 179
           VD D+L A   ++QTV   A VDGV  ALDIFNDED    V+++NP DA+ +R DA  +
Sbjct: 121 VDDDLLSAAKTTSQTVSTKANVDGVQAALDIFNDEDAQAYVLIVNPKDAAKIRKDANAKN 180

Query: 180 LGATEVGANRVVSGVYGEVLGVQIVRSRKCPKGTAYMVR----KGALRIMLKRNTMVETD 235
           +G +EVGAN +++G Y +VLG QIVRS+K  +G+A M +       AL+++LKR   VETD
Sbjct: 181 IG-SEVGANALINGTYADVLGAQIVRSKKLAEGSALMFKIVSNSPALKLVLKRGVQVETD 239

Query: 236 RDITKAINQIVANKHYGVYLYKAEKAVKIT 265
           RDI      I A++HY  YLY   K V IT
Sbjct: 240 RDIVTKITVITADEHYAAYLYDLTKVVNIT 269
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2580

A DNA sequence (GASx686) was identified in *S.pyogenes* <SEQ ID 7659> which encodes the amino acid sequence <SEQ ID 7660>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2581

A DNA sequence (GASx687) was identified in *S.pyogenes* <SEQ ID 7661> which encodes the amino acid sequence <SEQ ID 7662>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2942(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2582

A DNA sequence (GASx688) was identified in *S.pyogenes* <SEQ ID 7663> which encodes the amino acid sequence <SEQ ID 7664>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2844(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC00538 GB:L02496 unknown protein [Bacteriophage LL-H]
 Identities = 35/86 (40%), Positives = 48/86 (55%), Gaps = 6/86 (6%)

Query: 24  KLIMNNQVMMSMNPYVPYRDGALRGSSRANSVGVTWSGPHARAQFYGGAYNKYKSFKFKK 83
           +L + NQ+    M  YVP R G LR  S  N  G+ ++  +ARAQFYG         + +
Sbjct: 20  RLQVLNQMHQDMEQYVPKRAGFLRSQSFVNDTGIHYTAKYARAQFYGFV----NGHRVRN 75

Query: 84  YTTPGTGKRWDKRALANATIVKDWEK 109
           Y+TPGTG+RWD +  A A    DW+K
Sbjct: 76  YSTPGTGRRWDLK--AKAVYKADWQK 99
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2583

A DNA sequence (GASx689) was identified in *S.pyogenes* <SEQ ID 7665> which encodes the amino acid sequence <SEQ ID 7666>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2892(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA66741 GB:X98106 minor capsid protein [Bacteriophage phig1e]
```

```
 Identities = 36/109 (33%), Positives = 64/109 (58%), Gaps = 2/109 (1%)

Query: 17  DLGIKPRLDYLTRQEDLAIYPMPGGKVNNEYMDGTREISLPFEIAIKTKNQELASTVMWT 76
           +L +K  L YLT  + L++YP+PG +V +E   G ++  + +E+ ++TKNQ+ A+T +W
Sbjct: 16  NLPMKCTLGYLTAADSLSLYPLPGSRVLDEDYAGNQQWQMNYEVGMRTKNQQQANTTLWL 75

Query: 77  INSALSNFDL-KLPSLNHSYTFISLDVE-KPFLNDLSDQGFYIYVLDIT 123
           ++ AL       L S N S+ F SL +  +P +++   QG+  Y L  +
Sbjct: 76  VSQALDVLTADDLVSSNGSFEFESLTINGQPSISEQDTQGYSTYQLSFS 124
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2584

A DNA sequence (GASx690) was identified in *S.pyogenes* <SEQ ID 7667> which encodes the amino acid sequence <SEQ ID 7668>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1626(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB53798 GB:AJ242593 major tail shaft protein [Bacteriophage
          A118]
 Identities = 54/133 (40%), Positives = 77/133 (57%), Gaps = 9/133 (6%)

Query: 1   MRQKNALRGHFIAPYVKGEEKTEVTKEKLLELARWIKDISDDTDEKTEDEAYYDGDGTEE 60
           MR KNA   + +A  V G  + +  +     L++WI ++SDD  + TE++  YDGDG E+
Sbjct: 1   MRIKNAKTKYSVAEIVAGAGEPDWKR-----LSKWITNVSDDGSDNTEEQGDYDGDGNEK 55

Query: 61  TTVVGVKGAYTFEGTYDPEDKAQAHIASLKYKLGDERKVWHLIVSADGKTQWLGVATVTE 120
           T V+G   AYTFEGT+D ED+AQ  I + K +  + R +   I   D +T  +G ATV+E
Sbjct: 56  TVVLGYSEAYTFEGTHDREDEAQNLIVA-KRRTPENRSIMFKIEIPDTETA-IGKATVSE 113

Query: 121 I--IAGSGAAARF 131
           I   AG G A  F
Sbjct: 114 IKGSAGGGDATEF 126
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2585

A DNA sequence (GASx691) was identified in *S.pyogenes* <SEQ ID 7669> which encodes the amino acid sequence <SEQ ID 7670>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3521(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2586

A DNA sequence (GASx692) was identified in *S.pyogenes* <SEQ ID 7671> which encodes the amino acid sequence <SEQ ID 7672>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3438(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB53801 GB:AJ242593 gp15 [Bacteriophage A118]
 Identities = 67/191 (35%), Positives = 110/191 (57%), Gaps = 17/191 (8%)

Query: 11  FEFRGEIYPIDLSFNKVLDVFDVIDDDFLNEAEKCFLCLDILLDRTDLPFTYAVD----- 65
           +E+ G+ Y +DL+F+ VL V D+ +D+ L++  +  L +D+L    D+P+  + +
Sbjct: 12  YEYEGKEYKLDLAFDNVLRVIDLTEDNSLSDVFRANLAIDVLF-ADDMPWPRSNEEDEYA 70

Query: 66  -------LWVYIKTNFIDAERPEKPQLDIKGNPMPVVKEKEDNKKVI---DLSLDAEFIY 115
                  + + I TN+I  E  +    DI GN MP      D+ + I    L+ DA++IY
Sbjct: 71  NIEEKSLVLIDIFTNYIVKENDDGLLYDIDGNKMPSATNNNDDAEEIASYSLTQDADYIY 130

Query: 116 ASFRQAYQINLLKEQNRLSWIEFKALLNALPDDTVMQRIIAIRQWE-DDGEGSKKYRDNM 174
           ASF Q Y I+LL  + ++ W +F+ALL +L DDT ++ II IRQ E   G+G++K R+ +
Sbjct: 131 ASFLQDYNIDLLDSRGKMHWYKFRALLESLRDDTTIKTIIGIRQAELPSGKGTEKERNEL 190

Query: 175 RKLKAKYSLDE 185
            KLK +Y L +
Sbjct: 191 IKLKNRYKLKD 201
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2587

A DNA sequence (GASx694) was identified in *S.pyogenes* <SEQ ID 7673> which encodes the amino acid sequence <SEQ ID 7674>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4143(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG18639 GB:AY007505 unknown [Streptococcus mitis]
 Identities = 48/157 (30%), Positives = 85/157 (53%), Gaps = 10/157 (6%)

Query: 86  DLELSWEPDYIYKATHITPFSIKEVLRNFGRLKINFLIHPIKYLKTGKQEVPLVNG-GTL 144
           +LE S+ P+ ++ A H    S K    +  +LKI   + P +Y KT   E    NG GT+
Sbjct: 81  ELEFSYHPESVFYA-HFLTASYKPFGNHAWQLKIKLNMQPFRYQKTVNPES--YNGPGTI 137

Query: 145 QNPGNVQAKPILKIKGTGNGILTINDFETGLENVQSELVIDMERHLVYKDVLSAWDNIVR 204
            NPG + ++PI++++G G+  +TI   ET   NV+++  ID  +    +++ +A   +
Sbjct: 138 NNPGTIYSEPIIEVQGDGDVSITIGR-ETMYLNVKTKATIDCRQG--RQNIYNATGAVQN 194

Query: 205 TERHRMPLFDV--GQNKISWTGS-FTITAVPNWGVKV 238
           T R R   F++  G++ I++TG+   +  PNW  K+
Sbjct: 195 TLRKRGGFFEIPTGRSGITFTGNVLRLIIRPNWRYKI 231
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2588

A DNA sequence (GASx695R) was identified in *S.pyogenes* <SEQ ID 7675> which encodes the amino acid sequence <SEQ ID 7676>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60   Transmembrane   15 -  31 (  15 -  31)

----- Final Results -----
               bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2589

A DNA sequence (GASx697) was identified in *S.pyogenes* <SEQ ID 7677> which encodes the amino acid sequence <SEQ ID 7678>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3348(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA86895 GB:U28144 hyaluronidase [Streptococcus pyogenes]
 Identities = 326/337 (96%), Positives = 329/337 (96%)
```

```
Query: 1   MSENIPLRVQFKRMKAAEWARSDVILLESEIGFETDTGFARAGDGHNRFSDLGYISPLDY 60
           MSENIPLRVQFKRMKAAEWARSDVILLESEIGFETDTGFARAGDGHNRFSDLGYISPLDY
Sbjct: 1   MSENIPLRVQFKRMKAAEWARSDVILLESEIGFETDTGFARAGDGHNRFSDLGYISPLDY 60

Query: 61  NLLTNKPNIDGLATKVETAQKLQQKADKETVYTKAESKQELDKKLNLKGGVMTGQLKFKP 120
           NLLTNKPNIDGLATKVETAQKLQQKADKETVYTKAESKQELDKKLNLKGGVMTGQLKFKP
Sbjct: 61  NLLTNKPNIDGLATKVETAQKLQQKADKETVYTKAESKQELDKKLNLKGGVMTGQLKFKP 120

Query: 121 AATVAYSSSTGGAVNIDLSSTRGAGVVVYSDNDTSDGPLMSLRTGKETFNQSALFVDYKG 180
           AATVAYSSSTGGAVNIDLSSTRGAGVVVYSDNDTSDGPLMSLRTGKETFNQSALFVDYKG
Sbjct: 121 AATVAYSSSTGGAVNIDLSSTRGAGVVVYSDNDTSDGPLMSLRTGKETFNQSALFVDYKG 180

Query: 181 TTNAVNIAMRQPTTPNFSSALNITSGNENGSAMQLRGSEKALGTLKITHENPSIGADYDK 240
           TTNAVNIAMR  TTPNFSSALNITSGNENGSAMQLRGSEKALGTLKITHENPSIGADYDK
Sbjct: 181 TTNAVNIAMRHATTPNFSSALNITSGNENGSAMQLRGSEKALGTLKITHENPSIGADYDK 240

Query: 241 NAAALSIDIVKKTNGAGTAAQGIYINSTSGTTGKLLRIRNLSDDKFYVKSDGGFYAKETS 300
           NAA   + + K+ NGAGTAAQGIYINSTSGTTGKLLRIRNLSDDKFYVKSDGGFYAKETS
Sbjct: 241 NAARYPLILSKRQNGAGTAAQGIYINSTSGTTGKLLRIRNLSDDKFYVKSDGGFYAKETS 300

Query: 301 QIDGNLKLKDPTANDHAATKAYVDKAISELKKLILKK 337
           QIDGNLKLKDPTANDHAATKAYVDKAISELKKLILKK
Sbjct: 301 QIDGNLKLKDPTANDHAATKAYVDKAISELKKLILKK 337
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2590

A DNA sequence (GASx698) was identified in *S.pyogenes* <SEQ ID 7679> which encodes the amino acid sequence <SEQ ID 7680>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4208(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif 54-56
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA98102 GB:M19348 ORF [Streptococcus pyogenes phage H4489A]
 Identities = 250/648 (38%), Positives = 351/648 (53%), Gaps = 75/648 (11%)

Query: 1   MSRDPTLILDESNLVIGKDGRVHYTFTTEDDNPKVRLASKCLGTAHFNQLMIERGDQATS 60
           MSRDPT  ++E +L    DGR + TF  +  +  VRL S CLG     +L +E  +
Sbjct: 1   MSRDPTYTINEHDLSFA-DGRFYVTFKADKSSETVRLNSSCLGNTIIKKLQVEDDNTMHD 59

Query: 61  YVAPVVVEGTGNPTGLFKDLKEISLELTDTANSQLWSKIKLTNRGMLQEYYDGKIKTEIV 120
           +V P V   T    GL + +KE+ L+L D   S LW KIK  N+ ML EY + ++ + I
Sbjct: 60  FVKPKVT--TQQAFGLAQQVKELDLQLKDP-KSDLWGKIKFNNKAMLVEYANKEMSSAIA 116

Query: 121 NSARGVATRISEDTDKKLALINDTIDGIRREYRDADRKLSASYQAGIEGLKATMANDKIG 180
            SA  +  ++    D++ +    T++GI++  +
Sbjct: 117 QSAEQILLQVKSIDDERYSKFEQTLNGIKQTVKSES---------------------- 152

Query: 181 LQAEIKASAQGLSQKYDDELRKLSAKITTTSSGTTEAYESKLAGLRAEFTRSNQGTRTEL 240
               ++++   L+  +D  +  L  K +  S  T ++  S+L            G  + L
Sbjct: 153 ----VESARTQLASMFDSRISGLDGKYSRLSQ-TIDSLSSRLD--------DGVGNYSTL 199

Query: 241 ESQISGLRAVQQSTASQISQEIRDREGAVSRVQQSLESYQRRMQDAEENYSSLTHTVRGL 300
```

```
            ++SG           I   + +    VSR+ Q+ +  Q ++ +A +NYSSL+ TV+GL
Sbjct: 200 SQKVSG-----------IDLRVSNAANDVSRLSQTAQGLQSQITNANQNYSSLSQTVQGL 248

Query: 301 QSDVGSPTGKIQSRLTQLAGQIEQRVTRDGVMSIISGAGDSIKLAIQKAGGINAKMSGNE 360
           Q+ V        SR+ QL+  I  +VT+  V + I+ + D I  AI+     + KM+G+E
Sbjct: 249 QTTVRDNQSNATSRINQLSDLISTKVTKGDVETTIAQSYDKIAFAIRDKLPAS-KMTGSE 307

Query: 361 IISAINLNSYGVTIAGKHIALDGNTTVNGTFTTKIAEAIKIRADQIIAGTIDAARIRVIN 420
           IISAINL+  GV I GK+I LDGN+ ++      K A    + A +I  G ++A+RI
Sbjct: 308 IISAINLDRSGVKITGKNITLDGNSYISNA-VIKDAHIANMDAGKINTGYLNASRIAAEA 366

Query: 421 LNASSIVGLDANFIK--AKIGY---------------AIT---DLLEGKVIKARNGAMLI 460
           +    I    A F K  A  GY               A+T     + G V+ A NGA
Sbjct: 367 ITGDKIKMDYAFFNKLTANEGYFRTLFAKNIFTTSVQAVTTSASKITGGVLSATNGASRW 426

Query: 461 DLNTAKMDFNSDATINFNSKNNALVRKDGTHTAFVHFSNATPKGYTGSALYASIGITSSG 520
           DLN+A +DFN DATINFNSKNNALVRK GT+TAFVHFSNATPKGY GSALYASIGITSSG
Sbjct: 427 DLNSANIDFNRDATINFNSKNNALVRKSGTNTAFVHFSNATPKGYRGSALYASIGITSSG 486

Query: 521 DGVNSASSGRFAGLRSFRYATGYNHTAAVDQTEIYGDNVLVVDDFNITRGFKFRPDKMQK 580
           DG++SASSGRF G+R FRYA G  HTA VDQ EIYGD+++  DDFNI RGFK RP  M K
Sbjct: 487 DGIDSASSGRFCGVRFFRYAEGLQHTAKVDQAEIYGDDIVFSDDFNIDRGFKMRPSLMPK 546

Query: 581 MLDMNDLYAAVVALGRCWGHLANVGWNTAHSNFTSAVNRELNNYITKI 628
           M+D+N +Y A++ALGRCW H  N  W+  + +  SA+  E N +I  +
Sbjct: 547 MVDLNKMYQAILALGRCWLHANNTAWSW-NFDTRSAIIAEYNAHINNL 593
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2591

A DNA sequence (GASx699) was identified in *S.pyogenes* <SEQ ID 7681> which encodes the amino acid sequence <SEQ ID 7682>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3323(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2592

A DNA sequence (GASx701) was identified in *S.pyogenes* <SEQ ID 7683> which encodes the amino acid sequence <SEQ ID 7684>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1017(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2593

A DNA sequence (GASx702) was identified in *S.pyogenes* <SEQ ID 7685> which encodes the amino acid sequence <SEQ ID 7686>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.03   Transmembrane    2 -  18 (   1 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2594

A DNA sequence (GASx703) was identified in *S.pyogenes* <SEQ ID 7687> which encodes the amino acid sequence <SEQ ID 7688>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -3.45   Transmembrane   36 -  52 (  36 -  55)

----- Final Results -----
               bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC39287 GB:AF115103 orf87 gp [Streptococcus thermophilus
         bacteriophage Sfi21]
 Identities = 43/73 (58%), Positives = 61/73 (82%)

Query: 1  MINLKLRLQNKVTLMAILGAIFLLAQQLGIKLPSNIADIANTAVTLLVLLGVVTDPTTKG 60
          MIN KLRLQNK TL+A++ A+FL+ QQ G+ +P+NI +  NT V +LV+LG++TDPTTKG
Sbjct: 8  MINFKLRLQNKATLVALISAVFLMLQQFGLHVPNNIQEGINTLVGILVILGIITDPTTKG 67

Query: 61 LSDSEQALTYHEP 73
          ++DSE+AL+Y +P
Sbjct: 68 IADSERALSYIQP 80
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2595

A DNA sequence (GASx707R) was identified in *S.pyogenes* <SEQ ID 7689> which encodes the amino acid sequence <SEQ ID 7690>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.35   Transmembrane    9 -  25 (   1 -  27)

----- Final Results -----
               bacterial membrane --- Certainty=0.5140(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2596

A DNA sequence (GASx714R) was identified in *S.pyogenes* <SEQ ID 7691> which encodes the amino acid sequence <SEQ ID 7692>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1401(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2597

A DNA sequence (GASx715) was identified in *S.pyogenes* <SEQ ID 7693> which encodes the amino acid sequence <SEQ ID 7694>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0417(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2598

A DNA sequence (GASx726) was identified in *S.pyogenes* <SEQ ID 7695> which encodes the amino acid sequence <SEQ ID 7696>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.17   Transmembrane   18 -  34 (  18 -  35)

----- Final Results -----
               bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2599

A DNA sequence (GASx728R) was identified in *S.pyogenes* <SEQ ID 7697> which encodes the amino acid sequence <SEQ ID 7698>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1795(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF61314 GB:U96166 unknown [Streptococcus cristatus]
 Identities = 149/194 (76%), Positives = 162/194 (82%)

Query: 1   LSAIIRQSTSKRISDKRGIYLVEKLVSLAKQSYFTVTKTSPMIEEVRYYAKELLRLSERR 60
           L  IIRQSTSKRIS+KR  YL +KL+ LAKQS+  V KTSPM+EEVRYYA+ELLRLSERR
Sbjct: 56  LYEIIRQSTSKRISEKRIAYLTDKLIKLAKQSFCAVKKTSPMLEEVRYYAQELLRLSERR 115

Query: 61  QAIFDKMVASAQPLPEDKILRSIPSIVETTATSIIGELGAIRRFQSANQINAFIGIDFRH 120
           Q + + MVA AQPLPE  ILRSIP I ETTATSIIGELG I RFQS NQ NAFIGID RH
Sbjct: 116 QVVLNDMVALAQPLPEYDILRSIPGIAETTATSIIGELGDIHRFQSTNQFNAFIGIDLRH 175

Query: 121 YESGNYLAQEHITKRGNPYAPKILFKCIHDIAFASHTNPCHIADFYEKRKRQSQTASTKP 180
           YES N+LA+EHITKRGNPYA KILFKCIH+IA ASHTNPCHIADFYEKRKRQS  ASTKP
Sbjct: 176 YESRNFLAKEHITKRGNPYARKILFKCIHNIASASHTNPCHIADFYEKRKRQSTIASTKP 235

Query: 181 HTIASRHCLVRQCF 194
            TIAS H L+R  +
Sbjct: 236 LTIASIHRLIRTMY 249
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2600

A DNA sequence (GASx729R) was identified in *S.pyogenes* <SEQ ID 7699> which encodes the amino acid sequence <SEQ ID 7700>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2363(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2601

A DNA sequence (GASx730R) was identified in *S.pyogenes* <SEQ ID 7701> which encodes the amino acid sequence <SEQ ID 7702>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2602

A DNA sequence (GASx734) was identified in *S.pyogenes* <SEQ ID 7703> which encodes the amino acid sequence <SEQ ID 7704>. Analysis of this protein sequence reveals the following:

```
Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4001(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2603

A DNA sequence (GASx735) was identified in *S.pyogenes* <SEQ ID 7705> which encodes the amino acid sequence <SEQ ID 7706>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.66   Transmembrane  276 - 292 ( 274 - 292)

----- Final Results -----
               bacterial membrane --- Certainty=0.2466(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2604

A DNA sequence (GASx736) was identified in *S.pyogenes* <SEQ ID 7707> which encodes the amino acid sequence <SEQ ID 7708>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3998(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2605

A DNA sequence (GASx737) was identified in *S.pyogenes* <SEQ ID 7709> which encodes the amino acid sequence <SEQ ID 7710>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.74   Transmembrane   77 -  93 (  69 -  99)
   INTEGRAL    Likelihood = -4.14   Transmembrane  152 - 168 ( 151 - 170)
   INTEGRAL    Likelihood = -1.17   Transmembrane  196 - 212 ( 194 - 212)

----- Final Results -----
               bacterial membrane --- Certainty=0.6095(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2606

A DNA sequence (GASx738) was identified in *S.pyogenes* <SEQ ID 7711> which encodes the amino acid sequence <SEQ ID 7712>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-13.16   Transmembrane   44 -  60 (  39 -  71)
   INTEGRAL    Likelihood =-10.24   Transmembrane   94 - 110 (  81 - 114)
   INTEGRAL    Likelihood = -7.64   Transmembrane  185 - 201 ( 179 - 207)
   INTEGRAL    Likelihood = -7.48   Transmembrane  132 - 148 ( 130 - 158)
   INTEGRAL    Likelihood = -2.76   Transmembrane  208 - 224 ( 204 - 225)
   INTEGRAL    Likelihood = -0.06   Transmembrane  153 - 169 ( 152 - 169)

----- Final Results -----
               bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2607

A DNA sequence (GASx742) was identified in *S.pyogenes* <SEQ ID 7713> which encodes the amino acid sequence <SEQ ID 7714>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.80   Transmembrane  887 - 903 ( 882 - 906)
   INTEGRAL    Likelihood = -4.88   Transmembrane    6 -  22 (   5 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 877-881
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB46409 GB:AL096743 putative large secreted protein
          [Streptomyces coelicolor A3(2)]
 Identities = 231/599 (38%), Positives = 329/599 (54%), Gaps = 43/599 (7%)

Query: 278 TSSNSDASSRNIVKIGEIQGASHTSPLLKKAVTVEQVVVTYL---DDSTHFYVQDLNGDG 334
           T +++ +++   V+I ++QG++  SP   + VT     +VT +     S  F++QD   D
Sbjct: 28  TPAHAASAAAGPVRIHDVQGSTRLSPYAGEQVTDVAGIVTGVRGYGSSKGFWMQDPLPDA 87
```

```
Query: 335 DLATSDGIRVFAKNA-KVQVGDVLTISGEVEEFFGRGYEERKQTDLTITQIVAKAVTK-T 392
           D ATS+G+ VF   A +V VGD +T+SG V E+   G     Q+   +T+I    VT  +
Sbjct: 88  DPATSEGVFVFTSRAPEVAVGDAVTVSGTVSEYVPGGTSSGNQS---LTEITRPTVTVVS 144

Query: 393 GTAQVPSPLVLGKDRIAPANIIDNDGLR-------VFDPEEDAIDYWESMEGMLVAVDDA 445
           G   +P+   +    + A   + DG             P   A+DY+ES+EGM V V DA
Sbjct: 145 GGNAIPAATTVSARSVPRAYAPEGDGAANGSVNALPLRPGTYALDYYESLEGMNVRVADA 204

Query: 446 KILGPMKN-KEIYVLPGSSTRPLNNSGGVLLPANSYNTDVIPVLFKKGKQI----IKAGD 500
           +++G      E++V       P    G V    +  NT  + +    GK         GD
Sbjct: 205 RVVGASDPYTELWVTVKPWENPNRRGGTVYGSYDDQNTGRLQIQ-SLGKPADFPAADVGD 263

Query: 501 SYKGRLAGPVSYS-YGNYKVFVDDSKNMPSLMDGHLKPEKTNLQKDLSKLSIASYNIENF 559
           +  G  AGP+ Y+ YG Y +   +   + +L  G  + E T  Q     +L++A+YN+EN
Sbjct: 264 TLAGTTAGPLDYNQYGGYTLVASE---IGALESGGTERESTRRQS-ARELAVATYNVENL 319

Query: 560 SANPSSTKDEKVKRIAESFIHDLNAPDIIGLIEVQDNNGPTDDGTTDATQSAQRLIDAIK 619
             +PS   D+     AE+ +H L +PDI+ L E+QDNNG TDDGT  A  +  RLIDAI
Sbjct: 320 --DPS---DDTFTAHAETIVHRLKSPDIVSLEEIQDNNGATDDGTVAADATVGRLIDAIV 374

Query: 620 KLGGPTYRYVDIAPENNVDGGQPGGNIRTGFLYQPERVSLSDKPKGGARDA--LTWVNGE 677
             GGP Y +  I P +  DGGQPGGNIR   FL+ PERVS +D+  G A  A  +  V G+
Sbjct: 375 AAGGPRYDWRGIDPVDKADGGQPGGNIRQAFLFNPERVSFTDRAGGDATTATGVRKVRGK 434

Query: 678 --LNLSVGRIDPTNAAWKDVRKSLAAEFIFQGRKVVVVANHLNSKRGDNALYGCVQPVTF 735
             L  S GR+DP N AW+D RK LA EF+F+GR V VVANH NSK GD  L    QP +
Sbjct: 435 AALTHSPGRVDPANEAWEDSRKPLAGEFVFRGRTVFVVANHFNSKGGDQGLTAQYQPPSR 494

Query: 736 KSEQRRHVLANMLAQFAKE--GAKHQANIVMLGDFNDFEFTKTIQLIE-EGDMVNLVSRH 792
            SE +RH  A ++  F KE   A+  A++V LGD NDFEF++T +++E +G + + V
Sbjct: 495 GSETQRHAQAKVVNTFVKEILAAQKNADVVALGDINDFEFSRTARILEGDGALWSAVKSL 554

Query: 793 DISDRYSYFHQGNNQTLDNILVSRHLL--DHYEFDMVHVNSPFMEAHGRASDHDPLLLQ 849
             S+RYSY +QGN+Q LD ILVS  +    H  +D VHVN+ F   H + SDHDP +L+
Sbjct: 555 PRSERYSYVYQGNSQVLDQILVSPSVRRGGHLSYDSVHVNAEF---HDQISDHDPQVLR 610
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2608

A DNA sequence (GASx743) was identified in *S.pyogenes* <SEQ ID 7715> which encodes the amino acid sequence <SEQ ID 7716>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2437(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2609

A DNA sequence (GASx756) was identified in *S.pyogenes* <SEQ ID 7717> which encodes the amino acid sequence <SEQ ID 7718>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.30   Transmembrane   10 -  26 (   8 -  27)
   INTEGRAL    Likelihood = -3.08   Transmembrane   51 -  67 (  50 -  67)

----- Final Results -----
               bacterial membrane --- Certainty=0.2720(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2610

A repeated DNA sequence (GASx758) was identified in *S.pyogenes* <SEQ ID 7719> which encodes the amino acid sequence <SEQ ID 7720>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA38133 GB:X54225 7 kDa protein [Streptococcus pneumoniae]
 Identities = 31/61 (50%), Positives = 41/61 (66%)

Query: 1  MTNGLKYVLEQMLLLFIIAALACLFLAIGLMIGYSFMGDGQSPWHILSMDKWAELVNKFT 60
          M     YV++++LL+ I+  L  L L IGLM+GY   +G GQ PW ILS  KW EL++KFT
Sbjct: 3  MNKKSSYVVKRLLLVIIVLILGTLALGIGLMVGYGILGKGQDPWAILSPAKWQELIHKFT 62

Query: 61 G 61
          G
Sbjct: 63 G 63
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2611

A DNA sequence (GASx764) was identified in *S.pyogenes* <SEQ ID 7721> which encodes the amino acid sequence <SEQ ID 7722>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence
```

```
   INTEGRAL    Likelihood = -3.98   Transmembrane   47 -  63 (  46 -  67)

----- Final Results -----
                bacterial membrane --- Certainty=0.2593(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9149> which encodes the amino acid sequence <SEQ ID 9150>. Analysis of this protein sequence reveals the following:

```
     Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.98   Transmembrane   35 -  51 (  34 -  55)

----- Final Results -----
                bacterial membrane --- Certainty=0.2593(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2612

A DNA sequence (GASx783) was identified in *S.pyogenes* <SEQ ID 7723> which encodes the amino acid sequence <SEQ ID 7724>. Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.16   Transmembrane  142 - 158 ( 132 - 167)
   INTEGRAL    Likelihood =-12.26   Transmembrane  113 - 129 ( 101 - 140)
   INTEGRAL    Likelihood =-10.24   Transmembrane  238 - 254 ( 233 - 260)
   INTEGRAL    Likelihood = -2.76   Transmembrane   34 -  50 (  34 -  51)

----- Final Results -----
               bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA32091 GB:AB010970 ABC-transporter [Streptococcus mutans]
 Identities = 173/269 (64%), Positives = 214/269 (79%), Gaps = 2/269 (0%)

Query: 1   MNFLTKKNRILLREMVKTDFKLRYQGSAIGYLWSILKPLMMFTIMYLVFIRFLRLGGNVP 60
           M+F ++KNRILL+E++KTDFKLRYQGSAIGYLWSILKPLM+F IMY+VF+RFL LGG+VP
Sbjct: 1   MDFFSRKNRILLKELIKTDFKLRYQGSAIGYLWSILKPLMLFAIMYIVFVRFLPLGGDVP 60

Query: 61  HFPVALLLANVIWSFFSEATSMGMVSIVSRGDLLRKLNFSKHIIVFSAVLGALINFLINL 120
           H+PVALLL NVIW+FF E T MGMVS+V+RGDLLRKLNFSK  IVFSAV GA INF IN+
Sbjct: 61  HWPVALLLGNVIWTFFQETTMMGMVSVVTRGDLLRKLNFSKQTIVFSAVSGAAINFGINV 120

Query: 121 VVVLIFALINGVTIS--GYAYLSLFLFIELVVLVLGIALLLSNVFVYYRDLAQVWEVLLQ 178
           +VVLIFAL+NGVT +     +L + LF+EL++   GIA +LS ++V YRD+  VWEV+LQ
Sbjct: 121 IVVLIFALLNGVTFTFRWNLFLLIPLFLELLLFSTGIAFILSTLYVRYRDIGPVWEVILQ 180

Query: 179 AGMYATPIIYPITFVLDSHPLAAKLLMLNPVAQMIQDFRYLLIDRANVTIWQMSTNWFYI 238
```

```
            G Y TPIIY +T++      + AKLL+L+P+AQ+IQD R++LID ANVTIWQM  +
Sbjct: 181 GGFYGTPIIYSLTYIATRSVVGAKLLLLSPIAQIIQDMRHILIDPANVTIWQMINHKSIA 240

Query: 239 VIPYLVPFVILFIGIFVFKKNADRFAEII 267
           VIPYLVP  +  IG  VF  NA +FAEII
Sbjct: 241 VIPYLVPIFVFIIGFLVFNYNAKKFAEII 269
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2613

A DNA sequence (GASx786) was identified in *S.pyogenes* <SEQ ID 7725> which encodes the amino acid sequence <SEQ ID 7726>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3828(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA32094 GB:AB010970 rgpFc [Streptococcus mutans]
 Identities = 381/582 (65%), Positives = 475/582 (81%), Gaps = 1/582 (0%)

Query: 1   MNRILLYVHFNKYNKISAHVYYQLEQMRSLFSKIVFISNSKVSHEDLKRLKNHCLIDEFL 60
           M R+LLYVHFNKYN++S+HV YQL QMRSLFSK++FISNS+V+  D+K L+   LID+F+
Sbjct: 1   MKRLLLYVHFNKYNRVSSHVVYQLTQMRSLFSKVIFISNSQVADADVKMLREKHLIDDFI 60

Query: 61  QRKNKGFDFSAWHDGLIIMGFDKLEEFDSLTIMNDTCFGPIWEMAPYFENFEEKETVDFW 120
           QR+N GFDF+AW DG++ +GFD+L  +DS+T MNDTCFGP+WEM   ++ FE K TVDFW
Sbjct: 61  QRQNSGFDFAAWRDGMVFVGFDELVTYDSVTTMNDTCFGPLWEMYSIYQEFETKTTVDFW 120

Query: 121 GITNNRGTKAFKEHVQSYFMTFKNQVIQNKVFQQFWQSIIEYENVQEVIQHYETQLTSIL 180
           G+TNNR TK+F+EH+QSYF++FK  V+++  F+ FW++I EY++VQ+VI  YET++T+ L
Sbjct: 121 GLTNNRATKSFREHIQSYFISFKASVLRSTAFRDFWENIKEYQDVQKVIDQYETKVTTTL 180

Query: 181 LNEGFSYQTVFDTRKAESSFMPHPDFSYYNPTAILKHHVPFIKVKAIDANQHIAPYLLNL 240
           L+ GF Y  VFDT K ++S M H DFSYYNPTAIL H VPFIKVKAID NQHI PYLLN
Sbjct: 181 LDAGFQYDVVFDTTKEDASHMLHADFSYYNPTAILNHRVPFIKVKAIDNNQHITPYLLND 240

Query: 241 IRETTNYPIDLIVSHMSQISLPDTKYLLSQKYLNCQRLAKQTCQKVAVHLHVFYVDLLDE 300
           I++ + YPIDLIVSHMS+I+ PD  YLL  KY+  +        QKVAVHLHVFYVDLL+E
Sbjct: 241 IQKNSTYPIDLIVSHMSEINYPDFSYLLGHKYVKKRERVDLKNQKVAVHLHVFYVDLLEE 300

Query: 301 FLTAFENWNFHYDLFITTDSDIKRKEIKEILQRKGKTADIRVTGNRGRDIYPMLLLKDKL 360
           FLTAF+ ++F YDLFITTDSD K+ EI+EIL   G+ A + VTGN GRD+ PML LK+ L
Sbjct: 301 FLTAFKQFHFSYDLFITTDSDDKKAEIEEILSANGQEAQVFVTGNIGRDVLPMLKLKNYL 360

Query: 361 SQYDYIGHFHTKKSKEADFWAGESWRKELIDMLVKPADSILSAFETD-DIGIIIADIPSF 419
           S YD++GHFHTKKSKEADFWAG+SWR+ELIDMLVKPAD+IL+  + +  IG++IAD+P+F
Sbjct: 361 SAYDFVGHFHTKKSKEADFWAGQSWREELIDMLVKPADNILAQLQQNPKIGLVIADMPTF 420

Query: 421 FRFNKIVNAWNEHLIAQEMMSLWRKMDVKKQIDFQAMDTFVMSYGTFVWFKYDALKSLFD 479
           FR+NKIV+AWNEHLIA EM +LW+KM + K+IDF A  TFVMSYGTFVWFKYDALK LFD
Sbjct: 421 FRYNKIVDAWNEHLIAPEMNTLWQKMGMTKKIDFNAFHTFVMSYGTFVWFKYDALKPLFD 480

Query: 480 LELTQNDIPSEPLPQNSILHAIERLLVYIAWGDSYDFRIVKNPYELTPFIDNKLLNLRED 539
           L LT +D+P EPLPQNSILHAIERLL+YIAW + YDFRI KNP +LTPFIDNKLLN R +
Sbjct: 481 LNLTDDDVPEEPLPQNSILHAIERLLIYIAWNEHYDFRISKNPVDLTPFIDNKLLNERGN 540
```

```
Query: 540 EGAHTYVNFNQMGGIKGALKYIIVGPAKAMKYIFLRLMEKLK 581
           +T+V+FN MGGIKGA KYI +GPA+A+KYI  R ++K+K
Sbjct: 541 SAPNTFVDFNYMGGIKGAFKYIFIGPARAVKYILKRSLQKIK 582
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2614

A DNA sequence (GASx787) was identified in *S.pyogenes* <SEQ ID 7727> which encodes the amino acid sequence <SEQ ID 7728>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-15.66   Transmembrane  202 - 218 ( 191 - 224)
   INTEGRAL    Likelihood =-10.03   Transmembrane  340 - 356 ( 335 - 365)
   INTEGRAL    Likelihood = -9.08   Transmembrane  270 - 286 ( 263 - 289)
   INTEGRAL    Likelihood = -8.60   Transmembrane  124 - 140 ( 118 - 145)
   INTEGRAL    Likelihood = -4.94   Transmembrane  377 - 393 ( 375 - 395)
   INTEGRAL    Likelihood = -3.29   Transmembrane  291 - 307 ( 290 - 311)
   INTEGRAL    Likelihood = -2.87   Transmembrane  160 - 176 ( 159 - 180)
   INTEGRAL    Likelihood = -2.66   Transmembrane   50 -  66 (  48 -  66)
   INTEGRAL    Likelihood = -1.28   Transmembrane   77 -  93 (  76 -  93)
   INTEGRAL    Likelihood = -0.69   Transmembrane  229 - 245 ( 229 - 245)

----- Final Results -----
               bacterial membrane --- Certainty=0.7262(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA32095 GB:AB010970 ORF7 [Streptococcus mutans]
 Identities = 374/775 (48%), Positives = 525/775 (67%), Gaps = 7/775 (0%)

Query: 53  VSFVGYIISLIGLSYYLSRQVSRQLFLKTSFIVISYLIVSYWVQITQHLNDKRFDIWSLT 112
           V  V Y++S++GLS+YLS+ + +  F++       Y+++SY++ +T+ LN++ F IW L
Sbjct: 30  VCLVIYVLSILGLSFYLSKNLKKTFFIELLLGYGLYIVISYFLAVTRELNNESFKIWDLA 89

Query: 113 KNQFYQFQALPSLLIILV----MATLIKILAAYFAIEKDRFGLL-GYQGNTFSVALILAV 167
           KN F+Q   LP+L++I+     + LI++      + +    LL  +    F +  ++
Sbjct: 90  KNHFFQPYFLPTLVLIIACTFALNYLIRVKMKRSHLSRKMTLLLENFSETEFLLTGLIVS 149

Query: 168 VPINDIHLLKLISSRFSELVTAGNSQIALLKISGLLIVLLVIFATIIYVVLNALKHLKSN 227
             ++D   +KL+            +  +LL +  LL    L++F+ I+    NA + +K N
Sbjct: 150 FILSDTLYVKLLQESLRAYYHKPLAYESLLFLYTLLT--LILFSVIVEACFNAYRSIKLN 207

Query: 228 KPSFSVAATTSLFLALVFNYTFQYGVKGDEALLGYYVFPGATLFQIVAITLVALLAYVIT 287
           +P+ S+A  +SL  A +FNY FQYG+K D  LLG Y+ PGAT +QI+ +T      Y+I
Sbjct: 208 RPNLSLAFVSSLLFATIFNYAFQYGLKNDADLLGKYIVPGATAYQILVLTAAGFFLYLII 267

Query: 288 NRYWPTTFFLLILGTIISVVNDLKESMRSEPLLVTDFVWLQELGLVTSFVKKSVIVEMVV 347
           NRY   TF ++ILG+II+VVN LK  MR+EPLLVTDF W+  + L+   V  ++I   ++
Sbjct: 268 NRYLLVTFLIVILGSIITVVNVLKVGMRNEPLLVTDFAWVTNIRLLARSVNANIIFSTLL 327

Query: 348 GLAICIVVAWYLHGRVLAGKLFMSPVKRASAVLGLFIVSCSMLIPFSYEKEGKILSGLPI 407
            LA  I++  +L  R+L GK+  +  +  + + ++  S+ I F  EK  KI++G+P+
Sbjct: 328 ILAALILLYLFLRKRLLQGKITENYRLKVGLISSICLLGFSIFIIFRNEKGSKIVNGIPV 387

Query: 408 ISALNNDNDINWLGFSTNARYKSLAYVWTRQVTKKIMEKPTNYSQETIASIAQKYQKLAE 467
           IS +NN  DI + GF +NA YKSL YVWT+QVTK IM+KP++YS+E I  +A+KY  +A
Sbjct: 388 ISQVNNWVDIGYQGFYSNASYKSLMYVWTKQVTKSIMDKPSDYSKERILKLAKKYNNVAN 447
```

```
Query: 468 DINKDRKNNIADQTVIYLLSESLSDPDRVSNVTVSHDVLPNIKAIKNSTTAGLMQSDSYG 527
           INK R  NI++QTVIY+LSES SDPDRV  V +S DV+PNIK IK  TT+GLM SD YG
Sbjct: 448 KINKVRTENISNQTVIYILSESFSDPDRVQGVNLSRDVIPNIKQIKEKTTSGLMHSDGYG 507

Query: 528 GGTANMEFQTLTSLPFYNFSSSVSVLYSEVFPKMAKPHTISEFYQGKNRIAMHPASANNF 587
           GGTANMEFQ+LT LP+YNF+SSVS LY+EV P M+   +IS  ++ KNR+ +HP+SA+N+
Sbjct: 508 GGTANMEFQSLTGLPYYNFNSSVSTLYTEVVPDMSVFPSISNQFKSKNRVVIHPSSASNY 567

Query: 588 NRKTVYSNLGFSKFLALSGSKDKFKNIENVGLLTSDKTVYNNILSLINPSESQFFSVITM 647
           +RK VY  L F  F+A SG+ DK  + E VGL  SDKT Y NIL  INPS+SQFFSV+TM
Sbjct: 568 SRKYVYDKLKFPTFVASSGTSDKITHSEKVGLNVSDKTTYQNILDKINPSQSQFFSVMTM 627

Query: 648 QNHIPWSSDYPEEIVAEGKNFTEEENHNLTSYARLLSFTDKETRAFLEKLTQINKPITVV 707
           QNH+PW+SD P ++VA GK +T++EN +L+SYARLL++TDKET+ FL +L+Q+   +TVV
Sbjct: 628 QNHVPWASDEPSDVVATGKGYTKDENGSLSSYARLLTYTDKETKDFLAQLSQLKHKVTVV 687

Query: 708 FYGDHLPGLYPDSAFNKHIENKYLTDYFIWSNGTNEKKNHPLINSSDFTAALFEHTDSKV 767
           FYGDHLPGLYP+SAF K  +++Y TDYFIWSN  + NH  +NSSDFTA L EHT+SKV
Sbjct: 688 FYGDHLPGLYPESAFKKDPDSQYQTDYFIWSNYNTKTLNHSYVNSSDFTAELLEHTNSKV 747

Query: 768 SPYYALLTEVLNKASVDKSPDSPEVKAIQNDLKNIQYDVTIGKGYLLKHKTFFKI 822
           SPYYALLTEVL+  +V     + E K I NDLK IQYD+T+GKGY+  +K FF I
Sbjct: 748 SPYYALLTEVLDNTTVGHGKLTKEQKEIANDLKLIQYDITVGKGYIRNYKGFFDI 802
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2615

A DNA sequence (GASx789R) was identified in *S.pyogenes* <SEQ ID 7729> which encodes the amino acid sequence <SEQ ID 7730>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.06   Transmembrane   42 -  58 (  42 -  58)

----- Final Results -----
               bacterial membrane --- Certainty=0.1426(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2616

A DNA sequence (GASx790) was identified in *S.pyogenes* <SEQ ID 7731> which encodes the amino acid sequence <SEQ ID 7732>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2617

A DNA sequence (GASx791) was identified in *S.pyogenes* <SEQ ID 7733> which encodes the amino acid sequence <SEQ ID 7734>. Analysis of this protein sequence reveals the following:

```
Possible site: 48

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.42   Transmembrane  166 - 182 ( 157 - 188)
   INTEGRAL    Likelihood = -7.32   Transmembrane   85 - 101 (  79 - 104)
   INTEGRAL    Likelihood = -6.90   Transmembrane  397 - 413 ( 386 - 417)
   INTEGRAL    Likelihood = -6.05   Transmembrane  253 - 269 ( 252 - 273)
   INTEGRAL    Likelihood = -5.26   Transmembrane  301 - 317 ( 293 - 325)
   INTEGRAL    Likelihood = -3.35   Transmembrane  363 - 379 ( 362 - 379)
   INTEGRAL    Likelihood = -3.24   Transmembrane  335 - 351 ( 335 - 351)

----- Final Results -----
              bacterial membrane --- Certainty=0.5967(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA64645 GB:U10927 CapF [Staphylococcus aureus]
 Identities = 97/419 (23%), Positives = 186/419 (44%), Gaps = 40/419 (9%)

Query: 12  FLWNMLGSLSTAVISVILLMVVTRLLTSADSDIYAFAYSFANMMVVVGLFQVRNYQATDI 71
           F +  + ++ +A+    ++L+V+ RL T  D   Y +A      + +    ++R+   T
Sbjct: 5   FNYMFVANILSALCKFLILLVIVRLGTPEDVGRYNYALVITAPIFLFISLKIRSVIVT-- 62

Query: 72  NEKYSFSQYLVARLMTCLLMLAITVIYLTLTKTDSYKSTIVFLVCFYRSTDAFSDLYQGM 131
           N+KYS ++Y+ A L   ++ L    I++ +        T + +V   +  +   ++ G+
Sbjct: 63  NDKYSPNEYISAILSLNIITLIFVAIFVYVLGNGDL--TTILIVSLIKLFENIKEVPYGI 120

Query: 132 FQQHERLDIAGKSLAYRNTLIFMVVTAIILYSKNLTLALVAVCIVSLVFIMYYDIGHSKK 191
           +Q++E L + G S+   N L  +++  I  +S NL +AL+ + I  +      D  +  K
Sbjct: 121 YQKNESLKLLGISMGIYNILSLILFYIIYSFSHNLNMALLFLVISCIFSFAIIDRWYLSK 180

Query: 192 FQKLMFSELLSNISFQNSLKLLKESF----PLFLNGFLIIYIYTQPKYAIELMTTLGEVA 247
           +  +        + + N++   KE F    PL  +  L       P+  +E +   G+
Sbjct: 181 YYNI-------KLHYNNNIAKFKEIFILTIPLAFSSALGSLNTGIPRIVLENL--FGKYT 231

Query: 248 LGS-QTIFNILFMPAFVMNLLILFFRPHITQMAIALIRGQIK-EFNKIQVQLFAYLGVF- 304
           LG   TI  +L +     N +   F P + +     L + + K EF K+  ++  ++G+F
Sbjct: 232 LGIFSTIAYVLVIGGLFANSISQVFLPKLRK----LYKDEKKIEFEKLTRKM-VFIGIFI 286

Query: 305 SLIALVGSGLFGIPFLSILYG-----TNLTDYWVDF-MLIMLGGSIGSFATVIDNILTAM 358
            + +++ S   G   LS+L+G       N+    + F +L +L G       +
Sbjct: 287 GMCSVILSLFLGEALLSLLFGKEYGENNIILIILSFGLLFILSGIFLGTTIIATGKYNVN 346

Query: 359 RKQQLLLLIPYTGGFLISLLITNLFVMKYHILGAALSFLITMLVWLGLSIMIYLFIMNRF 417
            K  L+L+     F I L+ + L + KY +LGAAL+  I+  V L   I  Y F    F
Sbjct: 347 YKISLILL-----FCI-LIFSFLLIPKYSLLGAALTITISQFVAL---ISYYYFYKRIF 396
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2618

A DNA sequence (GASx792) was identified in *S.pyogenes* <SEQ ID 7735> which encodes the amino acid sequence <SEQ ID 7736>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.03   Transmembrane   64 -  80 ( 60 -  84)
   INTEGRAL    Likelihood = -9.66   Transmembrane   43 -  59 ( 37 -  63)
   INTEGRAL    Likelihood = -8.70   Transmembrane  232 - 248 ( 229 - 251)
   INTEGRAL    Likelihood = -8.28   Transmembrane  410 - 426 ( 402 - 432)
   INTEGRAL    Likelihood = -6.21   Transmembrane  298 - 314 ( 296 - 322)
   INTEGRAL    Likelihood = -6.21   Transmembrane  478 - 494 ( 471 - 496)
   INTEGRAL    Likelihood = -5.04   Transmembrane  265 - 281 ( 256 - 288)
   INTEGRAL    Likelihood = -3.29   Transmembrane  380 - 396 ( 378 - 397)
   INTEGRAL    Likelihood = -2.92   Transmembrane  210 - 226 ( 209 - 227)
   INTEGRAL    Likelihood = -2.60   Transmembrane  187 - 203 ( 187 - 204)
   INTEGRAL    Likelihood = -2.50   Transmembrane  442 - 458 ( 439 - 458)
   INTEGRAL    Likelihood = -1.65   Transmembrane   18 -  34 (  18 -  35)
   INTEGRAL    Likelihood = -1.38   Transmembrane  165 - 181 ( 165 - 181)

----- Final Results -----
              bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA19642 GB:AB002668 unnamed protein product [Actinobacillus
          actinomycetemcomitans]
 Identities = 116/459 (25%), Positives = 207/459 (44%), Gaps = 60/459 (13%)

Query: 69  FILVFGTISAIISPINDIPDEYVHYSRTVYISEGDINLTNNNKKLRISKDVDKLI----- 123
           FIL F  I  II+P    PDE+ H+ R   IS G I  ++  K   + K + K++
Sbjct: 16  FILTF-IIGVIITPPYQSPDEFYHFQRGYAISNGQIIPSSTEK---LDKAMMKMLSIYEG 71

Query: 124 ----KQSGKTFITSNLKATKHSTREYSYPYIKGTNAYYSFSYIPQALGILVGNALDLPIL 179
               ++  T    N        +EY       TN Y+   Y+PQALG  +G+ LDL +
Sbjct: 72  IPYRSENKVTHFLENEAQNVAWEKEYILDESANTNVYFPLIYLPQALGSFLGSTLDLSLY 131

Query: 180 LTYYFGRLCN-LISYAMLAFIAIKLSGSFKQVIAVVTLLPMNIYLAASFNQDGFAIGLVL 238
             YY  ++   L+S A+L F +++   S    + ++  LPM ++   S N D      ++
Sbjct: 132 NMYYLAKIFTLLVSIAILYFASVQYRLSIP--VLLILSLPMTMFQMGSTNPDS-----II 184

Query: 239 VTIGLFI-NLLSSKDKSNYNTKFFLYLVLCGLL------VLSKFTYFLLVCLPLFIPNEK 291
           ++ +FI +LL+    SNYN   F +   C LL      V  KF   +L+ LP FI   +
Sbjct: 185 FSLSVFIGSLLARGLDSNYN---FTHKDFCKLLFSIFLCVTVKFNMLVLLLLPFFISKRR 241

Query: 292 FGKNTKLVILKKLGGLLLIFLFAAMWFRLYGQVKTPYVADFLKEV----NVSQQVKNMLE 347
           ++  +  +  +L  + A          K   + +F   +    ++ +  KN L
Sbjct: 242 EIRHGSMYSIFIIILSILWIVLAMKLTEAQSHFKEGALHNFSYYIFHMDDLFEIFKNTLN 301

Query: 348 SPIVYSSIIIRHMVINLINMNNIFQFGA-LSYGITNLFPLYVCFFFFVYISNASKITINI 406
           + Y   ++R  +  L  ++  F     L +G T+L       + F++I N  K+   I
Sbjct: 302 --LTYLKSLLRMFLGVLGWVDTKFTINEYLFFGSTSLLA-----YIFLFIHNLYKLKYVI 354

Query: 407 VEKM--GIIFVISAIIGATVLAMYLTWTPVGSSTVLGVQSRYLIGIIPLVLLLFSS---- 460
          V  +  G++F+ +  I      + +T+  +G++ ++GVQ RY    IP++L++FSS
Sbjct: 355 VSVLLVGVVFLFTHFI------LLITYNEIGTTQIVGVQGRY---FIPIMLIIFSSFILK 405
```

```
Query: 461 QQQKFKQIEDILSDKLAIHVSLLFILAMLM--STIFRYY 497
          + +K    + I    + +    LFI + +   + + RYY
Sbjct: 406 KSEKTSNNKTISKYFIIVPFLFLFISSFITINTLVSRYY 444
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2619

A DNA sequence (GASx797) was identified in *S.pyogenes* <SEQ ID 7737> which encodes the amino acid sequence <SEQ ID 7738>. Analysis of this protein sequence reveals the following:

```
Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1491(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC83961 GB:L47648 cytidine monophosphate kinase [Bacillus subtilis]
 Identities = 116/220 (52%), Positives = 156/220 (70%), Gaps = 1/220 (0%)

Query: 2   KAIKIAIDGPASSGKSTVAKIIAKNLGYTYLDTGAMYRSATYIALTHGYTGKEVALILEE 61
           K + IAIDGPA++GKSTVAKI+A+   Y Y+DTGAMYR+ TY AL       +   + E
Sbjct: 3   KKLSIAIDGPAAAGKSTVAKIVAEKKSYIYIDTGAMYRAITYAALQENVDLTDEEKLAEL 62

Query: 62  LEKNPIFFKKAKDGSQLVFLGDEDVTLAIRQNDVTNNVSWISALPEIREELVHQQRRIAQ 121
           L++  I      KDG Q VF+    DVT AIR ++++N VS  +    +REE+V +Q+++ +
Sbjct: 63  LKRTDIELITTKDG-QKVFVNGTDVTEAIRTDEISNQVSIAAKHRSVREEMVKRQQQLGE 121

Query: 122 AGGIIMDGRDIGTVVLPDAELKIFLVASVEERAERRYKENLEKGIESDFETLKEEIAARD 181
            GG++MDGRDIGT VLP+AE+KIFL+ASVEERA+RRY+EN++KG + ++ETL EEIA RD
Sbjct: 122 KGGVVMDGRDIGTHVLPNAEVKIFLLASVEERAKRRYEENVKKGFDVNYETLIEEIARRD 181

Query: 182 YKDSHRKVSPLKAAEDALIFDTTGVSIDGVVQFIQEKAEK 221
             DS R+VSPL+ AEDAL  DTT +SI  V   I E  E+
Sbjct: 182 KLDSEREVSPLRKAEDALEIDTTSLSIQEVADKILEAVEQ 221
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2620

A DNA sequence (GASx799) was identified in *S.pyogenes* <SEQ ID 7739> which encodes the amino acid sequence <SEQ ID 7740>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4324(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA34313 GB:X16188 ribosomal protein L35 (AA 1-66) [Bacillus
          stearothermophilus]
 Identities = 46/65 (70%), Positives = 51/65 (77%)

Query: 1  MPKQKTHRASAKRFKRTGSGGLKRFRAFTSHRFHGKTKKQRRHLRKAGLVSSGDFKRIKA 60
          MPK KTHR SAKRFK+T SG LKR  A+TSH F  KTKKQ+RHLRKA LVS GDFKRI+
Sbjct: 1  MPKMKTHRGSAKRFKKTASGKLKRGHAYTSHLFANKTKKQKRHLRKATLVSPGDFKRIRQ 60

Query: 61 MVTGL 65
          M+  L
Sbjct: 61 MLDNL 65
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2621

A DNA sequence (GASx806R) was identified in *S.pyogenes* <SEQ ID 7741> which encodes the amino acid sequence <SEQ ID 7742>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5361(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2622

A DNA sequence (GASx809R) was identified in *S.pyogenes* <SEQ ID 7743> which encodes the amino acid sequence <SEQ ID 7744>. Analysis of this protein sequence reveals the following:

```
Possible site: 52

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.81   Transmembrane   33 -  49 (  28 -  53)

----- Final Results -----
              bacterial membrane --- Certainty=0.4524(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2623

A DNA sequence (GASx814R) was identified in *S.pyogenes* <SEQ ID 7745> which encodes the amino acid sequence <SEQ ID 7746>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0206(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2624

A DNA sequence (GASx817) was identified in *S.pyogenes* <SEQ ID 7747> which encodes the amino acid sequence <SEQ ID 7748>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.49   Transmembrane   16 -  32 (  15 -  32)

----- Final Results -----
            bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2625

A DNA sequence (GASx820) was identified in *S.pyogenes* <SEQ ID 7749> which encodes the amino acid sequence <SEQ ID 7750>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.11   Transmembrane   62 -  78 (  59 -  81)
   INTEGRAL    Likelihood = -6.00   Transmembrane  128 - 144 ( 123 - 147)
   INTEGRAL    Likelihood = -2.50   Transmembrane    5 -  21 (   3 -  26)

----- Final Results -----
            bacterial membrane --- Certainty=0.3845(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA26653 GB:M83994 prolipoprotein signal peptidase
          [Staphylococcus aureus]
 Identities = 57/153 (37%), Positives = 96/153 (62%), Gaps = 6/153 (3%)

Query: 1   MKKRLFVLSLILL----VALDQLSKFWIVSHIALGEVKPFIPGIVSLTYLQNNGAAFSIL 56
           M K+ F+ + IL+    V  DQ++K+ I + + +G+    IP  +++T  +NNGAA+ IL
Sbjct: 1   MHKKYFIGTSILIAVFVVIFDQVTKYIIATTMKIGDSFEVIPHFLNITSHRNNGAAWGIL 60

Query: 57  QDQQWFFVVITVLVIGYAIYYLATHPHLNIWKQLALLLIISGGIGNFIDRLRLAYVIDMI 116
             +  FF +IT++++   +Y+       N++ Q+A+ L+ +G +GNFIDR+    V+D I
Sbjct: 61  SGKMTFFFIITIIILIALVYFFIKDAQYNLFMQVAISLLFAGALGNFIDRILTGEVVDFI 120

Query: 117 HLDF--VDFAIFNVADSYLTVGVILLLICLWKE 147
               +    DF IFN+ADS LT+GVIL++I L K+
Sbjct: 121 DTNIFGYDFPIFNIADSSLTIGVILIIIALLKD 153
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2626

A DNA sequence (GASx822R) was identified in *S.pyogenes* <SEQ ID 7751> which encodes the amino acid sequence <SEQ ID 7752>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2638(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2627

A DNA sequence (GASx823R) was identified in *S.pyogenes* <SEQ ID 7753> which encodes the amino acid sequence <SEQ ID 7754>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3452(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2628

A DNA sequence (GASx828) was identified in *S.pyogenes* <SEQ ID 7755> which encodes the amino acid sequence <SEQ ID 7756>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2629

A DNA sequence (GASx836) was identified in *S.pyogenes* <SEQ ID 7757> which encodes the amino acid sequence <SEQ ID 7758>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4333(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2630

A DNA sequence (GASx853R) was identified in *S.pyogenes* <SEQ ID 7759> which encodes the amino acid sequence <SEQ ID 7760>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4906(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2631

A DNA sequence (GASx854R) was identified in *S.pyogenes* <SEQ ID 7761> which encodes the amino acid sequence <SEQ ID 7762>. Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3989(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9147> which encodes the amino acid sequence <SEQ ID 9148>. Analysis of this protein sequence reveals the following:

```
Possible site: 42

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty= 0.399(Affirmative) < succ>
               bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
                bacterial outside --- Certainty= 0.000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB59092 GB:M97157 pyrogenic exotoxin C [Streptococcus pyogenes]
 Identities = 39/67 (58%), Positives = 53/67 (78%)

Query: 1   LMESKEIYLTKSPYIRGSLEIHSKNRKHEKINLYDAKPNSTRSDVFKKYKDNKTINMKDF 60
           LM++ +IY   SPY+ G +EI +K+ KHE+I+L+D+    TRSD+F KYKDN+ INMK+F
Sbjct: 167 LMDNYKIYDATSPYVSGRIEIGTKDGKHEQIDLFDSPNEGTRSDIFAKYKDNRIINMKNF 226

Query: 61  SHFDIYL 67
           SHFDIYL
Sbjct: 227 SHFDIYL 233
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2632

A DNA sequence (GASx855R) was identified in *S.pyogenes* <SEQ ID 7763> which encodes the amino acid sequence <SEQ ID 7764>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2633

A DNA sequence (GASx856) was identified in *S.pyogenes* <SEQ ID 7765> which encodes the amino acid sequence <SEQ ID 7766>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4145(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2634

A DNA sequence (GASx862) was identified in *S.pyogenes* <SEQ ID 7767> which encodes the amino acid sequence <SEQ ID 7768>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.6285(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2635

A DNA sequence (GASx863) was identified in *S.pyogenes* <SEQ ID 7769> which encodes the amino acid sequence <SEQ ID 7770>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2636

A DNA sequence (GASx878) was identified in *S.pyogenes* <SEQ ID 7771> which encodes the amino acid sequence <SEQ ID 7772>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2637

A DNA sequence (GASx887R) was identified in *S.pyogenes* <SEQ ID 7773> which encodes the amino acid sequence <SEQ ID 7774>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1911(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2638

A DNA sequence (GASx910) was identified in *S.pyogenes* <SEQ ID 7775> which encodes the amino acid sequence <SEQ ID 7776>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4511(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2639

A DNA sequence (GASx911) was identified in *S.pyogenes* <SEQ ID 7777> which encodes the amino acid sequence <SEQ ID 7778>. Analysis of this protein sequence reveals the following:

```
Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2993(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC74707 GB:AE000259 glutathionine S-transferase [Escherichia
          coli]
 Identities = 29/137 (21%), Positives = 61/137 (44%), Gaps = 9/137 (6%)

Query: 1   LPFIAKQTLKSQLIPQDNLLAESRFNEIMDFLTGDFPLVFRPMINPHRYTISQDNQALEK 60
           + ++A      QL+   N ++  +  E ++++  +    F P+  P            E+
Sbjct: 70  MQYLADSVPDRQLLAPVNSISRYKTIEWLNYIATELHKGFTPLFRP---------DTPEE 120

Query: 61  VKQASYKRMDIAMTHLDSLIGESGHVYRDQQTIADAYAYAMALWSQKTPKSYENYPHLAA 120
            K     +++  + +++  + +   +   + TIADAY + +  W+     + E   H+AA
Sbjct: 121 YKPTVRAQLEKKLQYVNEALKDEHWICGQRFTIADAYLFTVLRWAYAVKLNLEGLEHIAA 180

Query: 121 FMAKMVEDSAVQQVLNA 137
           FM +M E   VQ  L+A
Sbjct: 181 FMQRMAERPEVQDALSA 197
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2640

A DNA sequence (GASx932R) was identified in *S.pyogenes* <SEQ ID 7779> which encodes the amino acid sequence <SEQ ID 7780>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4081(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2641

A DNA sequence (GASx935) was identified in *S.pyogenes* <SEQ ID 7781> which encodes the amino acid sequence <SEQ ID 7782>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.6304(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2642

A DNA sequence (GASx937) was identified in *S.pyogenes* <SEQ ID 7783> which encodes the amino acid sequence <SEQ ID 7784>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3503(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2643

A DNA sequence (GASx938R) was identified in *S.pyogenes* <SEQ ID 7785> which encodes the amino acid sequence <SEQ ID 7786>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2884(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2644

A DNA sequence (GASx939) was identified in *S.pyogenes* <SEQ ID 7787> which encodes the amino acid sequence <SEQ ID 7788>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2771(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2645

A DNA sequence (GASx941) was identified in *S.pyogenes* <SEQ ID 7789> which encodes the amino acid sequence <SEQ ID 7790>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2257(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2646

A DNA sequence (GASx942R) was identified in *S.pyogenes* <SEQ ID 7791> which encodes the amino acid sequence <SEQ ID 7792>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3255(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB91582 GB:AF242881 ymh [Agrobacterium tumefaciens] (ver 2)
 Identities = 75/223 (33%), Positives = 116/223 (51%), Gaps = 2/223 (0%)

Query: 38  DQNSGFNKHKRVHNLVSDILNRTQNTDNIKLVIEYVCNPLRYINEVSIFEQLRTAINIPL 97
           D +    K  R++N  +   N  +   +I   I     P R+  +   FE +R  +N  L
Sbjct: 39  DTDPQMTKRHRLYNAFASDQNSRKQRTHIIAFIRKAMKPERFARDSERFEPMRLNLNRAL 98

Query: 98  SLKGLIVSDSGQIVTTTTSKTLSEAKKRFETLDSRLKELKVHPHVLKFCTQELLQENYFH 157
           +  GL V  SG++      ++TLS+A +R    L + L     VHP VL+FC +ELL +NYFH
Sbjct: 99  AFAGLAVKASGELAAVDAAETLSQATRRALELRADLTSRGVHPDVLRFCREELLVDNYFH 158

Query: 158 AVFEASKGVFHRIRLLTGSAMDSASLIDQCFKPGEPIVIINGNKLQTLDEQSEYKGLKNL 217
           AV EA K V  +IR  TG   D A L+D+ F    P++ I  N+LQ+  E+ E +G  NL
Sbjct: 159 AVLEAVKSVADKIRQRTGLTDDGAVLVDRAFSGDAPMLAI--NELQSESEKGEQRGFSNL 216

Query: 218 LLAIAHLYRNSKAHKLKYYNPDNLNDALTALTLMSLAHNLLDS 260
           +     ++RN+ AH  + +   + DA    ++ SL H  +D+
Sbjct: 217 VKGTFSMFRNTTAHAPRIHWQMSKEDAEDLFSMFSLMHRRIDA 259
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2647

A DNA sequence (GASx943R) was identified in *S.pyogenes* <SEQ ID 7793> which encodes the amino acid sequence <SEQ ID 7794>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1526(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2648

A DNA sequence (GASx944) was identified in *S.pyogenes* <SEQ ID 7795> which encodes the amino acid sequence <SEQ ID 7796>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1427(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2649

A DNA sequence (GASx945) was identified in *S.pyogenes* <SEQ ID 7797> which encodes the amino acid sequence <SEQ ID 7798>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2578(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC98430 GB:L29324 excisionase [Streptococcus pneumoniae]
 Identities = 23/54 (42%), Positives = 40/54 (73%)

Query: 1  LIQQWEGLTVATAKQWATEMRDHPDFKQFVLNPTHRIVFIDYEGFKLFVQWKSR 54
          ++++W+GL   T  +W  EMR++  F  +V+NPTH++VFI+ EGF+ F++WK +
Sbjct: 21 ILKRWDGLNKYTLNRWIKEMRENRTFSMYVINPTHKLVFINLEGFESFLRWKQK 74
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2650

A DNA sequence (GASx946) was identified in *S.pyogenes* <SEQ ID 7799> which encodes the amino acid sequence <SEQ ID 7800>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -4.99   Transmembrane    3 -  19 (   1 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.2996(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2651

A DNA sequence (GASx950) was identified in *S.pyogenes* <SEQ ID 7801> which encodes the amino acid sequence <SEQ ID 7802>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2211(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2652

A DNA sequence (GASx951) was identified in *S.pyogenes* <SEQ ID 7803> which encodes the amino acid sequence <SEQ ID 7804>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4258(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2653

A DNA sequence (GASx952) was identified in *S.pyogenes* <SEQ ID 7805> which encodes the amino acid sequence <SEQ ID 7806>. Analysis of this protein sequence reveals the following:

```
Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2476(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF74110 GB:AF212847 ORF245 [Lactococcus lactis bacteriophage
          ul36.2]
 Identities = 82/265 (30%), Positives = 128/265 (47%), Gaps = 27/265 (10%)

Query: 1   MANQLSTQQVKRDITTDPTLLTGADIKKYFDPQNLLSEKQVGQALALCKGRNLNPFANEV 60
           MAN+L    V          L    IK+Y D     S+ ++   + LCK  N+NPF  EV
Sbjct: 1   MANELGIFSVDN--------LNMTTIKQYLDGGGKASDAELVLLINLCKQNNMNPFMKEV 52
```

```
Query: 61  YIVAYKNNSGTDFSLIVSKEAFMKRAERCEGYDGFEAGITVM-RNGEMVEIEGSLKLPDD 119
           Y + Y N        ++VS++ + KRA +    + G E G+ V+ ++G +    EG+ K  +
Sbjct: 53  YFIKYGNQPA---QIVVSRDFYRKRAFQNPNFVGIEVGVIVLNKDGVLEHNEGTFKTHEQ 109

Query: 120 VLIGGWAIVYRKDRSHRYKVTVDFNEYVKLDKYGNPRSTWKSMPGTMIRKTALVQTLREA 179
            L+G WA V+ K+       V V ++EYV++ K G+P   W + P TM+ K A  Q LR A
Sbjct: 110 ELVGAWARVHLKNTEIPVYVAVSYDEYVQM-KDGHPNKMWTNKPCTMLGKVAESQALRMA 168

Query: 180 FPDELGNMYTDIDGGDTFDAIKDVTPQETQEEVRARK---MAQIEQYKQEQ--TQKQTQK 234
           FP E    Y + +  +          P++   EV   K    AQIE + +E    +K  +
Sbjct: 169 FPAEFSGTYGEEEYPE---------PEKEPREVNGVKEPDRAQIESFDKEDYAAKKIEEL 219

Query: 235 ADTSYPVDEVSEHTDDPVQGELLDG 259
            + + P  EV E T + +  E L+G
Sbjct: 220 KEKAQPQKEVVEETGEVIDEEPLEG 244
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2654

A DNA sequence (GASx953) was identified in *S.pyogenes* <SEQ ID 7807> which encodes the amino acid sequence <SEQ ID 7808>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3413(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF74111 GB:AF212847 ORF364 [Lactococcus lactis bacteriophage
          ul36.2]
 Identities = 67/222 (30%), Positives = 120/222 (53%), Gaps = 3/222 (1%)

Query: 1   MQELQLKVTQAQVEIIDREKFEQNINEVVAKYQNYAVTAGTIKDDKQVLADLRKLKKQLS 60
           ++++++    A + I++ EKF+ +IN+VVA+Y  +  +   D++  A L KL  ++
Sbjct: 19  VKDIEIDFKPAIINILEEEKFKASINQVVAEYTGHVPSVENLTVDRKTRASLNKLITKIE 78

Query: 61  DERIKVKKELSKPADDIDGYIKQASKPLDDTIDKIATDVKEFEDHQKALRLDTVKSYLSN 120
             R ++KK ++ P  + +G+ K+A  P++  I+ I   +K+ E  QK  R   V   L
Sbjct: 79  TRRKEIKKSINVPYAEFEGWYKKAIAPMEKVIETIDAGIKKIEAEQKESRKKVVHELLVE 138

Query: 121 KASEYMLDPRIFDEKAMEYTKAGNFMADGVTLKKVTMKSLEDLVTFEYQKEQEVEKAKAT 180
             ++  +D RIF+    ++ K+ NF  + +  KK  + S+  ++  E QK  E + AK +
Sbjct: 139 LTTDTEVDSRIFENFVDDWAKSSNF--NDIKPKKQLIDSITYVIDGEKQKIAEYKSAKQS 196

Query: 181 ISGQCAEYGMTDQPYIRMLKE-MTLVEVLGQIKADYLAEKQK 221
           IS  C    +T  PYIRML    T+ E++  I  D L EKQ+
Sbjct: 197 ISDFCFGNNITSTPYIRMLDSGKTVSEIMAVITEDVLFEKQR 238
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2655

A DNA sequence (GASx954) was identified in *S.pyogenes* <SEQ ID 7809> which encodes the amino acid sequence <SEQ ID 7810>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3884(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2656

A DNA sequence (GASx955) was identified in *S.pyogenes* <SEQ ID 7811> which encodes the amino acid sequence <SEQ ID 7812>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1777(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2657

A DNA sequence (GASx956) was identified in *S.pyogenes* <SEQ ID 7813> which encodes the amino acid sequence <SEQ ID 7814>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.44   Transmembrane   82 -  98 (  81 -  98)

----- Final Results -----
             bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2658

A DNA sequence (GASx958) was identified in *S.pyogenes* <SEQ ID 7815> which encodes the amino acid sequence <SEQ ID 7816>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3673(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2659

A DNA sequence (GASx960) was identified in *S.pyogenes* <SEQ ID 7817> which encodes the amino acid sequence <SEQ ID 7818>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1852(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2660

A DNA sequence (GASx961) was identified in *S.pyogenes* <SEQ ID 7819> which encodes the amino acid sequence <SEQ ID 7820>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.7380(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF63071 GB:AF158600 gp137 [Streptococcus thermophilus
          bacteriophage Sfi11]
 Identities = 67/136 (49%), Positives = 97/136 (71%), Gaps = 2/136 (1%)

Query: 5   PEIDIQKTKSNAKRKLREYPRWRRIANDVDTQKVTATYSFEPRQPHGTPSKPVERLALNR 64
           PEID + T    KRKLREYPRWR IA+D   QK+T  ++F PR   G  +KPVE +A+ R
Sbjct: 4   PEIDEKATLKRCKRKLREYPRWREIAHDSAEQKITQEFTFMPRG--GGVNKPVENIAVRR 61

Query: 65  VSAEQELDTIERAVNGIFDPEYRLILIDKYLLTYPKTDCDIYTKLGYEKSQYYNMLDNAL 124
           V A  EL+ IE+AVNG++ P+YR ILI+KYL   PK +  I   +G+E++ +  +L+N++
Sbjct: 62  VDALNELEAIEQAVNGLYRPDYRRILIEKYLAYPPKPNWQIAQSIGFERTAFQELLNNSI 121

Query: 125 LSFSELYKEGMLLVEK 140
           L+F+ELY++G L+VE+
Sbjct: 122 LAFAELYRDGRLIVER 137
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2661

A DNA sequence (GASx962) was identified in *S.pyogenes* <SEQ ID 7821> which encodes the amino acid sequence <SEQ ID 7822>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3375(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2662

A DNA sequence (GASx963R) was identified in *S.pyogenes* <SEQ ID 7823> which encodes the amino acid sequence <SEQ ID 7824>. Analysis of this protein sequence reveals the following:

```
Possible site: 48

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2663

A DNA sequence (GASx964) was identified in *S.pyogenes* <SEQ ID 7825> which encodes the amino acid sequence <SEQ ID 7826>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.16   Transmembrane   90 - 106 (  89 - 111)
   INTEGRAL    Likelihood = -5.52   Transmembrane  131 - 147 ( 129 - 150)
   INTEGRAL    Likelihood = -0.43   Transmembrane   53 -  69 (  52 -  69)

----- Final Results -----
               bacterial membrane --- Certainty=0.3463(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2664

A DNA sequence (GASx965) was identified in *S.pyogenes* <SEQ ID 7827> which encodes the amino acid sequence <SEQ ID 7828>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3944(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA66779 GB:X98106 Rorf172 [Bacteriophage phig1e]
 Identities = 36/82 (43%), Positives = 52/82 (62%), Gaps = 3/82 (3%)

Query: 18 ELTEKQQRFVDKYITTFNATESAKQAGYSEKSAYSQGQRLLKNVEIQKAMKERFLEAKDT 77
          +LT KQQ+F D+YI + NA ++A++AGYS++SA S GQ  L   +I++ + ER     +
Sbjct: 4  KLTPKQQKFADEYIKSGNAADAARKAGYSKRSARSVGQENLTKPDIKQYIDERM---DEI 60

Query: 78 KGDRIQDVAETLEQDTSIARGE 99
             RI D  E +E  T IARGE
Sbjct: 61 ASKRIMDATEAVELLTRIARGE 82
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2665

A DNA sequence (GASx966) was identified in *S.pyogenes* <SEQ ID 7829> which encodes the amino acid sequence <SEQ ID 7830>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2389(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13115 GB:Z99110 PBSX defective prophage terminase (large
          subunit) [Bacillus subtilis]
 Identities = 117/417 (28%), Positives = 195/417 (46%), Gaps = 33/417 (7%)

Query: 31  YRVVKGSRGSKKSKTTALNFIVRLLKYPWANLLVIRRYSNTNKQSTYTDFKWACNQLKVT 90
           Y+ + G  GS KS  TAL  +++LLK      LVIR   +T++ ST+  F+    +L +T
Sbjct: 21  YQFLVGGYGSSKSYHTALKIVLKLLKEK-RTALVIREVFDTHRDSTFALFQEVIEELGLT 79

Query: 91  HLFKFNESLPEITVKATGQKILFRGLDDELKITSITVDVGALCWAWFEEAYQIETEDKFS 150
                S  ++     G +I+F+G+D+  K+ S    V  +   W EE  +++ E
Sbjct: 80  KAVASLSSPLQLRFH-NGSRIMFKGMDNPAKLKS----VHNISLIWIEECSEVKYEG--- 131

Query: 151 TVVESIRGSLDAPDFFKQITVTFNPWSERHWLKRVFFDEETKR---------------- 193
              + + G L  P+    +  T NP    +W  R FF +E K+
Sbjct: 132 --FKELIGRLRHPELKLHMICTTNPVGTSNWTYRHFFRDERKKRFVLDDSELYEKRTIVK 189

Query: 194 ADTFSGTTTFRVNEWLDDVDKRRYEDLYKTNPRRARIVCDGEWGVAEGLVFDNFEVVDFD 253
            DT+   +T   N +L +   ++ + L + +P   RI   G +GV    V   FEV+  D
Sbjct: 190 GDTYYHHSTANDNLFLPESYVKQLDGLKEYDPDLYRIARKGRFGVNGIRVLPQFEVLPHD 249

Query: 254 -VEKTIQRVKET--SAGMDFGFTQDPTTLICVAVDLANKELWLYNEHYQKAMLTDHIVKM 310
            V+K I  + +     GMDFGF +    ++ +AVD   K L++Y E+YQ  M  D   +
Sbjct: 250 QVKKCIAAISKPIFRTGMDFGFEESYNAVVRLAVDPEKKYLYIYWEYYQNKMTDDRTAEE 309

Query: 311 IRDKNLHRSYIAGDSAEKRLIAEIKSKGVSGIVPSIKGKGSIMQGIQFMQGF-KIYIHPS 369
           +R+    + I  DSAE + I   + +G   +V + K  GS +Q  + ++ F KI+
Sbjct: 310 LREFIETQELIKADSAEPKSIQYFRQQGFR-MVGARKFPGSRLQYTKKVKRFKKIFCSDR 368

Query: 370 CEHTIEEFNTYTFKQDKEGNWLNEPIDKNNHVIDAIRYALEKYHIRSNESNQFEVLR 426
           CE+ I E  T T+ +DK G  + +    + H + AI YAL+ Y +   +    + +R
Sbjct: 369 CENVIYELETLTYAKDKNGALIEDEFTIDPHTLSAIWYALDDYEVADMKETAHKRMR 425
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2666

A DNA sequence (GASx967) was identified in *S.pyogenes* <SEQ ID 7831> which encodes the amino acid sequence <SEQ ID 7832>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4899(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC34397 GB:AF158600 gp502 [Streptococcus thermophilus
```

```
           bacteriophage Sfi11]
 Identities = 67/114 (58%), Positives = 83/114 (72%)

Query: 6   FRDSTGKTKTLEFRFHREARMRYQAESLESLLTEKYKLLREMIEHHDKVQKPRIQELLDY 65
           F DSTG+   L  RFHRE+R+RY+A++LE L+   ++LL+  I HH   Q PRIQELLDY
Sbjct: 7   FTDSTGQDLVLNLRFHRESRIRYRADNLEELMVNNWELLKNFINHHKLRQAPRIQELLDY 66

Query: 66  AEGNNHTISEIGRRKDDDMADVRAVHNYGKYISTLKQGYLVGNPIRVEYIDGTE 119
           A G NH + + GRRKD++MAD RAVHNYG+ IS  K GYL GNPIRVEY D  +
Sbjct: 67  ARGENHDVLKSGRRKDNEMADKRAVHNYGRMISKFKTGYLAGNPIRVEYDDNED 120
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2667

A DNA sequence (GASx968) was identified in *S.pyogenes* <SEQ ID 7833> which encodes the amino acid sequence <SEQ ID 7834>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4007(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC34397 GB:AF158600 gp502 [Streptococcus thermophilus
           bacteriophage Sfi11]
 Identities = 172/319 (53%), Positives = 227/319 (70%), Gaps = 9/319 (2%)

Query: 1   LIYRSMDDKTEVVRLDPREVFVIYQNNLEQSSLAGVRYYNKNQLDGTTKIVELYTDNKIL 60
           +IYRS  D+T + RL P E FVIY N+LE +S+A VRYYN+  L      +VE+YT+  I
Sbjct: 157 VIYRSEYDETRIKRLSPLETFVIYDNSLEDNSIAAVRYYNRGTLQNAKDVVEIYTNQHIY 216

Query: 61  KFEYDGDLTPIGETSSHAFGSVPITEYLNTDDGMGDYETELSLIDLYDAAQSDTANYMQD 120
             +       I  T  HAFG+VPITE+LN  DG+GDYETEL LIDLYD+A+SDTAN+M D
Sbjct: 217 TLDASDSFNEISVTP-HAFGTVPITEFLNNADGIGDYETELYLIDLYDSAESDTANHMSD 275

Query: 121 LSDAILAIIGRVSFPGYVDTAEKAIEYLRKMRKARLLNLEPPVDQDGREGSVDAKYLYKQ 180
           ++DAILAI G ++ P  +  ++         M++ RL+ L+PP   DG+EG+V A+YL K
Sbjct: 276 MADAILAIYGDLALPQGMQASD--------MKRTRLMQLKPPKSADGKEGTVKAEYLTKS 327

Query: 181 YDVQGTEAYKNRIVSDIHKFTNTPDMTDSKFAGQQSGEALKWKVFGLDQERVDMQALFEQ 240
           YDV G EAYK R+  DIH FTNTPDM+D+ F+G  SGEALK+K+FGLDQ+RVD Q+ F Q
Sbjct: 328 YDVSGAEAYKTRLNKDIHVFTNTPDMSDNHFSGNASGEALKYKLFGLDQDRVDTQSQFTQ 387

Query: 241 SLKRRYKLIARVSQLLKEIDDFDISKLKITFTPNLPKSLQEKIEAFKALGGELSQETAMA 300
            LKRRY+L AR+  L+ E  DFD S+LKITFTPNLPKSL E++      LGG++SQETA++
Sbjct: 388 GLKRRYRLAARIGSLVNEFKDFDESRLKITFTPNLPKSLYEQVSILNDLGGQVSQETALS 447

Query: 301 ITDIVEDAKKEISLINSES 319
           ++ +VE+  +E+  IN ES
Sbjct: 448 LSGLVENPTEELDKINEES 466
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2668

A DNA sequence (GASx969) was identified in *S.pyogenes* <SEQ ID 7835> which encodes the amino acid sequence <SEQ ID 7836>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5307(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC79543 GB:U88974 ORF28 [Streptococcus thermophilus temperate
          bacteriophage O1205]
 Identities = 118/309 (38%), Positives = 183/309 (59%), Gaps = 18/309 (5%)

Query: 8   YWRDRIKKEMDAK-EADDISLEQSMKQLHDYHFRNIEKEIESFYQRYADKEKIDLSEARK 66
           YW  R  +E +A  +  +   ++ ++ L++     + KE++++ Q+YA+K  + +S+A++
Sbjct: 3   YWSKRTLREREASIKKGEAEFKKELEALYNLQLSQLRKELDAYIQKYANKNGLSVSDAKR 62

Query: 67  RASELDISAYQKKAKELVAKAEKLRREGKIVTRDDFTHQENADMSIYNLAMKTNALELLR 126
           +A   D+ A++ KAK  VA               DF+ + N ++  YN +M     ELL
Sbjct: 63  KADSFDVKAFETKAKRYVADK-------------DFSPKANRELQDYNFSMSVGRQELLI 109

Query: 127 LNIDLEMQELANGEHKLTKKFLDEGYRKETEFQAGLLGLSVASQASVKSLADAVINANFK 186
             ++LE+  L+  E +LT  +L  GY+ E   +  LL  +V S  +++    A +NANF+
Sbjct: 110 QELELELLALSESERQLTNDYLTNGYKSEV-VRESLLDQTVPSGKTLEKYMKAAVNANFE 168

Query: 187 GAKWSDNIWDRQDKLRSIISQSVQSAILKGKNGLTIARDIRREFDVSASYAKRLAITEHA 246
           GA+WS+ IW RQ++LR I+   V  A+++G+NGLTIAR IR+  D S + A+RLAITEHA
Sbjct: 169 GAEWSERIWKRQEQLRKIVKTEVTRALIRGENGLTIARRIRKHMDASRTEAERLAITEHA 228

Query: 247 RVQMEVGRLSMAENGFAMFDILPEPKACDVCKDIAKH---GPYHLDKWRIGENSPPFHPY 303
           RVQ       M ENGF  F ++PE +ACD+CKDI K     P  +    IG N+PP HPY
Sbjct: 229 RVQTLAQESIMKENGFEHFKLMPESRACDICKDIGKETEKNPVKIADMEIGTNAPPIHPY 288

Query: 304 CRCAIVGVD 312
           CRCA+V V+
Sbjct: 289 CRCAVVEVE 297
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2669

A DNA sequence (GASx970) was identified in *S.pyogenes* <SEQ ID 7837> which encodes the amino acid sequence <SEQ ID 7838>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2091(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2670

A DNA sequence (GASx971) was identified in *S.pyogenes* <SEQ ID 7839> which encodes the amino acid sequence <SEQ ID 7840>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2671

A DNA sequence (GASx972) was identified in *S.pyogenes* <SEQ ID 7841> which encodes the amino acid sequence <SEQ ID 7842>. Analysis of this protein sequence reveals the following:

```
Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3226(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2672

A DNA sequence (GASx973) was identified in *S.pyogenes* <SEQ ID 7843> which encodes the amino acid sequence <SEQ ID 7844>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1830(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2673

A DNA sequence (GASx975) was identified in *S.pyogenes* <SEQ ID 7845> which encodes the amino acid sequence <SEQ ID 7846>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4757(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB07248 GB:AP001519 unknown [Bacillus halodurans]
 Identities = 46/134 (34%), Positives = 73/134 (54%)

Query: 23  KQPQDEKKYTDADVDAIIDKKFAKWKSEQEAEKSEAKKMAKMNEKEKADYEKQKLLDELQ 82
           K  + E+ +T  +V+ I+  + A+   ++E    EA+K+AKMN ++K +YE +KL  E +
Sbjct: 66  KPNKTERLFTQEEVNRIVKDRLARALKDKEEAIKEAEKLAKMNAEQKREYELEKLRRENE 125

Query: 83  ELKNDKTRNELTAVARQMFAESEINVNDDVLGLVVTLDAEQTKANVTTLANAFAKVIADD 142
           +LK  + R EL   A +M  E+ I  +DDVL  VV  DAEQT+  V T  +   K+
Sbjct: 126 QLKKAQMRYELGREATKMLGEAGIMADDDVLSFVVRDDAEQTQEAVKTFISLVDKLADMR 185

Query: 143 RKALVRQTTPSTGG 156
            K  ++   P  G
Sbjct: 186 MKEKLKGRPPKKDG 199
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2674

A DNA sequence (GASx976) was identified in *S.pyogenes* <SEQ ID 7847> which encodes the amino acid sequence <SEQ ID 7848>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2478(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC79545 GB:U88974 ORF30 [Streptococcus thermophilus temperate
          bacteriophage O1205]
```

```
 Identities = 43/119 (36%), Positives = 66/119 (55%), Gaps = 16/119 (13%)

Query: 9   SKEILHNLDYEAISVTLDSNKIG-----KKVVPAGTILAGKDKSIFEDRKQKVETVTNEE 63
           +  I+ +L Y+A+S T+DS+  G      KK + AGT++AG   SIF+DR + V
Sbjct: 9   TSNIVRSLPYKAVSATVDSSYPGVLVDGKKYIKAGTLVAGNGGSIFDDRTKSV------- 61

Query: 64  VSTKEYVDGILLTDVDLTNGDAVGSCVYRGTINADKLADSSVAENYDDLEEVLPHIVFI 122
           V  K   +GI+L DVDLT  + V S +Y G +  DK+     +    D +++ LP + FI
Sbjct: 62  VENKTEPEGIVLYDVDLTIDNTV-SVLYAGEVYKDKVNGGDIT---DTVKKALPLVKFI 116
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2675

A DNA sequence (GASx978) was identified in *S.pyogenes* <SEQ ID 7849> which encodes the amino acid sequence <SEQ ID 7850>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4238(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC79546 GB:U88974 ORF31 [Streptococcus thermophilus temperate
          bacteriophage O1205]
 Identities = 195/343 (56%), Positives = 256/343 (73%), Gaps = 1/343 (0%)

Query: 1   MALIHEIITSENIKGFYNAKNENVENTLGEKAFPPKQQLGLKLSFIKGAAGKPVTLKAAA 60
           M LI++ +T+ NI G++NA  ENV +TLGE  FP ++QLG KLS+IKGA+G+ V LKAAA
Sbjct: 1   MGLIYDKVTASNIAGYFNALQENVSSTLGESIFPARKQLGTKLSYIKGASGQSVALKAAA 60

Query: 61  FDTKVPLRDRMAVELIDEEMPFFKEAMLVKEADRQQLNMLAQTKNNELIDTILASIYNDQ 120
           FDT V +RDR++ E+ DE+MPFFKEAMLVKE DRQQLN++  + N  L++TI+A I+ND
Sbjct: 61  FDTNVTIRDRVSAEMHDEQMPFFKEAMLVKENDRQQLNLVKDSGNAVLVNTIVAGIFNDN 120

Query: 121 ATLIAGAKARLEAMRMEVLSKGKIHIQSNGVMKDIDYGLAEDQTTKPDAKWDSAGTATPL 180
            TL+ GA+ARLEAMRM+VL+ GKI   S+GV KDIDYG+  D  +    W  G ATPL
Sbjct: 121 LTLVNGARARLEAMRMQVLATGKIAFTSDGVNKDIDYGVKPDHKKQVSKSWAEPG-ATPL 179

Query: 181 KDIEKAIEKMAERGFVPEAIIMNSKTFSLIKNAESTLDVVKPMAPNGAAVTKRDLNTYLE 240
            D+E AIE   E G  PE  +MN+KTF LI+ A ST+ V+KP+A +G+AVTK +L  Y+
Sbjct: 180 ADLEDAIETARELGLNPERAVMNAKTFGLIRKAASTVKVIKPLAGDGSAVTKAELENYIA 239

Query: 241 DELQIKVILKDGMFVGDDGESRKYFPDGFATLVPNGNLGYTVFGTTPEQSDLLGGEATDA 300
           D   + ++L++G +  D GE  K++PDG  TL+PNG LG TVFGTTPE+SDL      +A
Sbjct: 240 DNFGVSIVLENGTYRNDKGEVSKFYPDGHLTLIPNGPLGNTVFGTTPEESDLFADNTVNA 299

Query: 301 NVSIVETGIAITTTKTTDPVNVQTKVSMIALPSFERLEEVHII 343
            V IV+ GIA+TTTKTTDPVNVQTKVSM+ALPSFERL++V+++
Sbjct: 300 EVEIVDNGIAVTTTKTTDPVNVQTKVSMVALPSFERLDDVYML 342
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2676

A DNA sequence (GASx979) was identified in *S.pyogenes* <SEQ ID 7851> which encodes the amino acid sequence <SEQ ID 7852>. Analysis of this protein sequence reveals the following:

```
Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3319(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2677

A DNA sequence (GASx980) was identified in *S.pyogenes* <SEQ ID 7853> which encodes the amino acid sequence <SEQ ID 7854>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2385(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC34404 GB:AF158600 gp113 [Streptococcus thermophilus
          bacteriophage Sfi11]
 Identities = 53/109 (48%), Positives = 79/109 (71%), Gaps = 4/109 (3%)

Query: 11  IVKNVKLDLGIEDDNQDQLLEMLLNRITDHFKANYGVLEIDNAFSFVLEDCLIARFNRRG 70
           +++NV +DL I DDN   LL +LL RI +HFKA YGV E+D+  +F+ EDCL+ RFNRRG
Sbjct: 9   VIQNVSVDLNINDDN---LLGILLERIVNHFKAEYGVDEVDDNLAFIFEDCLVKRFNRRG 65

Query: 71  SERAKTEEVEGHKTTYYDHLNEFEPYDAMIMAKLNLIKDKSRKGGLYFL 119
           +E A++E ++GH  +YYD+ NEF+PYD M+  +L      ++++G + FL
Sbjct: 66  AEGARSESIDGHSMSYYDNENEFDPYDNMLQ-RLYGTSGQAKEGEVLFL 113
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2678

A DNA sequence (GASx981) was identified in *S.pyogenes* <SEQ ID 7855> which encodes the amino acid sequence <SEQ ID 7856>. Analysis of this protein sequence reveals the following:

```
Possible site: 49

>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5714(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA59188 GB:X84706 b3 [Bacteriophage B1]
 Identities = 28/82 (34%), Positives = 49/82 (59%), Gaps = 2/82 (2%)

Query: 1  MRYADRVTFVKTT-DEQYNPDLGEYTHTEVISITKPCFVMDMGMEKSVQIFGDYQKDRKV 59
          +RY D VTF+K + D  Y+PDLGE+   E          + D+G ++SV++FGD +K  KV
Sbjct: 1  LRYLDEVTFIKESPDSHYDPDLGEWVEKEPTRTVFSANITDIGTDRSVEVFGDIKKGAKV 60

Query: 60 IYLKQPYT-KAFDYCEYEGRRY 80
          + +   +    +DY E++ +++
Sbjct: 61 MRMMPLFNMPKYDYIEFDNKKW 82
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2679

A DNA sequence (GASx982) was identified in *S.pyogenes* <SEQ ID 7857> which encodes the amino acid sequence <SEQ ID 7858>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2509(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC34406 GB:AF158600 gp114 [Streptococcus thermophilus
          bacteriophage Sfi11]
 Identities = 44/103 (42%), Positives = 65/103 (62%), Gaps = 5/103 (4%)

Query: 17  GLKKKLELIIKKDAVKK---IVRDNGTQLQRKMINKAVFTKGYSTGATRRSITMQIGDGG 73
          GL +  + ++K  + +K   ++R  G++L+   +N+A F KGYSTGATRRSIT+Q+
Sbjct: 8   GLDEMAQSLLKNASPEKRSKVLRKYGSKLKEAAVNRAQFNKGYSTGATRRSITLQVESDK 67

Query: 74  LSVKVKPGTHYAGYLERGTRLMSKQPFVLPALKEQKVKFRKDL 116
           +V+    T Y+GYLE GTR M  QPF+ PAL E   K  ++L
Sbjct: 68  ATVEAL--TSYSGYLEVGTRKMEAQPFMKPALDEVAPKMVEEL 108
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2680

A DNA sequence (GASx983) was identified in *S.pyogenes* <SEQ ID 7859> which encodes the amino acid sequence <SEQ ID 7860>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3098(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA32612 GB:L31366 putative [Bacteriophage Tuc2009]
 Identities = 88/129 (68%), Positives = 108/129 (83%)

Query: 1   MIKTRDQSIFDEMFKRIQSLGFKVYDYKPMTEVPYPFVEMESTDAEYIPNKDDIKGSVEL 60
           MIKTRDQSIFDE+FKRIQ+LG+ VYDYKPM EV YPFVE+E+T   +  NK DIKG+V L
Sbjct: 1   MIKTRDQSIFDELFKRIQALGYTVYDYKPMNEVGYPFVELENTQTIHEANKTDIKGTVSL 60

Query: 61  MLSVWGVQKKRKQVSDMASAIFSQALTVESSDVFRWSLNTRQSSIQMLDDTTTVTPLKRA 120
            LSVWG+QKKRK+VSDMAS IF+QAL + ++D + W+LN++ S+IQMLDDTTT TPLKRA
Sbjct: 61  SLSVWGLQKKRKEVSDMASNIFNQALNISATDGYSWALNSQASTIQMLDDTTTHTPLKRA 120

Query: 121 IVTLRFNLR 129
           ++ L F LR
Sbjct: 121 LINLEFRLR 129
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2681

A DNA sequence (GASx984R) was identified in *S.pyogenes* <SEQ ID 7861> which encodes the amino acid sequence <SEQ ID 7862>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1736(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2682

A DNA sequence (GASx985) was identified in *S.pyogenes* <SEQ ID 7863> which encodes the amino acid sequence <SEQ ID 7864>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA32613 GB:L31366 structural protein [Bacteriophage Tuc2009]
 Identities = 81/185 (43%), Positives = 111/185 (59%), Gaps = 22/185 (11%)

Query: 4   QLEAKQGIHSILLFRLLKEASSEAATKLAFQTEHEVGKSRDVDGQKTKDGIIQSVGALEY 63
           +L AKQG   ILL+RLL +A+ EAA KLAFQTEH   K+RD +   TKDG I S+ A+EY
Sbjct: 3   ELTAKQGKDIILLYRLLSKATKEAAWKLAFQTEHSNEKTRDYNTTATKDGTIGSLAAIEY 62

Query: 64  DFKATSILAKGDVLAAKLEKAMENGELVEIWDIDLEETSKNGDSDNKLANVWGIDKNGTN 123
              ATSI A GD    +++KA ++GE++++W+ID  E
Sbjct: 63  SLSATSIAANGDPHLDEMDKAFDDGEIIDVWEIDKAEKG---------------------- 101

Query: 124 RGNGKYLATYYQGYISSFSAKKNAEENIEIEMEFAINGVGQKGFATLTDAQKAAVQYAFK 183
             +GKY A Y + Y++SFS + N+E+ +E+ +EF + G  QKG ATLT+ Q   VQY FK
Sbjct: 102 -SDGKYKAKYLRAYLTSFSYEPNSEDALELSLEFGVFGKPQKGQATLTEEQANVVQYVFK 160

Query: 184 DTTKG 188
           DT  G
Sbjct: 161 DTVAG 165
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2683

A DNA sequence (GASx986) was identified in *S.pyogenes* <SEQ ID 7865> which encodes the amino acid sequence <SEQ ID 7866>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2273(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA59192 GB:X84706 a2 [Bacteriophage B1]
 Identities = 54/111 (48%), Positives = 72/111 (64%), Gaps = 1/111 (0%)

Query: 1   MQLEIKGKTHNVKFGTRFVAEMDKNHIAERQGFKFGAGLQSSV-PFLIDHSVVTLAEVIY 59
           M+L IKGK  + KFG +FV E+DKN + E+ G  FG  L   + P L   ++ TL+ V++
Sbjct: 1   MELTIKGKQVHFKFGVKFVRELDKNLVIEQNGVSFGLALAVKIIPELEMANIATLSNVLF 60

Query: 60  TGTITEPPRPSLNDIYDYIDEVEDIEKLFDDVLDELRQSNASKLFMAQVEK 110
            G  TE P+ S  DI D+IDE EDIEKLFDDVL E+ +SN  KL  A++ K
Sbjct: 61  LGNRTETPKLSQGDIDDFIDECEDIEKLFDDVLKEITESNTGKLIKAKMTK 111
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2684

A DNA sequence (GASx987) was identified in *S.pyogenes* <SEQ ID 7867> which encodes the amino acid sequence <SEQ ID 7868>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2735(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA59193 GB:X84706 c2 [Bacteriophage B1]
 Identities = 40/111 (36%), Positives = 57/111 (51%), Gaps = 10/111 (9%)

Query: 2   IVLNCIRYLGMTDINEIGRLTLYEYDLLMTGKALAAVDESHKAHKQAWINHQVTATKLVG 61
           +++  +R  G+ D++   R+T+ EY +      L  +DE    ++QAW N QV ATK  G
Sbjct: 15  MMIRFLRCFGIQDLSVFERMTIREYSIRSIAFQLRTLDEEEFIYEQAWANWQVQATKQQG 74

Query: 62  GKKNKKEVPVYKKFKDFFD---YEEEIRKI-TQEIDEGYDKKGMDLLLKAN 108
            K      P+Y  FK FFD    E EI  I + E     D K +DL+ KAN
Sbjct: 75  KK------PLYPTFKKFFDKKKLENEILGIESPENKFKKDNKLIDLMKKAN 119
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2685

A DNA sequence (GASx989) was identified in *S.pyogenes* <SEQ ID 7869> which encodes the amino acid sequence <SEQ ID 7870>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2869(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA66560 GB:X97918 gene 19.1 [Bacteriophage SPP1]
 Identities = 66/232 (28%), Positives = 106/232 (45%), Gaps = 12/232 (5%)

Query: 38  FRTLTVSGRDVVDLEHQTTSVLGRNGEYFHNATVEVRKLEIKAKISGKDNKS-MRLQYEK 96
           F    V GR V  +E    ++ G +G      ++  R+LE+ A + G   ++ +R + E
Sbjct: 24  FLVQEVRGRSVYSIEMGKRTIAGVDGGVITTESLPARELEVDAIVFGDGTETDLRRRIEY 83

Query: 97  LNKLIVSHNQVFLSFSDEPDRNYLGIFKSKDVPEEVSNEQIIGLTFICYNPFK-----MS 151
           LN L+     V ++FSDEP R Y G ++     +E      + L F C +P K      +
Sbjct: 84  LNFLLHRDTDVPITFSDEPSRTYYGRYEFATEGDEKGGFHKVTLNFYCQDPLKYGPEVTT 143

Query: 152 DVKTKKGTSIQNGGLFQTKPIITLNLSSPTKEIKLLHVESQKYIRLT----GTYTTDEIK 207
           DV T   T ++N GL  T P I    S+   E ++  ++      G  T D +
Sbjct: 144 DV-TTASTPVKNTGLAVTNPTIRCVFSTSATEYEMQLLDGSTVVKFLKVKYGFNTGDTLV 202

Query: 208 IDMATGKITQNGRNILGDLDMINSRYFELLPGNNTLQCANAAITAEFREVYL 259
           ID     +T NG++I+  L +I S + +L P  NT   A    T  F E +L
Sbjct: 203 IDCHERSVTLNGQDIMPALLLIQSDWIQLKPQVNTYLKATQPSTIVFTEKFL 253
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2686

A DNA sequence (GASx990) was identified in *S.pyogenes* <SEQ ID 7871> which encodes the amino acid sequence <SEQ ID 7872>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2861(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04681 GB:AP001510 unknown conserved protein in others
          [Bacillus halodurans]
 Identities = 116/449 (25%), Positives = 198/449 (43%), Gaps = 79/449 (17%)

Query: 2   IYLFDKLERLVATVG-TDDLLSWHFKVKNNDWDQASFEVPVDYDVEPFVYFGFFNYDPHQ 60
           +++FD+ ++L+ T+  +  L+   F+ + N      F   ++   E   +      + HQ
Sbjct: 4   LFIFDREDQLLTTLTESTGLVRALFREELNRVPNQPFAFTIEASSEEAKHV----IEEHQ 59

Query: 61  -----KEDVFKLFKVIDYNLEDSKFYKG------LDKAESDLDTIAIIKDKRFRQSSADA 109
                KE   +LF + +  LED     G       + A  +L    I++     Q + +A
Sbjct: 60  VVFRDKEGDLRLFVIKE--LEDVDGLDGPQTTAICEPAFMELAEHMIVEQSVVNQPAHEA 117

Query: 110 CIDGALEGTGYQVGKVEGITNVRTLSYYYISPRAALIKIVEAFNCEFNVRYTF-INNKIT 168
            ++ AL+GT +  G VE      T  + Y+S   A+  I+  +  +F    TF   N+IT
Sbjct: 118 -LNVALQGTRW-TGSVEVNLGNATEHFSYVSAIEAVWNILVTWGGDFKDVVTFNAENRIT 175

Query: 169 SRYIDLKKRFGKPTGKQFEHGNNLLKVVYEESTDDIVTCLIGRGKGEEIQHEEAEPKDVE 228
           S  I + +R G   GK+FE  +N+ + +        VT L GRG    +Q E  E    +
Sbjct: 176 SHQIKIVQRRGVDRGKRFEIDHNI-EQIERTILSYPVTALYGRGAS--LQGENGE----D 228

Query: 229 GHLPQEERRQGYGRRIEFTDVVWSVEKGDPIDKPAGQNFVALDSAREEYGLSQNGELKHR 288
           G L              +F +V W    G P+DKP GQ +V    A ++YG   NG+L HR
Sbjct: 229 GSL-------------DFGEVEWRKSAGAPVDKPKGQLWVGDPEALQKYGRKHNGQLLHR 275

Query: 289 WGVFVNEEIEDKTELLKATWEELQRLSIPIRIYKAEILDIGPETWKGDSVAIIYDEVKIA 348
            G+F N  IED  ELL+ TWE+LQ+ S P   Y+  +                +++ +
Sbjct: 276 EGIFQNTNIEDPEELLEKTWEQLQKSSKPEVHYRLSVR--------------LFEHIS-- 319

Query: 349 FETRVDEIDIDKLNFNRSVVTLGDYSVVQNR------ESRSRKEAVQ-NMIDESLETITD 401
                      +     +LGD ++  +R       E +SR  A++ +++D     + +
Sbjct: 320 ------------GYEHEQASLGDTAIAIDRQFSRPIEIQSRIIAIEYDLVDIDGTGMVE 366

Query: 402 LGMTFQEFLQGIEKRIETGKKEMEDNWRK 430
           +G      L G+++R+E   +E+E N  K
Sbjct: 367 MGQFLS--LNGMDERLERIIEEIEKNQGK 393
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2687

A DNA sequence (GASx991) was identified in *S.pyogenes* <SEQ ID 7873> which encodes the amino acid sequence <SEQ ID 7874>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2584(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA98101 GB:M19348 hyaluronidase [Streptococcus pyogenes phage
          H4489A]
 Identities = 314/371 (84%), Positives = 338/371 (90%), Gaps = 1/371 (0%)

Query: 1   MAENIPLRVQFKRMKAAEWASSDVVLLEGEIGFETDTGFAKFGDGQNTFSKLKYLTGPKG 60
           M ENIPLRVQFKRM A EWA SDV+LLEGEIGFETDTGFAKFGDGQNTFSKLKYLTGPKG
Sbjct: 1   MTENIPLRVQFKRMSADEWARSDVILLEGEIGFETDTGFAKFGDGQNTFSKLKYLTGPKG 60

Query: 61  PKGDTGLQGKTGGTGSRGPAGKPGTTDYDQLQNKPDLGAFAQKEETNSKITKLESSKADK 120
           PKGDTGLQGKTGGTG RGPAGKPGTTDYDQLQNKPDLGAFAQKEETNSKITKLESSKADK
Sbjct: 61  PKGDTGLQGKTGGTGPRGPAGKPGTTDYDQLQNKPDLGAFAQKEETNSKITKLESSKADK 120

Query: 121 NAVYLKAESNAKLDEKLNLKGGVMTGQLQFKPN-SGIKPSSSVGGAINIDMSKSEGAAMV 179
           +AVY KAES  +LD+KL+L GG++TGQLQFKPN SGIKPSSSVGGAINIDMSKSEGAAMV
Sbjct: 121 SAVYSKAESKIELDKKLSLTGGIVTGQLQFKPNKSGIKPSSSVGGAINIDMSKSEGAAMV 180

Query: 180 MYTNKDTTDGPLMILRSNKDTFDQSVQFVDYKGTTNAVNIVMRQPTTPNFSSALNITSAN 239
           MYTNKDTTDGPLMILRS+KDTFDQS QFVDY G TNAVNIVMRQP+ PNFSSALNITSAN
Sbjct: 181 MYTNKDTTDGPLMILRSDKDTFDQSAQFVDYSGKTNAVNIVMRQPSAPNFSSALNITSAN 240

Query: 240 EGGSAMQIRGVEKALGTLKITHENPSVDKEYDKNAAALSIDIVKKQKGGKGTAAQGIYIN 299
           EGGSAMQIRGVEKALGTLKITHENP+V+ +YD+NAAALSIDIVKKQKGGKGTAAQGIYIN
Sbjct: 241 EGGSAMQIRGVEKALGTLKITHENPNVEAKYDENAAALSIDIVKKQKGGKGTAAQGIYIN 300

Query: 300 STSGTTGKLLRIRNLNDDKFYVKPDGGFYAKETSQIDGNLKLKDPIANDHAATKAYVDGE 359
           STSGT GK+LRIRN N+DKFYV PDGGF++    S + GNL +KDP +  HAATK YVD +
Sbjct: 301 STSGTAGKMLRIRNKNEDKFYVGPDGGFHSGANSTVAGNLTVKDPTSGKHAATKDYVDEK 360

Query: 360 VEKLKALLAAK 370
           + +LK L+  K
Sbjct: 361 IAELKKLILKK 371
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2688

A DNA sequence (GASx993) was identified in *S.pyogenes* <SEQ ID 7875> which encodes the amino acid sequence <SEQ ID 7876>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1358(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2689

A DNA sequence (GASx995) was identified in *S.pyogenes* <SEQ ID 7877> which encodes the amino acid sequence <SEQ ID 7878>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0855(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC34418 GB:AF158600 gp149 [Streptococcus thermophilus
          bacteriophage Sfi11]
 Identities = 27/95 (28%), Positives = 50/95 (52%), Gaps = 2/95 (2%)

Query: 9   KYPQLDGTGAVASTHIIIAAEDGAVIPQLIKQDLTSTNDTEIIKAALEEFKKSEYVEIAM 68
           K  + D +GA  +T +I+   DGA +P  +  +    ++TE++K ALE   +  + + A
Sbjct: 26  KSKEYDASGAAYATKVILKNRDGAYVPVFLPVEKIDLSNTELLKEALEVIYQENFPQRAE 85

Query: 69  GEAVQKVDDLEKISQETAKTAKTAQTAAGLAKVSA 103
            E   ++D  EKI +  A + K  +T A + + S+
Sbjct: 86  NEKFNELD--EKIKEYEALSKKATETIAKMEEASS 118
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2690

A DNA sequence (GASx996) was identified in *S.pyogenes* <SEQ ID 7879> which encodes the amino acid sequence <SEQ ID 7880>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -4.62   Transmembrane    9 -  25 (   7 -  26)

----- Final Results -----
             bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2691

A DNA sequence (GASx997) was identified in *S.pyogenes* <SEQ ID 7881> which encodes the amino acid sequence <SEQ ID 7882>. Analysis of this protein sequence reveals the following:

```
Possible site: 41

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.66   Transmembrane   38 -  54 (  35 -  55)

----- Final Results -----
               bacterial membrane --- Certainty=0.2466(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2692

A DNA sequence (GASx998R) was identified in *S.pyogenes* <SEQ ID 7883> which encodes the amino acid sequence <SEQ ID 7884>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.87   Transmembrane   47 -  63 (  41 -  72)

----- Final Results -----
               bacterial membrane --- Certainty=0.4949(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2693

A DNA sequence (GASx999) was identified in *S.pyogenes* <SEQ ID 7885> which encodes the amino acid sequence <SEQ ID 7886>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2694

A DNA sequence (GASx1001) was identified in *S.pyogenes* <SEQ ID 7887> which encodes the amino acid sequence <SEQ ID 7888>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.51   Transmembrane   18 -  34 (  16 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2695

A DNA sequence (GASx1002) was identified in *S.pyogenes* <SEQ ID 7889> which encodes the amino acid sequence <SEQ ID 7890>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.61   Transmembrane   12 -  28 (  11 -  33)

----- Final Results -----
               bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein is similar to AF186180 from *S.equi*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2696

A DNA sequence (GASx1003) was identified in *S.pyogenes* <SEQ ID 7891> which encodes the amino acid sequence <SEQ ID 7892>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein is similar to SeeH from *S.equi*:

```
>GP:AAF72809 GB:AF186180 SeeH [Streptococcus equi]        Length = 236
 Identities = 233/236 (98%), Positives = 234/236 (98%)

Query: 1   MRYNCRYSHIDKKIYSMIICLSFLLYSNVVQANSYNTTNRHNLESLYKHDSNLIEADSIK 60
           MRYNCRYSHIDKKIYSMIICLSFLLYSNVVQANSYNTTNRHNLESLYKHDSNLIEADSIK
Sbjct: 1   MRYNCRYSHIDKKIYSMIICLSFLLYSNVVQANSYNTTNRHNLESLYKHDSNLIEADSIK 60

Query: 61  NSPDIVTSHMLKYSVKDKNLSVFFEKDWISQEFKDKEVDIYALSAQEVCECPGKRYEAFG 120
           NSPDIVTSHMLKYSVKDKNLSVFFEKDWISQEFKDKEVDIYALSAQE CECPGKRYEAFG
Sbjct: 61  NSPDIVTSHMLKYSVKDKNLSVFFEKDWISQEFKDKEVDIYALSAQEACECPGKRYEAFG 120

Query: 121 GITLTNSEKKEIKVPVNVWDKSKQQPPMFITVNKPKVTAQEVDIKVRKLLIKKYDIYNNR 180
           GITLTNSEKKEIKVP+NVWDKSKQ PPMFITVNKPKVTAQEVDIKVRKLLIKKYDIYNNR
Sbjct: 121 GITLTNSEKKEIKVPINVWDKSKQHPPMFITVNKPKVTAQEVDIKVRKLLIKKYDIYNNR 180

Query: 181 EQKYSKGTVTLDLNSGKDIVFDLYYFGNGDFNSMLKIYSNNERIDSTQFHVDVSIS 236
           EQKYSKGTVTLDLNSGKDIVFDLYYFGNGDFNSMLKIYSNNERIDSTQFHVDVSIS
Sbjct: 181 EQKYSKGTVTLDLNSGKDIVFDLYYFGNGDFNSMLKIYSNNERIDSTQFHVDVSIS 236
```

There is also homology to a *S.aureus* enterotoxin:

```
>GP:AAA19777 GB:U11702 enterotoxin H [Staphylococcus aureus]
 Identities = 70/215 (32%), Positives = 108/215 (49%), Gaps = 19/215 (8%)

Query: 27  SNVVQANSYNTTNRHNLESLYKHDSNLIEADSI-KNSPDIVTSHMLKYSVKDKNLSVFFE 85
           +++  AN+Y   N   ++   K D    E D I +N  D      +K++  D
Sbjct: 34  TDLALANAYGQYNHPFIKENIKSDEISGEKDLIFRNQGDSGNDLRVKFATAD-------- 85

Query: 86  KDWISQEFKDKEVDIYALSAQEVCECPGKRYEA--FGGITLTNSEK--KEIKVPVNVWDK 141
              ++Q+FK+K VDIY  S    CE   +      +GG TL NSEK  +E  +  NVW
Sbjct: 86  ---LAQKFKNKNVDIYGASFYYKCEKISENISECLYGGTTL-NSEKLAQERVIGANVWVD 141

Query: 142 SKQQPPMFITVNKPKVTAQEVDIKVRKLLIKKYDIYNNREQKYSKGTVTLDLNSGKDIVF 201
             Q+    I  NK  VT QE+DIK+RK+L  KY IY  ++ + SKG +  D+ + +D  F
Sbjct: 142 GIQKETELIRTNKKNVTLQELDIKIRKILSDKYKIY-YKDSEISKGLIEFDMKTPRDYSF 200

Query: 202 DLYYFGNGDFNSMLKIYSNNERIDSTQF-HVDVSI 235
           D+Y     +   + KIY +N+ + S     H+DV++
Sbjct: 201 DIYDLKGENDYEIDKIYEDNKTLKSDDISHIDVNL 235

>GP:AAC26661 GB:AF064774 extracellular enterotoxin type I precursor
           [Staphylococcus aureus]
 Identities = 68/214 (31%), Positives = 109/214 (50%), Gaps = 27/214 (12%)

Query: 42  NLESLY-KHDSNLIEADSIKNSPDIVTSHMLKYSVKDKNLSVFFEKDWIS-QEFKDKEVD 99
           NL + Y KHD   ++   + KN P    ++ L++S    +L +    +W    +FK K++D
Sbjct: 32  NLRNFYTKHDYIDLKGVTDKNLP---IANQLEFSTGTNDL-ISESNNWDEISKFKGKKLD 87

Query: 100 IYALSAQEVCECPGKRYEAFGGITLTNSEKKEI-KVPVNVWDKSKQQPPMF--ITVNKPK 156
           I+ +     C    K    +GG TL+        K+P+N+W   K +        I  NK
Sbjct: 88  IFGIDYNGPC----KSKYMYGGATLSGQYLNSARKIPINLWVNGKHKTISTDKIATNKKL 143

Query: 157 VTAQEVDIKVRKLLIKKYDIYNNRE----------QKYSKGTVTLDLNSGKDIVFD 202
           VTAQE+D+K+R+ L ++Y+IY +                ++ G V   LN+ K   +D
Sbjct: 144 VTAQEIDVKLRRYLQEEYNIYGHNNTGKGKEYGYKSKFYSGFNNGKVLFHLNNEKSFSYD 203

Query: 203 LYYFGNGDFNSMLKIYSNNERIDSTQFHVDVSIS 236
           L+Y G+G    S LKIY +N+ I+S +FH+DV IS
Sbjct: 204 LFYTGDGLPVSFLKIYEDNKIIESEKFHLDVEIS 237

>GP:AAC28968 GB:U93688 enterotoxin [Staphylococcus aureus]
 Identities = 70/244 (28%), Positives = 127/244 (51%), Gaps = 27/244 (11%)

Query: 12  KKIYSMIICLSFLLYSNVVQANSYNTTNRHNLESLYKHDSNLIEADSIKNSPDIVTSHML 71
```

```
            KK+ S+++ ++ ++      A++       NL + Y    + ++   +K++ D   ++ L
Sbjct: 2    KKLISILL-INIIILGVSNNASAQGDIGIDNLRNFYTK-KDFVDLKDVKDN-DTPIANQL 58

Query: 72   KYSVKDKNLSVFFEKDWIS-QEFKDKEVDIYALSAQEVCECPGKRYEAFGGITLTNSE-K 129
            ++S +  +L +   KD+     FK K++D++ +S    C     +Y  +GG+T TN
Sbjct: 59   QFSNESYDL-ISESKDFNKFSNFKGKKLDVFGISYNGQCNT---KY-IYGGVTATNEYLD 113

Query: 130  KEIKVPVNVW--DKSKQQPPMFITVNKPKVTAQEVDIKVRKLLIKKYDIYNNREQK---- 183
             K   +P+N+W     K      ++ NK  VTAQE+D+K+RK L ++Y+IY +   K
Sbjct: 114  KSRNIPINIWINGNHKTISTNKVSTNKKLVTAQEIDVKLRKYLQEEYNIYGHNGTKKGEE 173

Query: 184  ----------YSKGTVTLDLNSGKDIVFDLYYFG-NGDFNSMLKIYSNNERIDSTQFHVD 232
                    ++ G VT  LN+     +DL+Y G +G   S LKIY +N+ ++S +FH+D
Sbjct: 174  YGHKSKFYSGFNIGKVTFHLNNNDTFSYDLFYTGDDGLPKSFLKIYEDNKTVESEKFHLD 233

Query: 233  VSIS 236
             V IS
Sbjct: 234  VDIS 237
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2697

A DNA sequence (GASx1004R) was identified in *S.pyogenes* <SEQ ID 7893> which encodes the amino acid sequence <SEQ ID 7894>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.18   Transmembrane   12 -  28 (  12 -  28)

----- Final Results -----
               bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2698

A DNA sequence (GASx1009) was identified in *S.pyogenes* <SEQ ID 7895> which encodes the amino acid sequence <SEQ ID 7896>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.6391(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2699

A DNA sequence (GASx1011) was identified in *S.pyogenes* <SEQ ID 7897> which encodes the amino acid sequence <SEQ ID 7898>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4528(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2700

A DNA sequence (GASx1024) was identified in *S.pyogenes* <SEQ ID 7899> which encodes the amino acid sequence <SEQ ID 7900>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2701

A DNA sequence (GASx1033) was identified in *S.pyogenes* <SEQ ID 7901> which encodes the amino acid sequence <SEQ ID 7902>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1652(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2702

A DNA sequence (GASx1039) was identified in *S.pyogenes* <SEQ ID 7903> which encodes the amino acid sequence <SEQ ID 7904>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.06   Transmembrane   15 -  31 (  15 -  31)

----- Final Results -----
               bacterial membrane --- Certainty=0.1426(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2703

A DNA sequence (GASx1058) was identified in *S.pyogenes* <SEQ ID 7905> which encodes the amino acid sequence <SEQ ID 7906>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.5484(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2704

A DNA sequence (GASx1077) was identified in *S.pyogenes* <SEQ ID 7907> which encodes the amino acid sequence <SEQ ID 7908>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4848(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2705

A DNA sequence (GASx1080) was identified in *S.pyogenes* <SEQ ID 7909> which encodes the amino acid sequence <SEQ ID 7910>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.42   Transmembrane  107 - 123 (  93 - 133)
   INTEGRAL    Likelihood =-11.20   Transmembrane   20 -  36 (  14 -  44)
   INTEGRAL    Likelihood = -8.39   Transmembrane  226 - 242 ( 218 - 246)
   INTEGRAL    Likelihood = -5.52   Transmembrane  129 - 145 ( 126 - 148)
   INTEGRAL    Likelihood = -4.46   Transmembrane  160 - 176 ( 159 - 183)
   INTEGRAL    Likelihood = -1.44   Transmembrane   55 -  71 (  55 -  72)

----- Final Results -----
               bacterial membrane --- Certainty=0.5967(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2706

A DNA sequence (GASx1081) was identified in *S.pyogenes* <SEQ ID 7911> which encodes the amino acid sequence <SEQ ID 7912>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-13.00   Transmembrane  103 - 119 (  91 - 129)
   INTEGRAL    Likelihood =-11.46   Transmembrane  208 - 224 ( 203 - 230)
   INTEGRAL    Likelihood = -8.28   Transmembrane   54 -  70 (  46 -  71)
   INTEGRAL    Likelihood = -5.79   Transmembrane  160 - 176 ( 155 - 181)
   INTEGRAL    Likelihood = -4.25   Transmembrane  127 - 143 ( 125 - 149)

----- Final Results -----
               bacterial membrane --- Certainty=0.6201(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2707

A DNA sequence (GASx1089) was identified in *S.pyogenes* <SEQ ID 7913> which encodes the amino acid sequence <SEQ ID 7914>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2999(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2708

A DNA sequence (GASx1109) was identified in *S.pyogenes* <SEQ ID 7915> which encodes the amino acid sequence <SEQ ID 7916>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1270(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2709

A DNA sequence (GASx1114R) was identified in *S.pyogenes* <SEQ ID 7917> which encodes the amino acid sequence <SEQ ID 7918>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4021(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2710

A DNA sequence (GASx1149) was identified in *S.pyogenes* <SEQ ID 7919> which encodes the amino acid sequence <SEQ ID 7920>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.70   Transmembrane   12 -  28 (  12 -  29)

----- Final Results -----
               bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2711

A DNA sequence (GASx1150) was identified in *S.pyogenes* <SEQ ID 7921> which encodes the amino acid sequence <SEQ ID 7922>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2712

A DNA sequence (GASx1160) was identified in *S.pyogenes* <SEQ ID 7923> which encodes the amino acid sequence <SEQ ID 7924>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.19   Transmembrane   15 -  31 (  15 -  31)

----- Final Results -----
              bacterial membrane --- Certainty=0.2275(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2713

A DNA sequence (GASx1167) was identified in *S.pyogenes* <SEQ ID 7925> which encodes the amino acid sequence <SEQ ID 7926>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1404(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB99233 GB:U67563 oxaloacetate decarboxylase alpha chain (oadA)
          [Methanococcus jannaschii]
 Identities = 250/453 (55%), Positives = 325/453 (71%), Gaps = 7/453 (1%)

Query: 13  VAITETVLRDGHQSLMATRLSIEDMLPVLTILDKIGYYSLECWGGATFDACIRFLNEDPW 72
           V I +T  RD  QSL+ATR+  EDMLP+   +D++G+YS+E WGGATFDACIR+LNEDPW
Sbjct: 2   VKIVDTTFRDAQQSLIATRMRTEDMLPIAEKMDEVGFYSMEVWGGATFDACIRYLNEDPW 61

Query: 73  ERLRTLKKGLPNTRLQMLLRGQNLLGYRHYADDIVDKFISLSAQNGIDVFRIFDALNDPR 132
           ERLR LKK + NT LQMLLRGQNL+GYRHY DDIV+KF+  + +NGID+FRIFDALND R
Sbjct: 62  ERLRALKKRIQNTPLQMLLRGQNLVGYRHYPDDIVEKFVIKAHENGIDIFRIFDALNDVR 121

Query: 133 NIQQALRAVKKTGKEAQLCIAYTTSPVHTLNYYLSLVKELVEMGADSICIKDMAGILTPK 192
           N++ A++  KK G E Q  I YT SPVHT++ Y+ L K+L EMG DSICIKDMAG+LTP
Sbjct: 122 NMETAIKTAKKVGAEVQGAICYTISPVHTIDQYVELAKKLEEMGCDSICIKDMAGLLTPY 181

Query: 193 AAKELVSGIKAMTNLPLIVHTHATSGISQMTYLAAVEAGADRIDTALSPFSEGTSQPATE 252
              ELV  +K   +LP+ VH+H TSG++ MTYL  +EAGAD +D A+SPF+ GTSQP TE
Sbjct: 182 EGYELVKRLKEEISLPIDVHSHCTSGLAPMTYLKVIEAGADMVDCAISPFAMGTSQPPTE 241

Query: 253 SMYLALKEASYDITLDETLLEQAANHLRQARQKYLADGILDPSLLFPDPRTLQYQVPGGM 312
           S+ +ALK   YD  LD  LL +  ++  + R+KY    +  P     D R L YQVPGGM
Sbjct: 242 SIVVALKGTKYDTGLDLKLLNEIRDYFMKVREKYKM--LFSPISQIVDARVLVYQVPGGM 299

Query: 313 LSNMLSQLKQANAESKLEEVLAEVPRVRKDLGYPPLVTPLSQMVGTQAAMNVILGKPYQM 372
           LSN++SQLK+  A  K EEVL E+PRVRKDLGYPPLVTP SQ+VGTQA +NV+  + Y++
Sbjct: 300 LSNLVSQLKEQGALDKFEEVLQEIPRVRKDLGYPPLVTPTSQIVGTQAVLNVLTEERYKI 359

Query: 373 VSKEIKQYLAGDYGKTPAPVNEDLKRSQI--GSAPVTTNRPADQLSPEFEVLK--AEVAD 428
           ++ E+  Y+ G YGK PAP+N +L +  +  G  P+T  RPAD L PE+E +K  AE
Sbjct: 360 ITNEVVNYVKGFYGKPPAPINPELLKRVLDEGEKPITC-RPADLLPPEWEKVKKEAEEKG 418

Query: 429 LAQTDEDVLTYALFPSVAKPFLTTKYQTDDVIK 461
           + + +ED+LTYAL+P +A  FL  + + + + K
Sbjct: 419 IVKKEEDILTYALYPQIAVKFLRGELKAEPIPK 451
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2714

A DNA sequence (GASx1168) was identified in *S.pyogenes* <SEQ ID 7927> which encodes the amino acid sequence <SEQ ID 7928>. Analysis of this protein sequence reveals the following:

```
Possible site: 38

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.11   Transmembrane   16 -  32 (   2 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.3845(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2715

A DNA sequence (GASx1170) was identified in *S.pyogenes* <SEQ ID 7929> which encodes the amino acid sequence <SEQ ID 7930>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.06   Transmembrane  211 - 227 ( 208 - 238)
   INTEGRAL    Likelihood = -5.84   Transmembrane  117 - 133 ( 110 - 136)
   INTEGRAL    Likelihood = -5.36   Transmembrane  256 - 272 ( 253 - 274)
   INTEGRAL    Likelihood = -4.67   Transmembrane   44 -  60 (  41 -  64)
   INTEGRAL    Likelihood = -4.19   Transmembrane  287 - 303 ( 287 - 306)
   INTEGRAL    Likelihood = -3.77   Transmembrane  358 - 374 ( 357 - 375)
   INTEGRAL    Likelihood = -2.18   Transmembrane   20 -  36 (  16 -  38)
   INTEGRAL    Likelihood = -0.85   Transmembrane   90 - 106 (  90 - 106)
   INTEGRAL    Likelihood = -0.53   Transmembrane  165 - 181 ( 164 - 181)

----- Final Results -----
               bacterial membrane --- Certainty=0.3824(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA05140 GB:AJ002015 methylmalonyl-CoA decarboxylase,
          beta-subunit [Propionigenium modestum]
 Identities = 231/395 (58%), Positives = 293/395 (73%), Gaps = 19/395 (4%)

Query: 1   MLDVLNQMVQSSGLAHLTVNNLIMICLASFFLYLGIKKEYEPYLMVPIAFGILLVNLPMA 60
           ML  +     S+G   L + ++IM+ +A  FLYL I KE+EP L+VPI+FGILL NLP A
Sbjct: 1   MLQAILDFYHSTGFYGLNMGSIIMMLVACVFLYLAIAKEFEPLLLVPISFGILLTNLPFA 60

Query: 61  GLMDHP---------ANG---------NPGGLLYYLYKGTSLGIYPPLIFLCLGASTDFG 102
           G+M  P         A+G          PGGLLYYL++G  LGI+PPLIFL +GA TDFG
Sbjct: 61  GMMAEPLLEVHEKLSASGAHLYTAHTAEPGGLLYYLFQGDHLGIFPPLIFLGVGAMTDFG 120

Query: 103 PLIANPKTILLGGAAQVGIFLAFFLAIMLGM-TPQEAASVGIIGGADGPTAIYVTTKLAP 161
           PLI+NPK++LLG AAQ GIF+ FF AI  G+ T QEAAS+GIIGGADGPTAI++++KLAP
Sbjct: 121 PLISNPKSLLLGAAAQFGIFVTFFGAIASGLFTAQEAASIGIIGGADGPTAIFLSSKLAP 180

Query: 162 DLLSTIALAAYSYMALVPIIQPPIIKLLTTKAERQVKMTQARTVSQKEKIIFPIMVTIFV 221
            L+  IA+AAYSYMALVPIIQPPI+  LT++ ER++KM+Q R VS++EKIIFPI+VTI V
Sbjct: 181 HLMGPIAVAAYSYMALVPIIQPPIMTALTSETERKIKMSQLRLVSKREKIIFPIVVTILV 240

Query: 222 SLLVPSATTLVGCLMLGNLVREIKIVPKIVENLQQVVMFCITIILGLTVGAKANGDLFLS 281
```

```
           SL+VP A TLVG LMLGNL RE  +V ++ +  +  ++  ITI LG+TVGA A  + FL
Sbjct: 241 SLIVPPAATLVGMLMLGNLFRECGVVGRLEDTAKNALINIITIFLGVTVGATATAEAFLK 300

Query: 282 ATTLKIIALGLIAFAAGTAGGVLMGKVMYYLSGNKVNPMIGAAGVSAVPMAARVVQKIGQ 341
             TL I+ LG++AF  GT  GVL+ K M  LS   +NP++G+AGVSAVPMAARV Q +GQ
Sbjct: 301 VETLAILGLGIVAFGIGTGSGVLLAKFMNKLSKEPINPLLGSAGVSAVPMAARVSQVVGQ 360

Query: 342 EEDPSNFLLMHAMGPNVAGVIGSAIASGALLAFFG 376
           + DP+NFLLMHAMGPNVAGVIGSA+++G LL+ FG
Sbjct: 361 KADPTNFLLMHAMGPNVAGVIGSAVSAGVLLSLFG 395
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2716

A DNA sequence (GASx1171R) was identified in *S.pyogenes* <SEQ ID 7931> which encodes the amino acid sequence <SEQ ID 7932>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0851(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF93965 GB:AE004165 citG protein [Vibrio cholerae]
 Identities = 100/287 (34%), Positives = 154/287 (52%), Gaps = 12/287 (4%)

Query: 9   ISQLALKALLYEVSLSPKPGLVDRFDNGAHDDMSFITFIDSMIALSPFFQAYIETGFAYA 68
           +  LA  A++ EV L+PKPGLVD  +NGAH DM   TFI S  A++P+  +++  G+  A
Sbjct: 32  VGHLAYHAMMLEVHLTPKPGLVDTANNGAHRDMDLNTFIASAEAIAPYLHSFVSAGWESA 91

Query: 69  KEEPLLLFNRLRQLGQKAEETMFCATQGINTHKGLNFSMALLLGATGAYLARTPHLMTDL 128
                L + LR +G +AE+ MF ATQG+NTHKG+ F + L+ G+ G   A
Sbjct: 92  GNPAAQLLSALRPIGIEAEQAMFAATQGVNTHKGMIFILGLICGSVGWLKANQ------- 144

Query: 129 GRFSKEDTLAICRLVKPMTAHLIQTDLGHLNTKKEFTYGEQLFVTYGIKGPRGEASEGFT 188
               K D   I   ++     L+  +L      +  T GE+++  YG+ G RGEA+ G
Sbjct: 145 ---LKIDAQHIGETIRQACQFLVIDELKAKRDCEPETAGERIYRQYGLTGARGEAASGLA 201

Query: 189 TLTDHALPYFRQMISQN-DPETSQLRLLVYLMSIVEDGNLIHRGGIEAWKGVKAD-MRLL 246
           +  HALP ++  +++     E +     L+ LM+    D NL+ RGG+      V+     +LL
Sbjct: 202 MVMQHALPAYQACLTKGASTEQALWHTLLVLMANNNDSNLVSRGGLAGLHFVQEQAQQLL 261

Query: 247 LQQDLSTTDLRLALSSYNQCLINQHLSPGGAADLLALTFYFAFLEKL 293
           +       ++  AL++ +  LI +HLSPGG+ADLLA T+    L +L
Sbjct: 262 AKGGFLYQEIEQALTALDSVLIEKHLSPGGSADLLAATWLIYELVQL 308
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2717

A DNA sequence (GASx1172R) was identified in *S.pyogenes* <SEQ ID 7933> which encodes the amino acid sequence <SEQ ID 7934>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2501(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12389 GB:Z99107 similar to transcriptional regulator (GntR
          family) [Bacillus subtilis]
 Identities = 60/205 (29%), Positives = 99/205 (48%), Gaps = 3/205 (1%)

Query: 19  PLKIAFYNALKKTIILRQIPVGSRINEKEFSIALNISRTPIRYALGLLSEEHLVEHIPKK 78
           P  + FYN LKK I       G RINE + + +  +SR+PIR A+ LL ++ L++   +
Sbjct: 11  PYYLQFYNQLKKMIFNGTFKPGERINETQLAKSFGVSRSPIREAMRLLEKDGLLKADDRN 70

Query: 79  GIIVKGVSIKDACEIFEIRKALETLATVQAMHLMTEEDFKVMHNLLEDCETFI--AEDDT 136
           G  +  ++ KD  EI++IR  LE LA    +   EE+  ++    LE+ E  I    +DT
Sbjct: 71  GFSITSLTAKDVDEIYKIRIPLEQLAVELVIDEADEEELTILEKQLEETEKAIHNGTEDT 130

Query: 137 NRILDNFNAFNNLIYSYSQMVRLKEIVTELQAYLVYFRKISISSVERRKRALSEHWMIYR 196
             I  N   F+ L+  +S    LK ++  +   + + R ++ +   R +  L EH  I+
Sbjct: 131 EIIRLN-QKFHELLVDFSHNRHLKNLLEHVNDLIHFCRILNYTGDHRAETILREHRRIFE 189

Query: 197 GMKNKDHEQITLITHEHLNSSLEFI 221
            +K K+ E        H N   E +
Sbjct: 190 EVKKKNKEAAKQHVLAHFNHDCEHL 214
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2718

A DNA sequence (GASx1173R) was identified in *S.pyogenes* <SEQ ID 7935> which encodes the amino acid sequence <SEQ ID 7936>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.99   Transmembrane  450 - 466 ( 445 - 473)
   INTEGRAL    Likelihood = -9.61   Transmembrane   33 -  49 (  30 -  55)
   INTEGRAL    Likelihood = -8.55   Transmembrane  326 - 342 ( 321 - 346)
   INTEGRAL    Likelihood = -7.01   Transmembrane  288 - 304 ( 286 - 311)
   INTEGRAL    Likelihood = -6.79   Transmembrane   95 - 111 (  88 - 114)
   INTEGRAL    Likelihood = -4.99   Transmembrane  265 - 281 ( 264 - 285)
   INTEGRAL    Likelihood = -4.62   Transmembrane  208 - 224 ( 204 - 228)
   INTEGRAL    Likelihood = -3.13   Transmembrane  126 - 142 ( 126 - 145)
   INTEGRAL    Likelihood = -2.81   Transmembrane  366 - 382 ( 365 - 383)
   INTEGRAL    Likelihood = -2.34   Transmembrane  419 - 435 ( 417 - 438)

----- Final Results -----
               bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9169> which encodes the amino acid sequence <SEQ ID 9170>. Analysis of this protein sequence reveals the following:

```
Possible cleavage site: 39
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood =-10.99   Transmembrane  443 - 459 ( 438 - 466)
```

```
INTEGRAL    Likelihood = -8.55   Transmembrane  319 - 335 ( 314 - 339)
INTEGRAL    Likelihood = -7.01   Transmembrane  281 - 297 ( 279 - 304)
INTEGRAL    Likelihood = -6.79   Transmembrane   88 - 104 (  81 - 107)
INTEGRAL    Likelihood = -4.99   Transmembrane  258 - 274 ( 257 - 278)
INTEGRAL    Likelihood = -4.62   Transmembrane  201 - 217 ( 197 - 221)
INTEGRAL    Likelihood = -3.13   Transmembrane  119 - 135 ( 119 - 138)
INTEGRAL    Likelihood = -2.81   Transmembrane  359 - 375 ( 358 - 376)
INTEGRAL    Likelihood = -2.34   Transmembrane  412 - 428 ( 410 - 431)

----- Final Results -----
               bacterial membrane --- Certainty=0.539(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG08853 GB:AE004959 probable citrate transporter [Pseudomonas aeruginosa]
 Identities = 199/468 (42%), Positives = 296/468 (62%), Gaps = 41/468 (8%)

Query: 9   LLTMLAYAMIIVFMYVVMKKKMTPFTALVMIPLIMTIAVILTGSADFNADAKFVAFVGDG 68
           +LT+LA+AM+  FM+++M K+++   AL+++P                    +AF   G
Sbjct: 1   MLTLLAFAMVATFMFLIMTKRLSALIALILVP--------------------IAFALIG 39

Query: 69  GIAKDLTAIGPMVMYGINNTAKTGIMLLFAILFFSVMLDAGLFDPITEKMIRFAKGDPMK 128
           G A  L   GPM++ GI   A TG+ML+FAIL+F++M+D+GLFDP   K++R  KGDP+K
Sbjct: 40  GFAAGL---GPMMLDGIRTLAPTGVMLMFAILYFAIMIDSGLFDPAVRKILRLVKGDPLK 96

Query: 129 VLIATAVVAAAVSLNGDGTTTTLICCSAFLPIYKKLDMKIMNLGVLIILQNTIMNLLPWG 188
           V + TA +A  VSL+GDG+TT +IC +A LP+Y +L M  + +  LI+L + ++N+ PWG
Sbjct: 97  VSLGTAALAMIVSLDGDGSTTYMICVAAVLPLYSRLGMSPLVMACLIMLSSGVLNMTPWG 156

Query: 189 GPTARAMSVLGVGP-EILGYLAPGMILSLL--YVICWVAPSMGRKERARLGVIDL--SEE 243
           GPTARA S L V P +I   + P MI  LL  + I W+    G++ERARLG + L    E
Sbjct: 157 GPTARAASALHVDPADIFVPMIPAMIAGLLAIFAIAWI---YGKRERARLGELHLPTDHE 213

Query: 244 DMRQLTDITDPDTLFIRRPKNFVFNAILTIGLITWLVAGSFNKSIAMAPLLLFAVGTCIA 303
           D+ +++     P+     RRPK   FNAILT+ L+  L+AG     + M  L + A G  IA
Sbjct: 214 DLAEISVSQYPEA---RRPKLLWFNAILTVVLMATLIAGL----LPMPVLFMIAFG--IA 264

Query: 304 LMVNYPVLKDQSKRIGDNAGDAVQVVILVFAAGIFMGLFQGSGMASALAQSFATIIPKQL 363
           ++VNYP +++Q KRIG +A + + VV L+FAAG+F G+  G+GM  A+++S   +IP  L
Sbjct: 265 MIVNYPCIQEQKKRIGAHAENILAVVSLIFAAGVFTGILSGTGMVDAMSKSLLAVIPPAL 324

Query: 364 AGFWGLVIALVSAPGTFFISNDGFYYGILPVLAEAGAEYGFSNMAMALASLMGQAFHLLS 423
             +   + ALVS P TFF+SND FYYG+LP+L +A AEYG + + MA AS++GQ  HLLS
Sbjct: 325 GPYLATITALVSMPFTFFMSNDAFYYGVLPILTQAAAEYGITPVEMARASIVGQPVHLLS 384

Query: 424 PLVAFIYLLLRLTGLDMGEWQKEAAKYALIIFVIFVVTIIAMGQMPLY 471
           PLV   YLL+ L  +D G+ Q+   K+A+++ +  +  + +G  PL+
Sbjct: 385 PLVPSTYLLVGLAKIDFGDHQRFTLKWAVLVCLAILAMALLLGLFPLF 432
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2719

A DNA sequence (GASx1174) was identified in *S.pyogenes* <SEQ ID 7937> which encodes the amino acid sequence <SEQ ID 7938>. Analysis of this protein sequence reveals the following:

```
Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
bacterial cytoplasm --- Certainty=0.3948(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2720

A DNA sequence (GASx1175) was identified in *S.pyogenes* <SEQ ID 7939> which encodes the amino acid sequence <SEQ ID 7940>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3519(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2721

A DNA sequence (GASx1177) was identified in *S.pyogenes* <SEQ ID 7941> which encodes the amino acid sequence <SEQ ID 7942>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.24   Transmembrane  115 - 131 ( 105 - 137)
   INTEGRAL    Likelihood = -8.92   Transmembrane  208 - 224 ( 204 - 238)
   INTEGRAL    Likelihood = -7.80   Transmembrane  282 - 298 ( 273 - 303)
   INTEGRAL    Likelihood = -4.94   Transmembrane   85 - 101 (  75 - 102)
   INTEGRAL    Likelihood = -4.04   Transmembrane   10 -  26 (   3 -  32)
   INTEGRAL    Likelihood = -3.61   Transmembrane  255 - 271 ( 253 - 271)

----- Final Results -----
             bacterial membrane --- Certainty=0.4694(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB89172 GB:AE000960 oxaloacetate decarboxylase, sodium ion pump
          subunit (oadB) [Archaeoglobus fulgidus]
 Identities = 190/354 (53%), Positives = 255/354 (71%), Gaps = 8/354 (2%)

Query: 16  IVMMVIGALLMYLGIKKEYEPTLLVPMGLGTILVNFPGSGVLTQVVNGVEQEGVFEALFN 75
           +VM+ +G LL+YLGI K+ EP LLVP+G+G ILVN PG G+         E+  +F+
Sbjct: 5   LVMIGVGLLLVYLGIVKKMEPLLLVPIGIGAILVNIPGGGL-------AEEGSIFDLFLK 57
```

```
Query: 76  FGIGTELFPLLIFIGIGAMIDFGPLLQNPFMLLFGDAAQFGIFFVVVVAVLAGFDIKEAA 135
           + I TE+ PLLIF+G+GA+ DF PLL NP   L G AAQ GIF  ++ A+  GF  +EAA
Sbjct: 58  YLIHTEIVPLLIFLGLGALTDFSPLLANPKTFLLGAAAQIGIFAALIAALFLGFTPQEAA 117

Query: 136 SIGIIGAADGPTSIFVANQLAKDLLGPITVAAYSYMALVPIIQPFAIKLVTTKKERRIRM 195
           SIGIIG ADGPT+I+    LA  LL    VAAYSYM+LVPIIQP  IK +T+ +ER+I+M
Sbjct: 118 SIGIIGGADGPTTIYTTTILAPHLLAATAVAAYSYMSLVPIIQPPIIKALTSSRERKIKM 177

Query: 196 TYKAENVSQMTKILFPIIITLVAGFIAPISLPLVGFLMFGNLLRECGVLDRLSQTAQNEL 255
             +   VS+  KILFPI   +++GF+AP +LPLVG LM GNL RE GV DRL++ A  EL
Sbjct: 178 R-QLRIVSKKEKILFPIATIIISGFLAPKALPLVGMLMTGNLFRESGVTDRLAKGASEEL 236

Query: 256 VNIISILLGLTISIKMQADLFLNVQTLLIIVFGLLAFIMDSIGGVMFAKFLNLFRKEKIN 315
           +NI++I+LGL++   M+A+ FL  +TLL++  G++AF   + GGV+ AK +NLF KEKIN
Sbjct: 237 MNIMTIILGLSVGSTMRAESFLTQKTLLVLALGVVAFAAATAGGVLLAKVMNLFLKEKIN 296

Query: 316 PMIGAAGISAFPMSSRVIQKMATDEDPQNFILMYAVGANVSGQIASVIAGGLLL 369
           PMIGAAG+SA PMS+RV+Q++A +EDP N ILM+A+G NV+G I S +A G+L+
Sbjct: 297 PMIGAAGVSAVPMSARVVQRLAIEEDPHNHILMHAMGPNVAGVIGSAVAAGVLI 350
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2722

A DNA sequence (GASx1178) was identified in *S.pyogenes* <SEQ ID 7943> which encodes the amino acid sequence <SEQ ID 7944>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.50   Transmembrane   21 -  37 (   8 -  43)

----- Final Results -----
               bacterial membrane --- Certainty=0.4800(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2723

A DNA sequence (GASx1179) was identified in *S.pyogenes* <SEQ ID 7945> which encodes the amino acid sequence <SEQ ID 7946>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1906(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF93961 GB:AE004165 citrate lyase, gamma subunit [Vibrio cholerae]
 Identities = 46/97 (47%), Positives = 64/97 (65%)

Query: 1  MDIKQTAVAGSLESSDLMITVSPNDEQTITITLDSSVEKQFGNHIRQLIHQTLVNLKVTA 60
          M I   A AG+LESSDL + + PN++  I + LDS+VE+QFG+ IRQ++  TL  ++V
Sbjct: 1  MKIAHPAFAGTLESSDLQVRIEPNNDGGIELVLDSTVEQQFGHAIRQVVLHTLDAMQVRD 60

Query: 61 AKVEAVDKGALDCTIQARTIAAVHRAAGIDQYDWKEI 97
           A V   DKGALDC I+AR  AAV RA  +   +W ++
Sbjct: 61 ALVTIEDKGALDCVIRARVQAAVMRACDVQNIEWSQL 97
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2724

A DNA sequence (GASx1181) was identified in *S.pyogenes* <SEQ ID 7947> which encodes the amino acid sequence <SEQ ID 7948>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.65   Transmembrane   74 -  90 (  74 -  90)

----- Final Results -----
               bacterial membrane --- Certainty=0.1659(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA71632 GB:Y10621 CILB, citryl-CoA lyase beta subunit
           [Leuconostoc mesenteroides]
 Identities = 187/293 (63%), Positives = 237/293 (80%), Gaps = 1/293 (0%)

Query: 2   ERLRRTMMFVPGANAAMLRDAPLFGADSIMFDLEDSVSLKEKDTSRALVHFALKTFDYSS 61
           ERLRRTMMFVPG N AM++DA +FGADSIMFDLED+VSL EKD++R LV+ AL+T DY S
Sbjct: 4   ERLRRTMMFVPGNNPAMVKDAGIFGADSIMFDLEDAVSLAEKDSARYLVYEALQTVDYGS 63

Query: 62  VETVVRVNGLDS-CGALDIEAVVLAGVNVIRLPKTETAQDIIDVEAVIERVERENSIEVG 120
            E VVR+NGLD+      DI+A+V AG++VIRLPK ETA  + ++E++I   E+E    VG
Sbjct: 64  SELVVRINGLDTPFYKNDIKAMVKAGIDVIRLPKVETAAMMHELESLITDAEKEFGRPVG 123

Query: 121 RTRMMAAIESAEGVLNAREIAKASKRLIGIALGAEDYVTNMKTRRYPDGQELFFARSMIL 180
            T MMAAIESA GV+NA EIA AS R+IGIAL AEDY T+MKT RYPDGQEL +AR++IL
Sbjct: 124 TTHMMAAIESALGVVNAVEIANASDRMIGIALSAEDYTTDMKTHRYPDGQELLYARNVIL 183

Query: 181 HAARAAGIAAIDTVYSDVNNTEGFQNEVRMIKQLGFDGKSVINPRQIPLVNEIYTPTKKE 240
           HAARAAGIAA DTV++++N+ EGF  E ++I QLGFDGKS+INPRQI +VN++Y PT+KE
Sbjct: 184 HAARAAGIAAFDTVFTNLNDEEGFYRETQLIHQLGFDGKSLINPRQIEMVNKVYAPTEKE 243

Query: 241 IDHAKQVIWAIREAESKGSGVISLNGKMVDKPIVERAERVIALATAAGVLSEE 293
           I++A+ VI AI EA+ KGSGVIS+NG+MVD+P+V RA+RV+ LA A  ++ E
Sbjct: 244 INNAQNVIAAIEEAKQKGSGVISMNGQMVDRPVVLRAQRVMKLANANHLVDSE 296
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2725

A DNA sequence (GASx1182) was identified in *S.pyogenes* <SEQ ID 7949> which encodes the amino acid sequence <SEQ ID 7950>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3554(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA71633 GB:Y10621 CILA, citrate CoA-transferase alpha subunit
          [Leuconostoc mesenteroides]
 Identities = 294/511 (57%), Positives = 378/511 (73%), Gaps = 7/511 (1%)

Query: 4   NKLGRDIPQPYADQY--GVFEGELANIKQYDESSRRIKPVKPGDSKLLGSVREAIEKTGL 61
           NK+  D+P     +Q     VFE        +       +++     G+SK+  S+ + +  T L
Sbjct: 3   NKVNIDVPDAILEQLDDSVFESTNYGNPEIQRVGPKVRATT-GESKVQSSIDDVLSNT-L 60

Query: 62  TDGMTISFHHHFREGDFIMNMVLEEIAKMGIKNLSIAPSSIANV-HEPLIDHIKNGVVTN 120
            DGMTISFHHHFREGDF+ N V+ +I  MG +NL++APSS+ NV ++ +I+ IK GVVTN
Sbjct: 61  KDGMTISFHHHFREGDFVFNKVMRKIIDMGYQNLTLAPSSLTNVMNDIVIEAIKKGVVTN 120

Query: 121 ITSSGLRDKVGAAISEGLMENPVVIRSHGGRARAIASGDIHIDVAFLGAPSSDAYGNVNG 180
           ITSSG+R  +G A+S G+++NPV+ RSHG RARAI SG+I IDVAFLG P+SD  GN NG
Sbjct: 121 ITSSGMRGTLGDAVSHGILKNPVIFRSHGARARAIESGEIKIDVAFLGVPNSDEMGNANG 180

Query: 181 TKGKATCGSLGYAMIDAKYADQVVILTDNLVPYPNTPISIPQTDVDYVVTVDAIGDPQGI 240
            G A  GSLGYA+IDA+YAD++V++TD ++PYPNTP SI QT VDYVV VD +GDP  I
Sbjct: 181 MNGDAAFGSLGYALIDAQYADKLVLITDTIMPYPNTPASIKQTQVDYVVKVDKVGDPDKI 240

Query: 241 AKGATRFTKNPKELLIAEYAAKVITNSPYFKEGFSFQTGTGGASLAVTRFMREAMIKENI 300
            GATRFTK+PKEL IA+    VI NS YFK  FSFQTG+GGA+LAVTRF+REAM+ +NI
Sbjct: 241 GSGATRFTKDPKELKIAKTVNDVIVNSKYFKNDFSFQTGSGGAALAVTRFLREAMMAQNI 300

Query: 301 KASFALGGITNAMVELLEEELVEKILDVQDFDHPSAVSLGKHAEHYEIDANMYASPLSKG 360
            ASFALGGIT   V+LL E LV +++DVQDFD  +A S+       EIDA+ YA P +KG
Sbjct: 301 MASFALGGITKPTVDLLNEGLVNRVMDVQDFDKGAASSMKLSPNQQEIDASWYADPANKG 360

Query: 361 AVINQLDTCILSALEVDTNFNVNVMTGSDGVIRGASGGHCDTAFAAKMSLVISPLIRGRI 420
           A++++LD  ILSALEVDTNFNVNVM+GSDGVIRGA GGH D A  AK++++  PL+RGRI
Sbjct: 361 AMVDKLDVAILSALEVDTNFNVNVMSGSDGVIRGAIGGHQDAA-TAKLTIISVPLVRGRI 419

Query: 421 PTFVDEVNTVITPGTSVDVIVTEVGIAINPNRQDLVDHFKSL-NVPQFSIEELKEKAYAI 479
            T V +VNTVITPG S+DV+VTEVGIAINP R DLV+  K +  +P +SIEEL++KA  I
Sbjct: 420 ATIVPKVNTVITPGDSIDVVVTEVGIAINPKRTDLVEQLKQVPGLPIYSIEELQQKAEKI 479

Query: 480 VGTPERIQYGDKVVALIEYRDGSLMDVVYNV 510
           VG P  +++ D+VVA+ EYRDGS++D++  V
Sbjct: 480 VGQPAPLKFTDRVVAVAEYRDGSVIDIIKEV 510
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2726

A DNA sequence (GASx1183) was identified in *S.pyogenes* <SEQ ID 7951> which encodes the amino acid sequence <SEQ ID 7952>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA71634 GB:Y10621 CILG, hypothetical protein [Leuconostoc
          mesenteroides]
 Identities = 65/176 (36%), Positives = 97/176 (54%), Gaps = 3/176 (1%)

Query: 21  DTYFSGEAIQLSDMLRAREERALRQLHLLKEYPEGSLLSVTMNIPGPIKTSPKLLEAFDI 80
           D +  GE + L  +L  RE R   Q  L+  +P   + SV +N+PGPIKTSPKL   F I
Sbjct: 2   DYFEGGERLNLMQVLDNREWREKYQKQLMASFPTAVITSVKLNLPGPIKTSPKLQSVFQI 61

Query: 81  VIKAIQTALADDKICYQLRLL-PTTGYEYYLITSLPSRDLKLKMIALETELPIGRLMDLD 139
           +I  +     D +I  +   +  TG + + +TS   + +K  MI  E    +GRL+DLD
Sbjct: 62  IINDLNPVFKDLQIIKEASFVDQITGPDIFFVTSGCLKLVKQIMITFEESHLLGRLLDLD 121

Query: 140 VLVLQNDLPHSISRTVLGGSPRQCFICSKEAKVCGRLRKHSVEEMQTAISKLLHSF 195
           V+    D    +SR  LG +PR+C +C K+AK C +   HS+ E  + I+K+LH+F
Sbjct: 122 VMCQNAD--KQLSREELGFAPRKCLLCGKDAKTCIKEGNHSLAEGYSQINKMLHNF 175
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2727

A DNA sequence (GASx1184) was identified in *S.pyogenes* <SEQ ID 7953> which encodes the amino acid sequence <SEQ ID 7954>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3730(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB99233 GB:U67563 oxaloacetate decarboxylase alpha chain (oadA)
          [Methanococcus jannaschii]
 Identities = 245/441 (55%), Positives = 336/441 (75%), Gaps = 5/441 (1%)

Query: 10  IRITETVLRDGQQSQIATRMTTKEMIPILETLDNAGYHALEMWGGATFDSCLRFLNEDPW 69
           ++I +T  RD QQS IATRM T++M+PI E +D  G++++E+WGGATFD+C+R+LNEDPW
Sbjct: 2   VKIVDTTFRDAQQSLIATRMRTEDMLPIAEKMDEVGFYSMEVWGGATFDACIRYLNEDPW 61

Query: 70  ERLRAIRKAVKKTKLQMLLRGQNLLGYRNYADDVVRSFIQKSIENGIDIVRIFDALNDPR 129
           ERLRA++K ++ T LQMLLRGQNL+GYR+Y DD+V  F+ K+ ENGIDI RIFDALND R
Sbjct: 62  ERLRALKKRIQNTPLQMLLRGQNLVGYRHYPDDIVEKFVIKAHENGIDIFRIFDALNDVR 121

Query: 130 NLQTAVSATKKFGGHAQVAISYTTSPVHTIDYFVELAKAYQAIGADSICIKDMAGVLTPE 189
           N++TA+    KK G   Q AI YT SPVHTID +VELAK  + +G DSICIKDMAG+LTP
Sbjct: 122 NMETAIKTAKKVGAEVQGAICYTISPVHTIDQYVELAKKLEEMGCDSICIKDMAGLLTPY 181
```

```
Query: 190 IGYQLVKCIKENTTIPLEVHTHATSGISEMTYLKVAEAGADIIDTAISSFSGGTSQPATE 249
           GY+LVK +KE  ++P++VH+H TSG++ MTYLKV EAGAD++D AIS F+ GTSQP TE
Sbjct: 182 EGYELVKRLKEEISLPIDVHSHCTSGLAPMTYLKVIEAGADMVDCAISPFAMGTSQPPTE 241

Query: 250 SMAIALTDLGFDTGLDMQEVAKVAEYFNTIRDHYREIGILNPKVKDTEPKTLIYQVPGGM 309
           S+ +AL    +DTGLD++ + ++ +YF  +R+ Y+   + +P  +  + + L+YQVPGGM
Sbjct: 242 SIVVALKGTKYDTGLDLKLLNEIRDYFMKVREKYKM--LFSPISQIVDARVLVYQVPGGM 299

Query: 310 LSNLLSQLTEQGLTDKYEEVLAEVPKVRADLGYPPLVTPLSQMVGTQALMNIISGERYKV 369
           LSNL+SQL EQG  DK+EEVL E+P+VR DLGYPPLVTP SQ+VGTQA++N+++ ERYK+
Sbjct: 300 LSNLVSQLKEQGALDKFEEVLQEIPRVRKDLGYPPLVTPTSQIVGTQAVLNVLTEERYKI 359

Query: 370 VPNEIKDYVRGLYGQSPAPLAEGIKEKIIGD-EAVITCRPADLIEPQMIYLRDEIAP--Y 426
           + NE+ +YV+G YG+ PAP+   + ++++ + E  ITCRPADL+ P+   ++ E
Sbjct: 360 ITNEVVNYVKGFYGKPPAPINPELLKRVLDEGEKPITCRPADLLPPEWEKVKKEAEEKGI 419

Query: 427 AHSEEDVLSYASFPQQARDFL 447
              EED+L+YA +PQ A  FL
Sbjct: 420 VKKEEDILTYALYPQIAVKFL 440
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2728

A DNA sequence (GASx1185R) was identified in *S.pyogenes* <SEQ ID 7955> which encodes the amino acid sequence <SEQ ID 7956>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2497(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF93960 GB:AE004165 citrate (pro-3S)-lyase ligase [Vibrio cholerae]
 Identities = 118/336 (35%), Positives = 183/336 (54%), Gaps = 5/336 (1%)

Query: 4   YTISKVFPSDKTTMASVKNLLHQEGIRLDAHLDYTCAIMNAQNDVIATGSYFGNSLRCLC 63
           YT S+V   ++T +  +K  L Q  + +D  +++   +    N +IA G   G+ L+ +
Sbjct: 10  YTFSRVSTKNRTKLLQIKEFLCQHQLTVDDDVEHF-VVAYGTNQIIACGGIAGHVLKSIA 68

Query: 64  VSSAYQGEGLLNRIVSHLIDEEYALGNYHLFVYTKTSSAAFFKDLGFTEIVHIDNHISFL 123
           VS A QG G   ++++ L +  Y +G + LF++TK ++   F+  GF  +  ++ HI+ L
Sbjct: 69  VSPALQGTGFALKLMTELTNFAYEMGRFSLFLFTKPANIDLFRQCGFFLVDKVEPHIALL 128

Query: 124 ENKKTGFQDYLMTLNKPEQTPGKVAAIVINANPFTLGHQFLVEKAARENDWVHLFMVSED 183
           EN      Y   L   + +  K+ +IV+NANPFTLGHQ+L+E+A  + DWVHLF+V  +
Sbjct: 129 ENSPNRLSVYCKQLQLLKMSGRKIGSIVMNANPFTLGHQYLIEQACEQCDWVHLFVVKAE 188

Query: 184 RSLIPFSVRKRLIQEGLAHLDNVIYHETGPYLISQATFPAYFQKEDNDVIKSQALLDTAI 243
                ++ R  +I+ G  HL N+  H     Y+IS+ATFP+YF K+   V +S   LD +I
Sbjct: 189 NKDFSYADRMAMIKAGSKHLLNLTIHSGSDYIISRATFPSYFIKDQQVVNQSHTALDLSI 248

Query: 244 FL-KIAQTLQITKRYVGEEPTSRVTAIYNEIM---AEQLQQAGILLDILPRKAINQQQDP 299
           F   IA  L IT R+VG EP   VT  YN+ M    E+   A  + ++  +  Q  P
Sbjct: 249 FRHSIAPALGITHRFVGSEPICTVTRHYNQAMRRWLEEAHDASAPIQVVEIERSQQASQP 308

Query: 300 ISASTARQALKDNDWDLLAKLLPKTSLDYFCSLKAQ 335
```

```
          ISAS  R  LK   +  +A L+PKT+  Y C   A+
Sbjct: 309 ISASRVRYLLKQFGFAAIADLVPKTTYSYLCQHYAE 344
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2729

A DNA sequence (GASx1187) was identified in *S.pyogenes* <SEQ ID 7957> which encodes the amino acid sequence <SEQ ID 7958>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4790(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2730

A DNA sequence (GASx1188R) was identified in *S.pyogenes* <SEQ ID 7959> which encodes the amino acid sequence <SEQ ID 7960>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3956(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2731

A DNA sequence (GASx1190) was identified in *S.pyogenes* <SEQ ID 7961> which encodes the amino acid sequence <SEQ ID 7962>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1274(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2732

A DNA sequence (GASx1196R) was identified in *S.pyogenes* <SEQ ID 7963> which encodes the amino acid sequence <SEQ ID 7964>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2733

A DNA sequence (GASx1211) was identified in *S.pyogenes* <SEQ ID 7965> which encodes the amino acid sequence <SEQ ID 7966>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1850(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2734

A DNA sequence (GASx1219R) was identified in *S.pyogenes* <SEQ ID 7967> which encodes the amino acid sequence <SEQ ID 7968>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2284(Affirmative) < succ>
```

```
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2735

A DNA sequence (GASx1225) was identified in *S.pyogenes* <SEQ ID 7969> which encodes the amino acid sequence <SEQ ID 7970>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2062(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2736

A DNA sequence (GASx1229) was identified in *S.pyogenes* <SEQ ID 7971> which encodes the amino acid sequence <SEQ ID 7972>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2755(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2737

A DNA sequence (GASx1247R) was identified in *S.pyogenes* <SEQ ID 7973> which encodes the amino acid sequence <SEQ ID 7974>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.32   Transmembrane   55 -  71 (  53 -  81)
```

```
INTEGRAL    Likelihood = -6.00   Transmembrane   74 -  90 (  72 -  95)
INTEGRAL    Likelihood = -2.18   Transmembrane   95 - 111 (  95 - 111)
INTEGRAL    Likelihood = -1.54   Transmembrane  124 - 140 ( 123 - 141)

----- Final Results -----
              bacterial membrane --- Certainty=0.3527(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14326 GB:Z99116 yqjA [Bacillus subtilis]
 Identities = 97/306 (31%), Positives = 154/306 (49%)

Query: 6   RTLKMTLATIVAILIAYQLHLDYAMSAGIIALLSVLDTRKSSLVVARNRLLSFFLAFGIA 65
           RT+K  L T +AI I+  LHL    SAGII +L +  T+K SL  +  R  +  LA   +
Sbjct: 7   RTIKTALGTALAIYISQLLHLQNFASAGIITILCIQITQKRSLQASWARFWACCLAIAFS 66

Query: 66  MMCFSLFGFTTVGFMCYLLIIIPLLYHFQIEAGLVPITVLVTHLIAKKSIALPILSNEFM 125
            + F L G+         LLI IP+    +I  G+V  +V++ HL     I    + NE
Sbjct: 67  YLFFELIGYHPFVIGALLLIFIPITVLLKINEGIVTSSVIILHLYMSGGITPTFIWNEVQ 126

Query: 126 LFFVGTSVALLFNAYMGPQDQQIRYYHQKVESDLKGILYRFESFLLEGKGQNEGLLIKNL 185
           L  VG  VALL N YM   D+++  Y +K+E +   I    E +LL G+    G  I
Sbjct: 127 LITVGIGVALLMNLYMPSLDRKLIAYRKKIEDNFAVIFAEIERYLLTGEQDWSGKEIPET 186

Query: 186 DKILDEALKLVYRERHNQLFQQTNYQVHYFEMRRQQNRLLGQMAINVNTLMRQSKESILL 245
            +++ EA  L YR+  N + +  N   HYF+MR +Q  ++ ++   V ++     +  ++
Sbjct: 187 HQLITEAKNLAYRDVQNHILRYENLHYHYFKMREKQFEIIERLLPKVTSISITVDQGKMI 246

Query: 246 SHLFHETACQLSEQNPALTLIDDIEQLLETFRHGDLPQTREEFERRAVLFQLLQDLERFI 305
           +   H+    +  N A  +  +  + + F    LP TREEFE RA LF LL ++E+++
Sbjct: 247 AEFIHDLREAIHPGNTAYKFLKRLADMRKEFEEMPLPATREEFEARAALFHLLGEMEQYL 306

Query: 306 LLKVEF 311
           ++K  F
Sbjct: 307 VIKSYF 312
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2738

A DNA sequence (GASx1261) was identified in *S.pyogenes* <SEQ ID 7975> which encodes the amino acid sequence <SEQ ID 7976>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.6082(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2739

A DNA sequence (GASx1262R) was identified in *S.pyogenes* <SEQ ID 7977> which encodes the amino acid sequence <SEQ ID 7978>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.06   Transmembrane   38 -  54 (  37 -  55)

----- Final Results -----
               bacterial membrane --- Certainty=0.3824(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2740

A DNA sequence (GASx1265R) was identified in *S.pyogenes* <SEQ ID 7979> which encodes the amino acid sequence <SEQ ID 7980>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2741

A DNA sequence (GASx1270) was identified in *S.pyogenes* <SEQ ID 7981> which encodes the amino acid sequence <SEQ ID 7982>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4063(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2742

A DNA sequence (GASx1290R) was identified in *S.pyogenes* <SEQ ID 7983> which encodes the amino acid sequence <SEQ ID 7984>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.37   Transmembrane  180 - 196 ( 172 - 207)
   INTEGRAL    Likelihood =-10.19   Transmembrane   34 -  50 (  30 -  53)
   INTEGRAL    Likelihood = -4.09   Transmembrane  233 - 249 ( 232 - 250)

----- Final Results -----
              bacterial membrane --- Certainty=0.5946(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB88010 GB:L21856 MalA [Streptococcus pneumoniae]
 Identities = 66/237 (27%), Positives = 105/237 (43%), Gaps = 28/237 (11%)

Query: 45  MIPVTLHYANMTTYPLERIVTKSLSPITDKTYQALTQGKIEKD---TFQGQSLIRRD--- 98
           M+P+ +  ++  TYPLE  +     P+TDK  Q L++     D   T+ G +
Sbjct: 1   MVPIAIQNSSQETYPLETFIDNVYEPLTDKVVQDLSEHATIVDGTLTYTGTASQAPSVVI 60

Query: 99  GELVLAVLPTKVDLEQLASESTRQIIVTKKEWRFVTPDGKEL-RAHVRGQQQSLADLTTV 157
           G   +  LP  + L       T +++++K        + KEL R   R  Q      T
Sbjct: 61  GPSQIKELPKDLQLHF----DTNELVISK--------ESKELTRISYRAIQ------TEG 102

Query: 158 KAVKDFVNQQWY---DSNKASVLGFLLLTFVLMVCVGTLIVIGLGAFFLTLTKRSRLFMI 214
              KD + Q +     +N+  +  FL+L    +  +    IV       L +TK+SRLF
Sbjct: 103 FKSKDSLTQAFIRLVPTNRVYISLFLVLGASFLFGLNFFIVSLGACLLLYITKKSRLFSF 162

Query: 215 RNFSEGLGLMVNCLAWPSLLAIALSFFIQDPVLIMNCQVFGTLLMLTWVFYKTQFRD 271
           R F E    ++NCL  P+L+ + L  F Q+   ++  Q    +L L  +FYKT FRD
Sbjct: 163 RTFKECYHFILNCLGLPTLITLILGLFGQNMTTLITVQNILFVLYLVTIFYKTHFRD 219
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2743

A DNA sequence (GASx1294) was identified in *S.pyogenes* <SEQ ID 7985> which encodes the amino acid sequence <SEQ ID 7986>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2104(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2744

A DNA sequence (GASx1303R) was identified in *S.pyogenes* <SEQ ID 7987> which encodes the amino acid sequence <SEQ ID 7988>. Analysis of this protein sequence reveals the following:

```
Possible site: 38

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.07   Transmembrane   13 -  29 (   8 -  38)

----- Final Results -----
               bacterial membrane --- Certainty=0.4227(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2745

A DNA sequence (GASx1307R) was identified in *S.pyogenes* <SEQ ID 7989> which encodes the amino acid sequence <SEQ ID 7990>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2746

A DNA sequence (GASx1312R) was identified in *S.pyogenes* <SEQ ID 7991> which encodes the amino acid sequence <SEQ ID 7992>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1996(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2747

A DNA sequence (GASx1316R) was identified in *S.pyogenes* <SEQ ID 7993> which encodes the amino acid sequence <SEQ ID 7994>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3504(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 271-273
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC66321 GB:AE000792 outer surface protein, putative [Borrelia
          burgdorferi]
 Identities = 127/365 (34%), Positives = 195/365 (52%), Gaps = 14/365 (3%)

Query: 1   MVDLGFSLYPERYDVTKSKAYIDLCHSYGAKRLFMSLLQLAPADHQMFHCYAELIAYANQ 60
           M ++G S+YP       K   Y++    +G  ++F SLL +   +   F  + EL++ AN+
Sbjct: 1   MKEIGISIYPNVSPKNKIIKYLEKSAHFGFTQVFTSLLYI---NGNEFDIFKELLSIANK 57

Query: 61  LGIRVIADVSPSFISQAGWSDQLIERA------HAFGLAGLRLDEALPLAEIVTLTRNPF 114
            G++ I DVSP    + G     +            G   +RLD       E   +T N
Sbjct: 58  NGMKPIIDVSPEIFKELGIDLSNLRNCPKLDYFKKLGAWAIRLDNTFTGIEESLMTFNDS 117

Query: 115 GLKIELNMSTDKQLLMSLLATDAERSNIIGCHNFYPHEFTGLSWQHFKDMSRFYHEHDIE 174
            LKI+LN+S   + + +++       N++GCHNFYPH++TGLS   FK+ ++ + +  + I
Sbjct: 118 DLKIQLNISNINKHIDTIMYFKPNIKNLLGCHNFYPHKYTGLSRNFFKETTKIFKHYSIP 177

Query: 175 TAAFITAQSASE-GPWLLAEGLPTVEDHRHLPIGLQVELMKAIGTIDNILISNQFISEEE 233
           TAAFI++ +A E       EG+PT+E HR   I  Q + +   G ID +LISN F SE E
Sbjct: 178 TAAFISSNNAEECARGKEKEGVPTLESHRSKDIETQAKDLFKEG-IDTVLISNCFPSETE 236

Query: 234 LAACTQALARPVTTIKVRPIIDLTEVEEQII-GYPHCYRGDVSDYVIRSTMPRLVYAQES 292
           L   ++ + R +  +K     D   VE++II    H  RGD++ Y IRSTMPR+ Y  +
Sbjct: 237 LKKVSK-VNRNILELKADLNPDANSVEKEIILENLHFNRGDINSYRIRSTMPRVYYNNKK 295

Query: 293 IAPRDQSKEVKRGSIIIDNDRYHRYKGELQIALKNFTVSSKANVVAEVREDYLSLLDDLR 352
             P     E+K+G I+ID+  Y  Y GELQIALK+   +   NVV ++  D + LL+ +
Sbjct: 296 F-PVHSPNEIKKGDILIDSSEYLGYTGELQIALKDTPNNGLVNVVGKIINDEIYLLEKIE 354

Query: 353 PWQEF 357
           PW++F
Sbjct: 355 PWEKF 359
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2748

A DNA sequence (GASx1319) was identified in *S.pyogenes* <SEQ ID 7995> which encodes the amino acid sequence <SEQ ID 7996>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.50   Transmembrane  127 - 143 ( 125 - 151)
   INTEGRAL    Likelihood = -7.43   Transmembrane   17 -  33 (  15 -  36)
   INTEGRAL    Likelihood = -5.68   Transmembrane   39 -  55 (  36 -  57)
   INTEGRAL    Likelihood = -1.86   Transmembrane   60 -  76 (  59 -  77)
   INTEGRAL    Likelihood = -0.59   Transmembrane   85 - 101 (  85 - 101)

----- Final Results -----
               bacterial membrane --- Certainty=0.4800(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2749

A DNA sequence (GASx1320) was identified in *S.pyogenes* <SEQ ID 7997> which encodes the amino acid sequence <SEQ ID 7998>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.81   Transmembrane   35 -  51 (  35 -  51)

----- Final Results -----
               bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2750

A DNA sequence (GASx1321) was identified in *S.pyogenes* <SEQ ID 7999> which encodes the amino acid sequence <SEQ ID 8000>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2751

A DNA sequence (GASx1329) was identified in *S.pyogenes* <SEQ ID 8001> which encodes the amino acid sequence <SEQ ID 8002>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.28   Transmembrane   64 -  80 (  64 -  80)

----- Final Results -----
               bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2752

A DNA sequence (GASx1332R) was identified in *S.pyogenes* <SEQ ID 8003> which encodes the amino acid sequence <SEQ ID 8004>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2753

A DNA sequence (GASx1333) was identified in *S.pyogenes* <SEQ ID 8005> which encodes the amino acid sequence <SEQ ID 8006>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2754

A DNA sequence (GASx1335R) was identified in *S.pyogenes* <SEQ ID 8007> which encodes the amino acid sequence <SEQ ID 8008>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF96047 GB:AE004354 uridine phosphorylase [Vibrio cholerae]
 Identities = 46/167 (27%), Positives = 72/167 (42%), Gaps = 12/167 (7%)

Query: 8   GVKEMISTGTCGVLVP-IAENRFLVPVKALRDEGTSYHYVAPSRYIDIDPKMLRLIEKTL 66
           G K ++  G+ G +   I     ++   A+RDEG S  Y+  +       +++  +++ L
Sbjct: 79  GAKAIVRVGSAGAMQSEIGLGELILVEGAVRDEGGSKAYIGAAYPAYSSFELVVEMQRFL 138

Query: 67  LAQGLAYQEVITWSTDGFYR-ETKEKVAHRQEEGCSVVEMECSALAAVAQLRG-----IL 120
             Q +     I  S D FY  E  E   +   +G    +ME SAL  V +LRG     +L
Sbjct: 139 AEQSVPIHRGIVRSHDSFYTDEEAELCRYWHRKGILAADMETSALLTVGRLRGLQVASVL 198

Query: 121 WGQLLFTADTLADVEVY---DQRNWGADSFSFALHLCLEVLNTLEKD 164
             +L+  D  A V  Y   DQR    +  + A    L  LN L+ D
Sbjct: 199 NNVVLYEQDVQAGVNQYVNADQRMMQGE--TLAARAALHALNALKFD 243
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2755

A DNA sequence (GASx1353) was identified in *S.pyogenes* <SEQ ID 8009> which encodes the amino acid sequence <SEQ ID 8010>. Analysis of this protein sequence reveals the following:

```
Possible site: 42

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -5.79   Transmembrane  241 - 257 ( 234 - 260)
   INTEGRAL    Likelihood = -5.15   Transmembrane   44 -  60 (  43 -  65)
   INTEGRAL    Likelihood = -4.78   Transmembrane   74 -  90 (  72 -  92)

----- Final Results -----
               bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2756

A DNA sequence (GASx1354R) was identified in *S.pyogenes* <SEQ ID 8011> which encodes the amino acid sequence <SEQ ID 8012>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -3.45   Transmembrane   68 -  84 (  65 -  86)

----- Final Results -----
               bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB83831 GB:AL162753 putative integral membrane protein
          [Neisseria meningitidis]
 Identities = 31/72 (43%), Positives = 46/72 (63%), Gaps = 6/72 (8%)

Query: 17  FVIYAFDKRKAIKKKRRISERKLLVITVLFGGF-GALLAAKKYHHKTRKWYFVI----TC 71
           F +Y  DKR+A++ KRRI E +LL +  LFGG+ GA L ++ + HKT K  FV+     T
Sbjct: 38  FALYGIDKRRAVRGKRRIPEHRLL-LPALFGGWAGAYLGSRIFRHKTAKKRFVVLFRLTV 96

Query: 72  YTSILLTLLVTY 83
             ++L TL++ Y
Sbjct: 97  SGNVLATLILIY 108
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2757

A DNA sequence (GASx1363R) was identified in *S.pyogenes* <SEQ ID 8013> which encodes the amino acid sequence <SEQ ID 8014>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2758

A DNA sequence (GASx1367) was identified in *S.pyogenes* <SEQ ID 8015> which encodes the amino acid sequence <SEQ ID 8016>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have an uncleavable N-term signal seq
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA63508 GB:X92946 hypothetical protein [Lactococcus lactis]
 Identities = 64/96 (66%), Positives = 77/96 (79%)

Query: 1  MPRKTFDKAFKLSAVKLILEEEQPVKMVSSTLEIHPNSLYQWIQEYEKYGESAFPGHGSA 60
          M R+ FDK FK SAVKLILEE   VK VS  LE+H NSLY+W+QE E+YGESAFPG+G+A
Sbjct: 1  MARRKFDKQFKNSAVKLILEEGYSVKEVSQELEVHANSLYRWVQEVEEYGESAFPGNGTA 60

Query: 61 LRHAQFKTKKLEKEHKLLQEELALLKKFQVFLKPNR 96
          L +AQ K K LEKE++ LQEEL LLKKF+VFLK ++
Sbjct: 61 LANAQHKIKLLEKENRYLQEELELLKKFRVFLKRSK 96
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2759

A DNA sequence (GASx1374R) was identified in *S.pyogenes* <SEQ ID 8017> which encodes the amino acid sequence <SEQ ID 8018>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2585(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2760

A DNA sequence (GASx1382R) was identified in *S.pyogenes* <SEQ ID 8019> which encodes the amino acid sequence <SEQ ID 8020>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.39   Transmembrane    3 -  19 (   3 -  19)

----- Final Results -----
              bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2761

A DNA sequence (GASx1391R) was identified in *S.pyogenes* <SEQ ID 8021> which encodes the amino acid sequence <SEQ ID 8022>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> May be a lipoprotein

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2762

A DNA sequence (GASx1404) was identified in *S.pyogenes* <SEQ ID 8023> which encodes the amino acid sequence <SEQ ID 8024>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3046(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2763

A DNA sequence (GASx1412R) was identified in *S.pyogenes* <SEQ ID 8025> which encodes the amino acid sequence <SEQ ID 8026>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1590(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2764

A DNA sequence (GASx1414R) was identified in *S.pyogenes* <SEQ ID 8027> which encodes the amino acid sequence <SEQ ID 8028>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2816(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2765

A DNA sequence (GASx1416) was identified in *S.pyogenes* <SEQ ID 8029> which encodes the amino acid sequence <SEQ ID 8030>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1744(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2766

A DNA sequence (GASx1417) was identified in *S.pyogenes* <SEQ ID 8031> which encodes the amino acid sequence <SEQ ID 8032>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3771(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2767

A DNA sequence (GASx1419R) was identified in *S.pyogenes* <SEQ ID 8033> which encodes the amino acid sequence <SEQ ID 8034>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.93   Transmembrane    4 -  20 (   1 -  25)

----- Final Results -----
               bacterial membrane --- Certainty=0.5373(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2768

A DNA sequence (GASx1423) was identified in *S.pyogenes* <SEQ ID 8035> which encodes the amino acid sequence <SEQ ID 8036>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.97   Transmembrane   30 -  46 (  25 -  49)
   INTEGRAL    Likelihood = -7.80   Transmembrane   52 -  68 (  50 -  72)
   INTEGRAL    Likelihood = -6.95   Transmembrane  129 - 145 ( 125 - 146)

----- Final Results -----
               bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2769

A DNA sequence (GASx1426R) was identified in *S.pyogenes* <SEQ ID 8037> which encodes the amino acid sequence <SEQ ID 8038>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
```

```
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -3.45   Transmembrane   36 -  52 (  36 -  55)

----- Final Results -----
               bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC39287 GB:AF115103 orf87 gp [Streptococcus thermophilus
         bacteriophage Sfi21]
 Identities = 43/73 (58%), Positives = 61/73 (82%)

Query: 1  MINLKLRLQNKVTLMAILGAIFLLAQQLGIKLPSNIADIANTAVTLLVLLGVVTDPTTKG 60
          MIN KLRLQNK TL+A++ A+FL+ QQ G+ +P+NI +  NT V +LV+LG++TDPTTKG
Sbjct: 8  MINFKLRLQNKATLVALISAVFLMLQQFGLHVPNNIQEGINTLVGILVILGIITDPTTKG 67

Query: 61 LSDSEQALTYHEP 73
          ++DSE+AL+Y +P
Sbjct: 68 IADSERALSYIQP 80
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2770

A DNA sequence (GASx1427R) was identified in *S.pyogenes* <SEQ ID 8039> which encodes the amino acid sequence <SEQ ID 8040>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.03   Transmembrane    2 -  18 (   1 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2771

A DNA sequence (GASx1428R) was identified in *S.pyogenes* <SEQ ID 8041> which encodes the amino acid sequence <SEQ ID 8042>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1017(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2772

A DNA sequence (GASx1429R) was identified in *S.pyogenes* <SEQ ID 8043> which encodes the amino acid sequence <SEQ ID 8044>. Analysis of this protein sequence reveals the following:

```
Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3097(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2773

A DNA sequence (GASx1431R) was identified in *S.pyogenes* <SEQ ID 8045> which encodes the amino acid sequence <SEQ ID 8046>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2584(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA98101 GB:M19348 hyaluronidase [Streptococcus pyogenes phage
          H4489A]
 Identities = 337/371 (90%), Positives = 351/371 (93%), Gaps = 1/371 (0%)

Query: 1   MAENIPLRVQFKRMKAAEWASSDVVLLEGEIGFETDTGFAKFGDGQNTFSKLKYLTGPKG 60
           M ENIPLRVQFKRM A EWA SDV+LLEGEIGFETDTGFAKFGDGQNTFSKLKYLTGPKG
Sbjct: 1   MTENIPLRVQFKRMSADEWARSDVILLEGEIGFETDTGFAKFGDGQNTFSKLKYLTGPKG 60

Query: 61  PKGDTGLQGKTGGTGSRGPAGKPGTTDYDQLQNKPDLGAFAQKEETNSKITKLESSKADK 120
           PKGDTGLQGKTGGTG RGPAGKPGTTDYDQLQNKPDLGAFAQKEETNSKITKLESSKADK
Sbjct: 61  PKGDTGLQGKTGGTGPRGPAGKPGTTDYDQLQNKPDLGAFAQKEETNSKITKLESSKADK 120

Query: 121 NAVYLKAESNAKLDEKLNLKGGVMTGQLQFKPN-SGIKPSSSVGGAINIDMSKSEGAAMV 179
           +AVY KAES  +LD+KL+L GG++TGQLQFKPN SGIKPSSSVGGAINIDMSKSEGAAMV
Sbjct: 121 SAVYSKAESKIELDKKLSLTGGIVTGQLQFKPNKSGIKPSSSVGGAINIDMSKSEGAAMV 180
```

```
Query: 180 MYTNKDTTDGPLMILRSNKDTFDQSVQFVDYKGTTNAVNIVMRQPTTPNFSSALNITSAN 239
           MYTNKDTTDGPLMILRS+KDTFDQS QFVDY G TNAVNIVMRQP+ PNFSSALNITSAN
Sbjct: 181 MYTNKDTTDGPLMILRSDKDTFDQSAQFVDYSGKTNAVNIVMRQPSAPNFSSALNITSAN 240

Query: 240 EGGSAMQIRGVEKALGTLKITHENPSVDKEYDENAAALSIDIVKKQKGGKGTAAQGIYIN 299
           EGGSAMQIRGVEKALGTLKITHENP+V+ +YDENAAALSIDIVKKQKGGKGTAAQGIYIN
Sbjct: 241 EGGSAMQIRGVEKALGTLKITHENPNVEAKYDENAAALSIDIVKKQKGGKGTAAQGIYIN 300

Query: 300 STSGTAGKMLRIRNKNKDKFYVGPDGDFWSCASSIVDGNLTVKDPTSGKHAATKDYVDEK 359
           STSGTAGKMLRIRNKN+DKFYVGPDG F S A+S V GNLTVKDPTSGKHAATKDYVDEK
Sbjct: 301 STSGTAGKMLRIRNKNEDKFYVGPDGGFHSGANSTVAGNLTVKDPTSGKHAATKDYVDEK 360

Query: 360 IAELKKLILKK 370
           IAELKKLILKK
Sbjct: 361 IAELKKLILKK 371
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2774

A DNA sequence (GASx1438R) was identified in *S.pyogenes* <SEQ ID 8047> which encodes the amino acid sequence <SEQ ID 8048>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence <SEQ ID 10439> was identified in GBS which encodes amino acid sequence <SEQ ID 10440>.

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB18711 GB:U38906 ORF36 [Bacteriophage r1t]
 Identities = 70/111 (63%), Positives = 88/111 (79%)

Query: 1   LIEVIIKKYLDEHLDVPSFFEHQKDEPARFIILEKTSGAKQNHLLSSTFAFQSYAESLYE 60
           +IE+IIK +LD HL V SF E + + P  +I+ EKT  +K NHLLSSTFAFQSYA S+YE
Sbjct: 1   MIEIIIKNFLDTHLSVSSFLEKKGEMPLSYILFEKTGSSKSNHLLSSTFAFQSYAPSMYE 60

Query: 61  AALLNDKVKQVIEQLDVLPQVSGVHLNADYNFTDTATKRYRYQAVFDINHY 111
           AA LN+++K+V+E+L  L ++S V LN+DYNFTDT TK YRYQAVFDINHY
Sbjct: 61  AAKLNEQLKEVVERLIELNEISNVSLNSDYNFTDTETKEYRYQAVFDINHY 111
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2775

A DNA sequence (GASx1442R) was identified in *S.pyogenes* <SEQ ID 8049> which encodes the amino acid sequence <SEQ ID 8050>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1241(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2776

A DNA sequence (GASx1444R) was identified in *S.pyogenes* <SEQ ID 8051> which encodes the amino acid sequence <SEQ ID 8052>. Analysis of this protein sequence reveals the following:

```
Possible site: 42

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4547(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2777

A DNA sequence (GASx1447R) was identified in *S.pyogenes* <SEQ ID 8053> which encodes the amino acid sequence <SEQ ID 8054>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2778

A DNA sequence (GASx1448R) was identified in *S.pyogenes* <SEQ ID 8055> which encodes the amino acid sequence <SEQ ID 8056>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3221(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2779

A DNA sequence (GASx1449R) was identified in *S.pyogenes* <SEQ ID 8057> which encodes the amino acid sequence <SEQ ID 8058>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.6356(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2780

A DNA sequence (GASx1453R) was identified in *S.pyogenes* <SEQ ID 8059> which encodes the amino acid sequence <SEQ ID 8060>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2869(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2781

A DNA sequence (GASx1455R) was identified in *S.pyogenes* <SEQ ID 8061> which encodes the amino acid sequence <SEQ ID 8062>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1787(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF43512 GB:AF145054 ORF19 [Streptococcus thermophilus
          bacteriophage 7201]
 Identities = 47/126 (37%), Positives = 86/126 (67%), Gaps = 2/126 (1%)

Query: 8   LKDLRNLDLYIASLIRRRDKIEASLL--SSPKWSSDKVNGGIKRKQDDVYVELIATAKDI 65
           ++ ++ LD YI S I +  ++E+  L  +S    +D V GG ++ +DD+YVELI   +++
Sbjct: 7   IQQIKALDRYIESQIEQIKRLESQALKVTSGSMHTDMVQGGKRKGKDDIYVELITAREEV 66

Query: 66  EKKTAEAIRKQRELQNLIDSLENTDSQTILSMVYIDKMTRWQVIDELNCSESTYFRLLRV 125
           E+ TAEAI+++ E +  I ++E+ D++++L MVYID+++ WQ+ D++  S++TY+  LR
Sbjct: 67  ERFTAEAIKQKLEFRRQIANIEDIDARSLLQMVYIDQLSIWQICDKMGISKATYYVKLRQ 126

Query: 126 ATKELN 131
           A K L+
Sbjct: 127 AEKYLD 132
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2782

A DNA sequence (GASx1456R) was identified in *S.pyogenes* <SEQ ID 8063> which encodes the amino acid sequence <SEQ ID 8064>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2883(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB18697 GB:U38906 ORF22 [Bacteriophage r1t]
 Identities = 78/207 (37%), Positives = 123/207 (58%), Gaps = 2/207 (0%)

Query: 6   EIHRILGIDEVYKAPKRLTDILFDKDSREDIFRQFLKYETDVSYDWFMQYFEEEQADRKN 65
           + + +L +DE      R+ +++FDK  RE+ + + L    D+  D+F  YF    A
Sbjct: 7   QFYDMLNVDEHMNFTNRIQELVFDKKGREEFYSKILNIHHDMGVDFFRDYFMAHSAVSA- 65

Query: 66  KKQDFTPKSVSTLLSKIISGNQYYEVA-VGTGGILIQAWQEQRLNDSPFTYRPSKYWYHV 124
           K Q +TP  +  L + ++ G+   ++   GTG ++IQ WQ+ R+N   F Y PS YWY
Sbjct: 66  KGQHYTPDELGKLTALLVGGSGGADLTGAGTGTLIIQKWQDDRMNTDFFNYLPSNYWYQA 125

Query: 125 EELSDKAVPFLLFNMSIRGINGVVVHGDSLTRQVKNIYFLQNTKDDMLSFSDINVMPRTQ 184
            ELSD+A+ FL+   +IRG+NGVV+HGD+L   VK +YF+QN+ ++ + FS+INV+P ++
Sbjct: 126 LELSDEAISFLIHAFAIRGMNGVVIHGDALEMAVKQVYFIQNSANNPIGFSEINVIPHSK 185
```

```
Query: 185 DIEREFNVKEWIGDGIEHIENPLIEWI 211
           D      + EW    IEHIE+   +WI
Sbjct: 186 DAMEFLGIHEWTEQAIEHIESKFPDWI 212
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2783

A DNA sequence (GASx1459R) was identified in *S.pyogenes* <SEQ ID 8065> which encodes the amino acid sequence <SEQ ID 8066>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.44   Transmembrane   82 -  98 (  81 -  98)

----- Final Results -----
               bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2784

A DNA sequence (GASx1460R) was identified in *S.pyogenes* <SEQ ID 8067> which encodes the amino acid sequence <SEQ ID 8068>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3368(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2785

A DNA sequence (GASx1461R) was identified in *S.pyogenes* <SEQ ID 8069> which encodes the amino acid sequence <SEQ ID 8070>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2834(Affirmative) < succ>
```

```
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2786

A DNA sequence (GASx1462R) was identified in *S.pyogenes* <SEQ ID 8071> which encodes the amino acid sequence <SEQ ID 8072>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3531(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2787

A DNA sequence (GASx1463R) was identified in *S.pyogenes* <SEQ ID 8073> which encodes the amino acid sequence <SEQ ID 8074>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2483(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14569 GB:Z99117 similar to phage-related protein [Bacillus subtilis]
 Identities = 98/252 (38%), Positives = 152/252 (59%), Gaps = 29/252 (11%)

Query: 16  SPAVKNRIEQVVGARAEQFTTSLLSIISNNNLLAKATSESIMGAAMKAAVLNLPIEPSLG 75
           SP+V  R E+V+G RA QFT S+LS+ ++  +L K    S++ +AM AA L+LPI+ +LG
Sbjct: 33  SPSVIKRFEEVLGKRATQFTASILSLYNSEQMLQKTDPMSVISSAMVAATLDLPIDKNLG 92

Query: 76  FAYVVPYNRNYKDGNRWITVNEAQFQIGYRGLIQLAQRSGQVRNIEHGIIYEEEFLGYDK 135
           +A++VPY              +AQFQ+GY+G IQLA R+GQ ++I    I+E E   ++
Sbjct: 93  YAWIVPYG------------GKAQFQLGYKGYIQLALRTGQYKSINCIPIHEGELQKWNP 140

Query: 136 IRGQLKLTGDYVDSGVVKGYFASLELISGFYKMIFWPKEKVYEHAKKYSKTFDKKTGDFK 195
           +  ++++  +  +S  V GY A  ELI+GF K ++W K +V +H KK+SK+       DF
Sbjct: 141 LTEEIEIDFEKRESDAVIGYAAYFELINGFRKTVYWTKAQVEKHKKKFSKS------DF- 193
```

```
Query: 196 PGTPWATEFDPMAIKTLLKELLSKYAPLSVEMQDA-LEADNADSTIVIPKDVTPQETNSL 254
                W  ++D MA+KT+LK +LSK+  LSVEMQ A +E D     I    D+T +  +S
Sbjct: 194 ---GWKNDWDAMALKTVLKAVLSKWGILSVEMQKAVIEEDETRERI----DITNEADSS- 245

Query: 255 DDLIGTQNEKKD 266
           ++I ++   KD
Sbjct: 246 -EIIDSEPSNKD 256
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2788

A DNA sequence (GASx1464R) was identified in *S.pyogenes* <SEQ ID 8075> which encodes the amino acid sequence <SEQ ID 8076>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4258(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2789

A DNA sequence (GASx1465R) was identified in *S.pyogenes* <SEQ ID 8077> which encodes the amino acid sequence <SEQ ID 8078>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2045(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2790

A DNA sequence (GASx1469R) was identified in *S.pyogenes* <SEQ ID 8079> which encodes the amino acid sequence <SEQ ID 8080>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
```

```
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2791

A DNA sequence (GASx1470R) was identified in *S.pyogenes* <SEQ ID 8081> which encodes the amino acid sequence <SEQ ID 8082>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3577(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC98430 GB:L29324 excisionase [Streptococcus pneumoniae]
 Identities = 23/56 (41%), Positives = 41/56 (73%)

Query: 23 KHLIQQWEGLTVATAKQWATEMRDHPDFKQFVLNPTHRIVFIDYKGFKLFVQWKSR 78
          K ++++W+GL   T  +W  EMR++  F  +V+NPTH++VFI+ +GF+ F++WK +
Sbjct: 19 KGILKRWDGLNKYTLNRWIKEMRENRTFSMYVINPTHKLVFINLEGFESFLRWKQK 74
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2792

A DNA sequence (GASx1473) was identified in *S.pyogenes* <SEQ ID 8083> which encodes the amino acid sequence <SEQ ID 8084>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2725(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2793

A DNA sequence (GASx1476) was identified in *S.pyogenes* <SEQ ID 8085> which encodes the amino acid sequence <SEQ ID 8086>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1422(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2794

A DNA sequence (GASx1480R) was identified in *S.pyogenes* <SEQ ID 8087> which encodes the amino acid sequence <SEQ ID 8088>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -4.04   Transmembrane  291 - 307 ( 290 - 309)

----- Final Results -----
              bacterial membrane --- Certainty=0.2614(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2795

A DNA sequence (GASx1489R) was identified in *S.pyogenes* <SEQ ID 8089> which encodes the amino acid sequence <SEQ ID 8090>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2278(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2796

A DNA sequence (GASx1490R) was identified in *S.pyogenes* <SEQ ID 8091> which encodes the amino acid sequence <SEQ ID 8092>:

```
SFITSVLAFRKLLKCEGIDLYLMYGDLMTCFEQLLTQLKDWTDVYFNYDESGYGRLRDQKAAQFFKKNGIAVHTYQDHYLHGSQEIINQSG
QPYKVFTPYYRIWQNYPKETPIKVELSQGRWLNLETPDDVLRTVESFKDEKYQDVATFDEASKQLNRFIQDQLAAYHANRDFPAQLGTSRL
SPFLRIGAIGIRTVYHAVRQAPNSLGQATFLKELAWRDFYNMVYVAYPDQKTQPIQKAFSQIEWVNNPDWFQLWKEGKTGYPIVDAAMLQL
QKTGWMHNRLRMIVASFLTKDLLCDWRLGEQYFQQQLIDYDAASNIGGWQWAASTGTDAVPYFRIFNPVTQGKRFDPKGEFIKAYLPQLEH
VPEKYLHEPWKMPKNLQESVSCIIGTDYPQPIVDHAKQREQAIAKYEWAKEKAKIE
```

Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA22361 GB:M94110 DNA photolyase [Bacillus firmus]
 Identities = 175/338 (51%), Positives = 228/338 (66%), Gaps = 6/338 (1%)

Query: 145 EIINQSGQPYKVFTPYYRIWQNYPKETP--IKVELSQGRWLNLETPDDVLRTVES--FKD 200
           +++ + G PYKVFTPYY+ W    K TP  IK ++  G       PD    T+ +   K
Sbjct: 2   QVLKKDGTPYKVFTPYYKAWAKERKRTPAVIKRDVLLGSVHKGTAPDREAETLFNNLIKK 61

Query: 201 EKYQDVATFDE-ASKQLNRFIQDQLAAYHANRDFPAQLGTSRLSPFLRIGAIGIRTVY-H 258
             Y   A  +E A K+L  F + +L+ Y ANRDFP+  GTSRLSP+++ GA+  R++Y H
Sbjct: 62  CSYDWSAIGEEHAIKRLQMFTKKRLSGYKANRDFPSITGTSRLSPYIKTGAVSSRSIYYH 121

Query: 259 AVRQAPNSLGQATFLKELAWRDFYNMVYVAYPDQKTQPIQKAFSQIEWVNNPDWFQLWKE 318
            +    +S    TFLKELAWRDFY MV+   PD K + I + + ++ W ++ D    WK
Sbjct: 122 ILNAEADSYSAETFLKELAWRDFYRMVHFYEPDCKDREIMEGYRELNWSHDQDDLTSWKR 181

Query: 319 GKTGYPIVDAAMLQLQKTGWMHNRLRMIVASFLTKDLLCDWRLGEQYFQQQLIDYDAASN 378
           G+TG+PIVDA M QL   GWMHNRLRMI ASFLTKDLL DWRLGE+YF++ LIDYD +SN
Sbjct: 182 GETGFPIVDAGMRQLLNEGWMHNRLRMITASFLTKDLLIDWRLGERYFERMLIDYDPSSN 241

Query: 379 IGGWQWAASTGTDAVPYFRIFNPVTQGKRFDPKGEFIKAYLPQLEHVPEKYLHEPWKMPK 438
           IGGWQWAAS GTDAVPYFRIFNPVTQ KRFD  G +I+ Y+P+L HVP+ Y+HEPWKM +
Sbjct: 242 IGGWQWAASVGTDAVPYFRIFNPVTQSKRFDENGTYIRTYIPELNHVPDHYIHEPWKMSE 301

Query: 439 NLQESVSCIIGTDYPQPIVDHAKQREQAIAKYEWAKEK 476
             Q    C +  DYP PIVDH+KQR++A++ ++   E+
Sbjct: 302 EEQVKYKCRLDEDYPLPIVDHSKQRKKALSFFKGDDEE 339
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2797

A DNA sequence (GASx1493R) was identified in *S.pyogenes* <SEQ ID 8093> which encodes the amino acid sequence <SEQ ID 8094>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2748(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2798

A DNA sequence (GASx1501R) was identified in *S.pyogenes* <SEQ ID 8095> which encodes the amino acid sequence <SEQ ID 8096>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.27   Transmembrane   64 -  80 (  53 -  83)

----- Final Results -----
             bacterial membrane --- Certainty=0.3909(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC95443 GB:AF068901 YlmG [Streptococcus pneumoniae]
 Identities = 35/81 (43%), Positives = 58/81 (71%)

Query: 1  MILILSILLRLIKVYTYLLIAYALMSWFPGAYDSKIGRLISGIVEPILKPFRAFNLQFAG 60
          MI ++ ++    + +Y+ +L+A+A+MSWFPGAY+S +GR I  +V+P+L P +   LQ AG
Sbjct: 1  MIFLIRMIYNAVDIYSLILVAFAVMSWFPGAYESSLGRWIVALVKPVLAPLQRLPLQIAG 60

Query: 61 LDFTIFVVIISLNFLAQVLVR 81
          LD +++V I+ + FL + LVR
Sbjct: 61 LDLSVWVAIVLVRFLGENLVR 81
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2799

A DNA sequence (GASx1502) was identified in *S.pyogenes* <SEQ ID 8097> which encodes the amino acid sequence <SEQ ID 8098>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.39   Transmembrane   17 -  33 (  17 -  33)

----- Final Results -----
             bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2800

A DNA sequence (GASx1507) was identified in *S.pyogenes* <SEQ ID 8099> which encodes the amino acid sequence <SEQ ID 8100>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0865(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2801

A DNA sequence (GASx1511R) was identified in *S.pyogenes* <SEQ ID 8101> which encodes the amino acid sequence <SEQ ID 8102>. Analysis of this protein sequence reveals the following:

```
Possible site: 47

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.83   Transmembrane   31 -  47 (  22 -  53)
   INTEGRAL    Likelihood = -0.96   Transmembrane    2 -  18 (   1 -  18)

----- Final Results -----
              bacterial membrane --- Certainty=0.5734(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2802

A DNA sequence (GASx1516R) was identified in *S.pyogenes* <SEQ ID 8103> which encodes the amino acid sequence <SEQ ID 8104>. Analysis of this protein sequence reveals the following:

```
Possible site: 42

>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2729(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA96472 GB:AB036428 Dpr [Streptococcus mutans]
 Identities = 132/175 (75%), Positives = 153/175 (87%)

Query: 1   MTNTLVENIYASVTHNISKKEASKNEKTKAVLNQAVADLSVAASIVHQVHWYMRGPGFLY 60
           MTNT+ ENIYAS+ H + KKE S NEKTKAVLNQAVADLS AASIVHQVHWYMRG GFLY
Sbjct: 1   MTNTITENIYASIIHQVEKKENSGNEKTKAVLNQAVADLSKAASIVHQVHWYMRGSGFLY 60

Query: 61  LHPKMDELLDSLNANLDEMSERLITIGGAPYSTLAEFSKHSKLDEAKGTYDKTVAQHLAR 120
           LHPKMDEL+D+LN +LDE+SERLITIGGAP+STL EF ++S+L+E  GT+DK++  HL R
Sbjct: 61  LHPKMDELMDALNGHLDEISERLITIGGAPFSTLKEFDENSRLEETVGTWDKSITDHLKR 120

Query: 121 LVEVYLYLSSLYQVGLDITDEEGDAGTNDLFTAAKTEAEKTIWMLQAERGQGPAL 175
           LV+VY YLSSLYQVGLD+TDEE DA +ND+FTAA+TEA+KTIWMLQAE GQ P L
Sbjct: 121 LVQVYDYLSSLYQVGLDVTDEEDDAVSNDIFTAAQTEAQKTIWMLQAELGQAPGL 175
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2803

A DNA sequence (GASx1517) was identified in *S.pyogenes* <SEQ ID 8105> which encodes the amino acid sequence <SEQ ID 8106>. Analysis of this protein sequence reveals the following:

```
Possible site: 46

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.32   Transmembrane  109 - 125 ( 106 - 126)
   INTEGRAL    Likelihood = -5.26   Transmembrane   63 -  79 (  61 -  81)
   INTEGRAL    Likelihood = -5.20   Transmembrane  154 - 170 ( 151 - 176)
   INTEGRAL    Likelihood = -4.14   Transmembrane  189 - 205 ( 189 - 205)
   INTEGRAL    Likelihood = -3.50   Transmembrane  130 - 146 ( 127 - 147)
   INTEGRAL    Likelihood = -2.92   Transmembrane    6 -  22 (   1 -  24)
   INTEGRAL    Likelihood = -2.23   Transmembrane   83 -  99 (  83 - 101)

----- Final Results -----
             bacterial membrane --- Certainty=0.3527(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA96471 GB:AB036428 type IV prepilin peptidase homologue
          [Streptococcus mutans]
 Identities = 55/127 (43%), Positives = 78/127 (61%), Gaps = 3/127 (2%)

Query: 83  VSASYCYLLLFSLLFSLFDWRSQEYPFILWLFSFVSLLLFYSINYLSLILLLLGLLAHLR 142
           ++ S   LL   +L SL+D + Q YP  LW+     L+  Y +N +SLIL L G+ A L+
Sbjct: 91  LTTSQVCLLFMGVLLSLYDLQDQSYPLTLWIGFTFLLMFIYPLNLISLILFLFGIFAALK 150

Query: 143 PFSIGAGDFFYLASLALVLDLTSLIWLIQLASLAGITACLLLGIKRIP--FIPYLSFGLF 200
             +IG+GDFFYLA+LAL L+L  +IW+IQ+ASL GI   LL    + P  F+P+L  G
Sbjct: 151 NINIGSGDFFYLATLALSLNLQQIIWIIQIASLLGILYSLLFQKHKEPFAFVPFLFLG-H 209

Query: 201 WIVLLEH 207
```

```
            I++  H
Sbjct: 210 LIIIFSH 216
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2804

A DNA sequence (GASx1538R) was identified in *S.pyogenes* <SEQ ID 8107> which encodes the amino acid sequence <SEQ ID 8108>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1186(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2805

A DNA sequence (GASx1539R) was identified in *S.pyogenes* <SEQ ID 8109> which encodes the amino acid sequence <SEQ ID 8110>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.73   Transmembrane    6 -  22 (   3 -  32)

----- Final Results -----
              bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF31453 GB:AF221126 putative histidine kinase [Streptococcus pneumoniae]
 Identities = 141/301 (46%), Positives = 210/301 (68%), Gaps = 7/301 (2%)

Query: 1   MKRYPLLVQLISYVFVIVIALITTLGLLYYQTSSRNIRQLIERDTRQSIRQSSQFIDAYI 60
            MKR  LLV+++  +F++ + L+  +G  YYQ+SS  I   IE +++ +I Q+S FI +YI
Sbjct: 1   MKRSSLLVRMVISIFLVFLILLALVGTFYYQSSSSAIEATIEGNSQTTISQTSHFIQSYI 60

Query: 61  KPLKETTSVLAKNTEIQAFASQIHQENDKQVLQLMKMVLATNSDLQAAVLVTKDGRTVST 120
            K L+ T++ L + T++ A+A    Q+  + +  L   +L ++ DL+  VLVTK G+ +ST
Sbjct: 61  KKLETTSTGLTQQTDVLAYAENPSQDKVEGIRDLFLTILKSDKDLKTVVLVTKSGQVIST 120

Query: 121 NSQLTMKTSSDMMAEPWYKAAIDRQAMPILTPARQLSLSSKKEWVVSVTQEVVDRAGHNL 180
            +  + MKTSSDMMAE WY+ AI + AMP+LTPAR+     S  +WV+SVTQE+VD  G NL
Sbjct: 121 DDSVQMKTSSDMMAEDWYQKAIHQGAMPVLTPARK----SDSQWVISVTQELVDAKGANL 176

Query: 181 GVLRLDIAYPTIKASLDQLQLGRQGFAFIVNDKHEFVYHPKKSVYSSSKEMAAMKPYLAI 240
            GVLRLDI+Y T++A L+QLQLG+QGFAFI+N+ HEFVYHP+ +VYSSS +M AMKPY+
```

```
Sbjct: 177 GVLRLDISYETLEAYLNQLQLGQQGFAFIINENHEFVYHPQHTVYSSSSKMEAMKPYIDT 236

Query: 241 QNGYTKDKTSFVYQKLIPNSQWTLVGVASLDQLHRVQRQIFWSFSWNRASTLSDLWLCNCL 301
             GYT    S+V Q+ I  + WT++GV+SL++L +V+ Q+ W+      ++++ L +C CL
Sbjct: 237 GQGYTPGHKSYVSQEKIAGTDWTVLGVSSLEKLDQVRSQLLWTL---LGASVTSLLVCLCL 294
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2806

A DNA sequence (GASx1542R) was identified in *S.pyogenes* <SEQ ID 8111> which encodes the amino acid sequence <SEQ ID 8112>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC23101 GB:U32823 conserved hypothetical protein [Haemophilus influenzae Rd]
 Identities = 56/128 (43%), Positives = 87/128 (67%)

Query: 73  DFELKGIDGKTYRLSEFKGKKVYLKFWASWCSICLSTLADTEDLAKMSDKDYVVLTVVSP 132
           D +LK ++ +   LS++KGK VY+K WASWC ICL+ LA+ +DL+   D+++ V+T+VSP
Sbjct: 24  DVQLKDLNNQPVTLSQYKGKPVYVKMWASWCPICLAGLAEIDDLSAEKDRNFEVITIVSP 83

Query: 133 GHQGEKSEADFKKWFQGTDYKDLPVLLDPDGKLLEAYGVRSYPTEVFIGSDGVLAKKHIG 192
            H+GEK  ADF +W++G +YK++ VLLD  G++++   VR YP  +F+ SD  L K   G
Sbjct: 84  DHKGEKDTADFIEWYKGLEYKNITVLLDEKGEIIDKARVRGYPFNLFLDSDLNLKKTVPG 143

Query: 193 YAKKSDIK 200
           +     I+
Sbjct: 144 HLGAEQIR 151
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2807

A DNA sequence (GASx1543R) was identified in *S.pyogenes* <SEQ ID 8113> which encodes the amino acid sequence <SEQ ID 8114>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.75   Transmembrane  171 - 187 ( 169 - 191)
   INTEGRAL    Likelihood = -6.26   Transmembrane  205 - 221 ( 203 - 232)
   INTEGRAL    Likelihood = -5.73   Transmembrane   56 -  72 (  54 -  81)
   INTEGRAL    Likelihood = -5.36   Transmembrane   92 - 108 (  91 - 113)
   INTEGRAL    Likelihood = -3.45   Transmembrane   20 -  36 (  14 -  39)
   INTEGRAL    Likelihood = -1.17   Transmembrane  147 - 163 ( 144 - 163)

----- Final Results -----
               bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC23102 GB:U32823 cytochrome C-type biogenesis protein
          [Haemophilus influenzae Rd]
 Identities = 106/224 (47%), Positives = 138/224 (61%), Gaps = 16/224 (7%)

Query: 6   VLMVSVFGAGLLSFFSPCIFPVLPVYLGILLDADDSKTITIFGKKLYWYGIVKTLAFIFG 65
           +L+ +VF AGL SF SPCIFP++P+Y GIL            GKK     ++ T  FI G
Sbjct: 6   LLIGTVFLAGLASFLSPCIFPIIPIYFGILSKG---------GKK-----VLNTFLFILG 51

Query: 66  LSTIFVILGYGAGFLGNILYAVWFRYLLGALVIILGIHQMGLITIKSLQFQKSLTFHNNK 125
           LS  FV LG+  GFLGNIL++   R + G +VIILGIHQ+G+  I  L+  K +    +
Sbjct: 52  LSLTFVSLGFSFGFLGNILFSNTTRIIAGVIVIILGIHQLGIFKIGLLERTKLVEIKTSG 111

Query: 126 NRNGLFNAFILGLTFSFGWTPCVGPVLSSVLALVASGGNGAWQGGVLMIIYTLGLGIPFL 185
               L  AF+LGLTFS GWTPC+GP+L+SVLAL    G+ A  G  +M +Y LGL  PF+
Sbjct: 112 KSTAL-EAFVLGLTFSLGWTPCIGPILASVLALSGDEGS-ALYGASMMFVYVLGLATPFV 169

Query: 186 LISFASGIVLKQFNKLKPHILLLKKVGGVLIIVMGILLMTGTLN 229
           L SF S  +LK+   L  H+   K  GG+LIIVMGILL+T   +
Sbjct: 170 LFSFFSDSLLKRAKGLNKHLDKFKIGGGILIIVMGILLITNNFS 213
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2808

A DNA sequence (GASx1544) was identified in *S.pyogenes* <SEQ ID 8115> which encodes the amino acid sequence <SEQ ID 8116>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1493(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2809

A DNA sequence (GASx1546R) was identified in *S.pyogenes* <SEQ ID 8117> which encodes the amino acid sequence <SEQ ID 8118>. Analysis of this protein sequence reveals the following:

```
Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4658(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04061 GB:AP001508 unknown conserved protein in others
        [Bacillus halodurans]
 Identities = 48/89 (53%), Positives = 61/89 (67%)

Query: 1  MMVLVTYDVNTETPAGRKRLRHVAKLCVDYGQRVQNSVFECSVTPAEFVDIKHRLTQIID 60
          M+VL+TYDV T +  G KRLR VAK C +YGQRVQNSVFEC V   +   +K  LT +ID
Sbjct: 1  MLVLITYDVQTSSMGGTKRLRKVAKACQNYGQRVQNSVFECIVDSTQLTSLKLELTSLID 60

Query: 61 EKTDSIRFYLLGKNWQRRVETLGRSDSYD 89
          E+ DS+R Y LG N++ +VE +G   S D
Sbjct: 61 EEKDSLRIYRLGNNYKTKVEHIGAKPSID 89
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2810

A DNA sequence (GASx1547R) was identified in *S.pyogenes* <SEQ ID 8119> which encodes the amino acid sequence <SEQ ID 8120>. Analysis of this protein sequence reveals the following:

```
Possible site: 57

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.70   Transmembrane   44 -  60 (  43 -  60)

----- Final Results -----
               bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 330-332
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04060 GB:AP001508 unknown conserved protein in others
        [Bacillus halodurans]
 Identities = 162/341 (47%), Positives = 231/341 (67%), Gaps = 1/341 (0%)

Query: 1   MKKLLNTLYLTQEDFYVTKEGDNIVIKQEGKVLKRFPFRIIDGIVCFSYLGVSSALVKLC 60
           MKKLLNTLY+TQ D Y++ +GDN+V+ +E + L R P   ++ IV F Y G S AL+  C
Sbjct: 1   MKKLLNTLYVTQPDTYLSLDGDNVVLLKEQEKLGRLPLHNLEAIVGFGYT>FEATURESALMGYC 60

Query: 61  TENQINLSFHTPQGRFCGRYIGSTNGNVLLRREHYRLSDRE-ESLEYAKRFILAKISNSR 119
            E  I+++F T  GRF  R +G + GNV+LR+  YR+S+ + ES + A+ FI  K+ NS+
Sbjct: 61  AERNISITFLTKNGRFLARVVGESRGNVVLRKTQYRISENDQESTKIARNFITGKVYNSK 120

Query: 120 KYLLRFKRDHRQQIDTKLFEAVNDELIWALEMVQAADNKDSLRGIEGQAANQYFRIFNDL 179
             L R  R+H  +++ + F+A +  L   ++ ++  D+ +SLRG EGQAA  Y ++F+ +
Sbjct: 121 WMLERMTREHPLRVNVEQFKATSQLLSVMMQEIRNCDSLESLRGWEGQAAINYNKVFDQM 180

Query: 180 VLTDKKTFYFQGRSKRPPLDCVNALLSFGYSLLTFECQSALEAVGLDSYVGFFHTDRPGR 239
           +L  K+ F F GRS+RPP D VNA+LSF Y+LL  +  +ALE VGLD+YVGF H DRPGR
Sbjct: 181 ILQQKEEFAFHGRSRRPPKDNVNAMLSFAYTLLANDVAAALETVGLDAYVGFMHQDRPGR 240

Query: 240 ASLALDLVEEFRSYIVDRFVFSLINKGQLQKKHFEVKENGSILLTENGRAIFIDLWQKRK 299
           ASLALDL+EE R    DRFV SLIN+ ++    F  KENG++L+T+  R  F+  WQ +K
Sbjct: 241 ASLALDLMEELRGLYADRFVLSLINRKEMTADGFYKKENGAVLMTDEARKTFLKAWQTKK 300

Query: 300 HTEVEHPFTKEKVKLMLLPYVQAQLLAKAIRGDLESYPPFM 340
             ++ HP+  EK+   L+PYVQA LLA+ +RGDL+ YPPF+
Sbjct: 301 QEKITHPYLGEKMSWGLVPYVQALLLARFLRGDLDEYPPFL 341
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2811

A DNA sequence (GASx1548R) was identified in *S.pyogenes* <SEQ ID 8121> which encodes the amino acid sequence <SEQ ID 8122>. Analysis of this protein sequence reveals the following:

```
Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2247(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04059 GB:AP001508 unknown [Bacillus halodurans]
 Identities = 90/169 (53%), Positives = 111/169 (65%), Gaps = 1/169 (0%)

Query: 45  LHTKADNPYIKEKRKELLVSRAMPISSAELGLSGIMDVVEFYKDDQGVSLRGKRGKWLPK 104
           +H KAD P++KEKR   L  RAMPI S  L +SGI DVVEF +D +G+ L G  G +
Sbjct: 1   MHKKADQPFMKEKRGSKLTVRAMPIQSKNLQISGICDVVEFVQDSEGIELSGVSGSYKAF 60

Query: 105 VVEYKRGKPKKDTRDIVQLVAQTMCLEETLDCDINEGCLYYHSVNQRVIVPMTSALRQEV 164
            VEYKRGKPKK   DIVQLVAQ MCLEE L C I++G L+Y+ +  RV VP+T ALR +V
Sbjct: 61  PVEYKRGKPKKGDEDIVQLVAQAMCLEEMLVCRIDKGYLFYNEIKHRVEVPITDALRDKV 120

Query: 165 KELAAEMHEVYQSQMLPKAAYFKNCQLCSLVDICKPRLSKKTRSVSRYI 213
           ++A EMH  Y+++  PK      C  CSL  IC P+L  K RSV RYI
Sbjct: 121 VQMAKEMHHYYENRHTPKVKTGPFCNNCSLQSICLPKLMNK-RSVKRYI 168
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2812

A DNA sequence (GASx1549R) was identified in *S.pyogenes* <SEQ ID 8123> which encodes the amino acid sequence <SEQ ID 8124>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1399(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04058 GB:AP001508 unknown conserved protein in others
          [Bacillus halodurans]
 Identities = 148/290 (51%), Positives = 190/290 (65%), Gaps = 19/290 (6%)

Query: 6   MLEHKIDFMVTLEVKEANANGDPLNGNMPRTDAKGYGVMSDVSIKRKIRNRLQDMGKSIF 65
```

```
            +L+HKIDF V L V +AN NGDPLNGN PR +  G+G +SDV+IKRKIRNRL DM + IF
Sbjct: 3    ILDHKIDFAVILSVTKANPNGDPLNGNRPRQNYDGHGEISDVAIKRKIRNRLLDMEEPIF 62

Query: 66   VQANERIEDDFRSLEKRFSQH----FTAKTPDKEIEEKANAL---WFDVRAFGQVFTYLK 118
            VQ+++R  D F+SL  R   +       K  +  ++E A      W DVR+FGQVF +
Sbjct: 63   VQSDDRKADSFKSLRDRADSNPELAKMLKAKNASVDEFAKIACQEWMDVRSFGQVFAFKG 122

Query: 119  K--SIGVRGPVSISMAKSLEPIVISSLQITRSTNGMEAKNNSGRSSDTMGTKHFVDYGVY 176
               S+GVRGPVSI  A S++PI I S QIT+S N +       RSSDTMG KH VD+GVY
Sbjct: 123  SNLSVGVRGPVSIHTATSIDPIDIVSTQITKSVNSVTGDK---RSSDTMGMKHRVDFGVY 179

Query: 177  VLKGSINAYFAEKTGFSQEDAEAIKEVLVSLFENDASSARPEGSMRVCEVFWFTHSSKLG 236
            V KGSIN   AEKTGF+ EDAE IK  L++LFEND+SSARP+GSM V +V+W+ HSSKLG
Sbjct: 180  VFKGSINTQLAEKTGFTNEDAEKIKRALITLFENDSSSARPDGSMEVHKVYWWEHSSKLG 239

Query: 237  NVSSARVFDLLEYHQSIEEKSTYDAYQIHLNQEKLAKYEAKGLTLEILEG 286
              SSA+V   L+     +  ++D Y + L       YE  GL +E+++G
Sbjct: 240  QYSSAKVHRSLKIESKTDTPKSFDDYAVEL-------YELDGLGVEVIDG 282
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2813

A DNA sequence (GASx1550R) was identified in *S.pyogenes* <SEQ ID 8125> which encodes the amino acid sequence <SEQ ID 8126>. Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2882(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04057 GB:AP001508 unknown [Bacillus halodurans]
 Identities = 176/671 (26%), Positives = 311/671 (46%), Gaps = 87/671 (12%)

Query: 1    MDFFTSLLKTYEKAELADLVDHQKR--NNEPVLLPIYHTSLKSNGKNIISVKLDKDGQFH 58
            M +   L +TYE A L  +    K+  + E  LLPI HT+  ++    I V LD+DG F
Sbjct: 1    MSWLLHLYETYE-ANLDQVGKTVKKGEDREYTLLPISHTTQNAH----IEVTLDEDGDFL 55

Query: 59   KAEFMADKQMIIFPVTADSVARSGSHPAPHPLVDKFAYYSAEM----GQIQ------YDS 108
            +A+ +  K+  + P T ++ +RSGS  AP+PL DK +Y + +      G+I+       +D+
Sbjct: 56   RAKALT-KESTLIPCTEEAASRSGSKVAPYPLHDKLSYVAGDFVKYGGKIKNQDDAPFDT 114

Query: 109  FHKQLNNWID--YCEEGDVKKFLTFVQQFILKPEFLTLILDSLIGPDYQHNQLKVTFCDA 166
            + K L  W +  Y  E  VK   T++++  L + +  L       NQ  +   +
Sbjct: 115  YIKNLGEWANSPYATE-KVKCIYTYLKKGRLIEDLVDAGVLKL-----DENQQLIEKWEK 168

Query: 167  TGKEKLIDLSACFLEFSIDQ------FQGFKNESVSTF---KALHQSYISFVEANRENLG 217
              +E L +  A F   + DQ      F  F  ES+       K +  S+ISF
Sbjct: 169  RYEELLGEKPAIFSSGATDQASAFVRFNVFHPESIDDVWKDKEMFDSFISFYNDKLGEED 228

Query: 218  ICNISGREEQLTDKH----RGLMGNAKIISVS-NKREAYKGRFRERDVFSVGYETSEKI 272
            IC ++G     T++H     R     AK+IS + N     ++GRF+   +   + YE S+K
Sbjct: 229  ICFVTGNRLPSTERHANKIRHAADKAKLISANDNSGFTFRGRFKTSREAVGISYEVSQKA 288

Query: 273  HLMLKYLLENKNTSTWLGSSQYLINWFSDD-LTNDSRLDIVSPIFDDGLEEDDDDDTPPV 331
            H  LK+L+  ++ S      +  + W +D+ L  +   D V  +     E + D DT  +
Sbjct: 289  HNALKWLIHRQSKSI---DDRVFLVWSNDNSLVPNPDEDAVDIMKHANRELERDPDTGQI 345
```

```
Query: 332 ITLATEDNKRIGKSFIKGQKLFANDATY----YVAILNKTSNGRIALKYFRQLQASQLLT 387
             A E  K IG          + +D  Y    ++ +L+  + GR+A+ Y+R L     L
Sbjct: 346 F--AGEVKKAIGG--------YRSDLNYQPEVHILVLDSATTGRMAVLYYRSLNKELYLN 395

Query: 388 NLNKWQETYSWESRSKFGKSRLRT----PTFHDILNVSYGVDRDRFLELDNDNFKSDQIQ 443
            L  W ++ +WE R +  +    +    P   DI   +YG          ++    D ++
Sbjct: 396 RLEAWHDSCAWEHRYRRDEKEFISFYGAPATKDIAFAAYGPRA-------SEKVIKDLME 448

Query: 444 KLVASLIDGKPMPQSIVKKL---GNNVKERHRYRKHWYQVEQVCLAILHK---QNGEEFS 497
            +++  ++DG+ +P+ IV+      +N     R+   W +   + A++ K   +  EE+
Sbjct: 449 RMLPCIVDGRRVPKDIVRSAFQRASNPVSMERWE--WEKTLSITCALIRKMHIEQKEEWG 506

Query: 498 PMLDHTNQNRSYLFGRLLAIFELIETLRYGLDGNNNDRITNAERYWTAYTGQPTKLMMLL 557
             LD ++ +RSYLFGRLLA+ +++E    G  G +  R TNA RY  +Y+  P +    +
Sbjct: 507 VPLDKSSTDRSYLFGRLLAVADVLER---GALGKDETRATNAIRYMNSYSKNPGRTWKTI 563

Query: 558 ENKIKPYEEPLKLNRRGSWMKLEKEKEEILELLNPLLETETMEKPLDYRFIFGYYAEKNY 617
            +  ++PY+    KL  + ++  L K  +EI +   P     +   PL  +++ G+Y+++
Sbjct: 564 QESLQPYQ--AKLGTKATY--LSKLVDEIGDQFEP---GDFNNNPLTEQYLLGFYSQRRE 616

Query: 618 YYTKQNTEVTE 628
            Y K+  E  +
Sbjct: 617 LYKKKEEETNQ 627
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2814

A DNA sequence (GASx1551R) was identified in *S.pyogenes* <SEQ ID 8127> which encodes the amino acid sequence <SEQ ID 8128>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3035(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04056 GB:AP001508 unknown [Bacillus halodurans]
 Identities = 90/218 (41%), Positives = 127/218 (57%), Gaps = 7/218 (3%)

Query: 13  GQRALFTNPATKGGSERSSYSVPTRQALNGIVDAIYYKPTFTNIVTEVKVINQIQTELQG 72
           G  ALFT+P TK G E+ SYSVPT QAL GI ++IY+KPT   ++ E++V+  IQ E +G
Sbjct: 11  GDYALFTDPLTKIGGEKLSYSVPTYQALKGIAESIYWKPTIVFVIDELRVMKPIQMESKG 70

Query: 73  VRALLHDYSADLSYVSYLSDVVYLIKFHFVWNEDRKDLNSDRLPAKHEAIMERSIRKGGR 132
           VR + +     L++ +YL DV Y +K HF +N  R DL  DR   KH +I++RS++ GGR
Sbjct: 71  VRPIEYGGGNTLAHYTYLKDVHYQVKAHFEFNLHRPDLAFDRNEGKHYSILQRSLKAGGR 130

Query: 133 RDVFLGTRECLGLVDDISQEEYETTVSYYNGV-NIDLGIMFHSFAYPKDK-KTPLKSYFT 190
           RD+FLG REC G V   + E+ +   +Y+G     LG M H F YP +  +  L
Sbjct: 131 RDIFLGARECQGYV---APCEFGSGDGFYDGQGKYHLGTMVHGFNYPDETGQHQLDVRLW 187

Query: 191 KTVMKNGVITFKAQSECDIVNTLSSYAFKA--PEEIKS 226
             VM+NG I F    +C IV  +     K   P+ ++S
Sbjct: 188 SAVMENGYIQFPRPEDCPIVRPVKEMEPKIFNPDNVQS 225
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2815

A DNA sequence (GASx1552R) was identified in *S.pyogenes* <SEQ ID 8129> which encodes the amino acid sequence <SEQ ID 8130>. Analysis of this protein sequence reveals the following:

```
Possible site: 53

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2770(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04055 GB:AP001508 unknown conserved protein in others
          [Bacillus halodurans]
 Identities = 252/836 (30%), Positives = 404/836 (48%), Gaps = 90/836 (10%)

Query: 3   MILAHYDCKKDKKQSLDEHLWHVACSSRQEASIIGQGDVLFLIGLYHDLGKADRTFQD-- 60
           M +AH         Q+L EHL  V C +    + +    V  L GL HDLGK    F+D
Sbjct: 1   MYIAHIREVDKVIQTLKEHLCGVQCLAETFGAKLRLQHVAGLAGLLHDLGKYTNEFKDYI 60

Query: 61  -------KLLNNPNRHVDHSYAGAKYLCSIIGPHLKNRGVDKNERMTFNEMVGYVISAHH 113
                +L       VDHS AG + L  +    L +R    +E++   E+VG  I +HH
Sbjct: 61  YKAVFEPELAEKKRGQVDHSTAGGRLLYQM----LHDRENSFHEKL-LAEVVGNAIISHH 115

Query: 114 GMYDLCYYFDDAEYYGFNKFKNRINRDLDGYHYHEDIKGYALKLEKKLCDYGYK-DLREL 172
                    +Y       N + R L+     +++  Y   +E+   +   + +L
Sbjct: 116 SNLQ--------DYISPTIESNFLTRVLE-----KELPEYESAVERFFQEVMTEAELARY 162

Query: 173 IDKAFDNYQQAMSSLNWQDKSEWDYYQSCMVRLYLSLLKNADILDTVNAYGLKISPMDKT 232
           + KA D  +Q   +   Q      Y  SC++         +AD  +T   +  +    + T
Sbjct: 163 VAKAVDEIKQFTDNSPTQSFFLTKYIFSCLI--------DADRTNT-RMFDEQAREEEPT 213

Query: 233 ERSFLKHSYLAAIEQKYASFGQPNNQ---LNTIRTEIAERVKERGKRDSKGIYRLDLPTG 289
           +  L   Y   +    AS  + ++     +N +R+ ++E+ +     R S GIY L +PTG
Sbjct: 214 QPQQLFEHYHQQLLNHLASLKESDSAQKPINVLRSAMSEQCESFAMRPS-GIYTLSIPTG 272

Query: 290 AGKTNLSMRYAFHQLVHHDKSRFFYITPFLSVLEQNASEIRKVTGD-LGVLEHHSNVVKQ 348
            GKT  S+RYA       ++K R  YI PF +++EQNA E+R + GD   +LEHHSNVV+
Sbjct: 273 GGKTLASLRYALKHAQEYNKQRIIYIVPFTTIIEQNAQEVRNILGDDENILEHHSNVVED 332

Query: 349 ANEDDDDKDSLLSA-----YLSDSWDSQVVLTSMVQFFQTLFKTKSANLRRFSSLINSVV 403
           +   D+ +D +++         D+WD  ++ T++VQF    +   + N RR  +L +SV+
Sbjct: 333 SENGDEQEDGVITKKERLRLARDNWDRPIIFTTLVQFLNVFYAKGNRNTRRLHNLSHSVL 392

Query: 404 ILDEVQSLPIEVTTLFNLTMNFLNKVMDTTIVLCTATQPAYDSSEIDHRICYGGNLGELA 463
           I DEVQ +P +  +LFN  +NFL +    +I+LCTATQP  ++  + H +     +
Sbjct: 393 IFDEVQKVPTKCVSLFNEALNFLKEFAHCSILLCTATQPTLEN--VKHSLLKDRD----G 446

Query: 464 EIVELTIEEKQIFSRTELRKFDDSDQKVHLTDVINLILGEE---NSVLAIFNTKKTVHNC 520
           EIV+   E  + F R E+   D +DQ +    +  + E     S L I NTKK V +
Sbjct: 447 EIVQNLTEVSEAFKRVEI--LDKTDQPMTNERLAEWVRDEAPSWGSTLIILNTKKVVKDL 504

Query: 521 YTMLKDMTDRPVYQLSTNMCAQHRLDLIAKIKTELQNNIPIICISTQLIEAGVDVDFHRV 580
           Y  L+     PV+ LST+MCA HR D + +I+  L+    P IC++TQLIEAGVDV F  V
Sbjct: 505 YEKLEG-GPLPVFHLSTSMCAAHRKDQLDEIRALLKEGTPFICVTTQLIEAGVDVSFKCV 563

Query: 581 IRSYSGIDSIVQAAGRCNREGKRDKGQVTLVNLTNEEENISRLTEIKTKKEATESILHKI 640
```

```
          IRS +G+DSI QAAGRCNR G+     V +++  + EE +S+L EI+  +E   ++L +
Sbjct: 564 IRSLAGLDSIAQAAGRCNRHGEEQLQYVYVID--HAEETLSKLKEIEVGQEIAGNVLARF 621

Query: 641 GSPIDISTLN-------RDFFEYYYANNQGLMDYPLED-----NLSIYDYLSLNIYQTAN 688
               +    N       R++F YYY+     ++Y +++         +  +   N Y T
Sbjct: 622 KKKAEKYEGNLLSQAAMREYFRYYYSKMDANLNYFVKEVDKDMTKLLMSHAVENSYVTYY 681

Query: 689 KKFKGK-----LKQAFKTAGAKMNLINNDMIGILVPYGEAEKKLAYLEELGVSHFLSAKD 743
           +K  G      L  ++KTA     +I+ +    +VPYGE +  +A L          S +
Sbjct: 682 QKNTGTHFPLLLNGSYKTAADHFRVIDQNTTSAIVPYGEGQDIIAQLN--------SGEW 733

Query: 744 YQTIKSLLKELQPFTVNV--RENDPLFE--TTKSYLNGQILVLTSEYYDTERGVKY 795
              +  +LK+ Q +TVN+  +E D L +      +L+G +  L   +Y  + GV +
Sbjct: 734 VDDLSKVLKKAQQYTVNLYSQEIDQLKKEGAIVMHLDGMVYELKESWYSHQYGVDF 789
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2816

A DNA sequence (GASx1558) was identified in *S.pyogenes* <SEQ ID 8131> which encodes the amino acid sequence <SEQ ID 8132>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1050(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2817

A DNA sequence (GASx1563) was identified in *S.pyogenes* <SEQ ID 8133> which encodes the amino acid sequence <SEQ ID 8134>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1872(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2818

A DNA sequence (GASx1564R) was identified in *S.pyogenes* <SEQ ID 8135> which encodes the amino acid sequence <SEQ ID 8136>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2173(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2819

A DNA sequence (GASx1566R) was identified in *S.pyogenes* <SEQ ID 8137> which encodes the amino acid sequence <SEQ ID 8138>. Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3486(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2820

A DNA sequence (GASx1568) was identified in *S.pyogenes* <SEQ ID 8139> which encodes the amino acid sequence <SEQ ID 8140>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2711(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2821

A DNA sequence (GASx1569) was identified in *S.pyogenes* <SEQ ID 8141> which encodes the amino acid sequence <SEQ ID 8142>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2822

A DNA sequence (GASx1576R) was identified in *S.pyogenes* <SEQ ID 8143> which encodes the amino acid sequence <SEQ ID 8144>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4042(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2823

A DNA sequence (GASx1577R) was identified in *S.pyogenes* <SEQ ID 8145> which encodes the amino acid sequence <SEQ ID 8146>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3342(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04515 GB:AP001509 unknown [Bacillus halodurans]
 Identities = 36/104 (34%), Positives = 55/104 (52%)
```

```
Query: 2   HMGAWNTGNNKILYTQESVTDDMIAKRDQSIKDAKESPILGFTVDTKVIKTELSNISNVM 61
           +M ++  GN  IL   E    D     +   + A  SP LGF  D+  ++TE++ ISNV
Sbjct: 392 NMPSFAIGNQLILKLYEDDPQDKWEAFEAFNESAIPSPALGFYFDSNPVRTEIAAISNVT 451

Query: 62  NRYKASINTGTVDPDEALPKLLADLKGAGWDKVQKEVQKQLDDF 105
           + +  ++  G VDP+E LP     L  AG  KV  E+Q+Q D++
Sbjct: 452 SEFSPALLKGAVDPEEYLPLFNDKLNEAGLQKVIDEMQRQFDEW 495
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2824

A DNA sequence (GASx1578R) was identified in *S.pyogenes* <SEQ ID 8147> which encodes the amino acid sequence <SEQ ID 8148>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04515 GB:AP001509 unknown [Bacillus halodurans]
 Identities = 134/346 (38%), Positives = 206/346 (58%), Gaps = 10/346 (2%)

Query: 21  AACESKSASKDSDVKLLMYQVGDKPDNFDELMTIANKRIKEKTGATVDLQYIGWGDWDDK 80
           +A E+++   D  V L  Y +G    + + +M   N   +EK  ATVDL+ + WG++D++
Sbjct: 42  SANETEATDLDH-VTLTWYMIGTPQPDLELVMEEVNAYTEEKINATVDLRMLDWGEYDER 100

Query: 81  MSTIIASGENYDIAF----ANNYVVNAQKGAFADLTTLMPKYAKKTYKNLDPAYIKGNTI 136
           M  I  SGE YDIAF    ANNY +NA++GAF +L  L+ ++ ++  + +DPA+++G  +
Sbjct: 101 MQVITTSGEAYDIAFTSSWANNYALNARRGAFLELNDLLDEHGQEMKELIDPAFLEGAQV 160

Query: 137 DGKLYAFPVDANVYAQQMLSFNKELVDKYGLDISNIKSYADAENVLKQFHEKEPNTAAFA 196
           DGKLYA P +  V  Q +LSFN ELV+K+ LD+S++ S AD E +L    E+E +    A
Sbjct: 161 DGKLYAVPTNKEVGQQAVLSFNNELVEKHNLDLSSVHSLADLEPLLAVIKEEESDVTPIA 220

Query: 197 IGQVFSMSGDYDYPLTKTQPFAVKIDEGKPTIINQYEDESFKNNLRLMHKWYKEGLIPTD 256
               F     +D  L +  PFA +++     +IN+YE++     L+ MH +YK+G I  D
Sbjct: 221 ---TFDAYLPFDSILQEEMPFAFRLEGNTNEVINKYEEDITMETLKTMHDYYKKGYIRPD 277

Query: 257 AATNTEGYPLEGNTWFMREETQGPMDYGDTILTNAAGKDIVSRPLTKPLKTTSQAQMANF 316
           AAT+T+ +PLE   WF+R+E   P  Y + I T  AG +I +RPL +P    +    +
Sbjct: 278 AATSTDSWPLETPNWFVRKELYQP--YAELIWTRTAGYEIATRPLHEPYIFNNSVTGSMQ 335

Query: 317 VVSSVSKNKEKAVEVLSLLNSDPELLNGLVYGVEGKAWEKIGDKKI 362
            +S+ SKN E+A+  L+LLNSDP L N L  G+EG  +E+++ D  I
Sbjct: 336 AISATSKNPERAMMFLNLLNSDPYLRNLLDKGIEGVHYEELEDGTI 381
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2825

A DNA sequence (GASx1582) was identified in *S.pyogenes* <SEQ ID 8149> which encodes the amino acid sequence <SEQ ID 8150>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0454(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2826

A DNA sequence (GASx1584R) was identified in *S.pyogenes* <SEQ ID 8151> which encodes the amino acid sequence <SEQ ID 8152>. Analysis of this protein sequence reveals the following:

```
Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3105(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 3-5
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG21428 GB:AF307332 meningioma-expressed antigen 5s splice
          variant [Homo sapiens]
 Identities = 94/271 (34%), Positives = 148/271 (53%), Gaps = 14/271 (5%)

Query: 120 GIIEGFYGTPWTREERLDCLRFIGNKRMNTYMYAPKDDDYQRKLWRDLYPEDWVTYFKEL 179
           G++EGFYG PW  E+R +  R +     +NTY+YAPKDD   R  WR++Y  +       L
Sbjct: 63  GVVEGFYGRPWVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMYSVEEAEQLMTL 122

Query: 180 LAVAKEEGLDFWYMISPGLDFDYTKEADYQLLYQKLQQLLALGVCHFGLLLDDIDYQIVD 239
           ++ A+E  ++F Y ISPGLD  ++   +   L +KL Q+   G   F LL DDID+ +
Sbjct: 123 ISAAREYEIEFIYAISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFDDIDHNMCA 182

Query: 240 AVERRFKKTAYAQAHLATEVHHFLNQQHAAPELVICPTE------YDNHHDSIYLQELSE 293
           A +  F   A+AQ  +  E++ +L +       + CPTE      Y N   S YL+ + E
Sbjct: 183 ADKEVFSSFAHAQVSITNEIYQYLGEPET---FLFCPTEYCGTFCYPNVSQSPYLRTVGE 239

Query: 294 RIPKEVAFFWTGPSTLASQISQADIETMAAVYQRPIIIWDNIPVNDYQKDPERLFLTPFA 353
           ++   +   WTGP  ++ +I    IE ++ + +R  +IWDNI  NDY  D +RLFL P+
Sbjct: 240 KLLPGIEVLWTGPKVVSKEIPVESIEEVSKIIKRAPVIWDNIHANDY--DQKRLFLGPYK 297

Query: 354 NRSPFLCQPDYQVKGIVSNPMISWELSKLTL 384
            RS  L     ++KG+++NP    +E + + +
Sbjct: 298 GRSTELIP---RLKGVLTNPNCEFEANYVAI 325
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2827

A DNA sequence (GASx1585R) was identified in *S.pyogenes* <SEQ ID 8153> which encodes the amino acid sequence <SEQ ID 8154>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4469(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2828

A DNA sequence (GASx1587) was identified in *S.pyogenes* <SEQ ID 8155> which encodes the amino acid sequence <SEQ ID 8156>. Analysis of this protein sequence reveals the following:

```
Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3082(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04509 GB:AP001509 unknown conserved protein in others
          [Bacillus halodurans]
 Identities = 221/425 (52%), Positives = 296/425 (69%), Gaps = 4/425 (0%)

Query: 12  RPIPTSVSQFMAKVESLCGDQHPDWALNFKTSFTNTLETTLKTYEDGTSFLLTGDIPAMW 71
           + IP S+   +A+V++   D      L F+  F NT  TT++  E GT F++TGDIPAMW
Sbjct: 4   KKIPRSLQAIIAQVKAHYADDQELQTL-FEQCFLNTYLTTIQEDEQGT-FVVTGDIPAMW 61

Query: 72  LRDSTAQMKPYLFLAKEDEEIRKIIAGLVKRQFRYICIDPYANAFNEEANEKGHQTDHTQ 131
           LRDS+AQ++PYL + KED ++ ++I G+++RQ+RYI  DPYANAFN+ AN++GHQ D T+
Sbjct: 62  LRDSSAQVRPYLTVVKEDADMARMIKGVIERQWRYILHDPYANAFNQTANKQGHQQDRTE 121

Query: 132 MNPWIWERKYEIDCLCYPIQLAYLLYRETGSTDQFNDDFHRGVELILDLWTVEQDH-AQS 190
           M+P +WERKYE+D LCYPIQLAYL ++ TG          + +E I  +W +EQDH A+S
Sbjct: 122 MSPLVWERKYELDSLCYPIQLAYLYWKATGDDSVLQPTLKQVLETIYRIWKIEQDHEAKS 181

Query: 191 PYLFERDTWRKEDTLTHAGKGSPVAPTGMTWSGFRPSDDACQYGYLIPSNMFAVVVLSYL 250
            Y FERD  R  DTL   GKG    PTGMTWSGFRPSDDAC YGYLIP+NMFAVVV +Y
Sbjct: 182 SYSFERDDCRVSDTLLRKGKGGYSVPTGMTWSGFRPSDDACLYGYLIPANMFAVVVSNYA 241

Query: 251 EDLYNNLFHNEPVATRAKQLKEAIQSGIADHALVQNSKGETIYAYEVDGLGQFSIMDDAN 310
           +L   +     +A   ++L+  I+ GI  +  + +     IY YE DG G+ ++MDDAN
```

```
Sbjct: 242 VELLTAM-EEIKLAEEFRELEADIRQGIGQYGKMDHPVYGEIYVYETDGNGRVNLMDDAN 300

Query: 311 IPSLLAAPYLGFCTKDDPIYLATRRTILSQENPYYYQGNAAAGIGSSHTPENYIWHIALA 370
           +PSLLA PYLG+ T DDP+Y  TRR ILS++NPYYY+G+ A G+GS HTP++Y+WHI+LA
Sbjct: 301 VPSLLAIPYLGYTTADDPVYQNTRRFILSRDNPYYYEGSYAKGVGSPHTPDHYVWHISLA 360

Query: 371 LQGLTALDQDSKKEMLDLLVATDAGTHLMHEGFDVNDPYQYTREWFSWANMMFCELLLDY 430
           +QG+TA+D   KK+++ +   T A T+ MHEGFDV+ P QYTR WF+WAN MF E LL
Sbjct: 361 IQGMTAIDSKEKKQIVAMFKQTHADTYFMHEGFDVDRPEQYTRSWFAWANSMFSEFLLSE 420

Query: 431 LGFSI 435
            G  +
Sbjct: 421 AGIYV 425
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2829

A DNA sequence (GASx1588) was identified in *S.pyogenes* <SEQ ID 8157> which encodes the amino acid sequence <SEQ ID 8158>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5250(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04508 GB:AP001509 unknown conserved protein in others
          (divided) [Bacillus halodurans]
 Identities = 312/737 (42%), Positives = 426/737 (57%), Gaps = 21/737 (2%)

Query: 123 FPDTFGNMGQTPQLMLKAGLQAAAFGRGIRPTGFNNQVDTSEKYSSQFSEISWQGPDNSR 182
           FPDTFG  GQ PQL+ +AG++AA FGRG+ PTGFNNQV   + YSS FSE+ W+ PD S+
Sbjct: 4   FPDTFGIYGQAPQLLAQAGIRAAVFGRGVTPTGFNNQVQHDD-YSSPFSELIWEAPDGSQ 62

Query: 183 ILGLLFANWYSNGNEIPTTEAEARLFWDKKLADAERFASTKHLLMMNGCDHQPVQLDVTK 242
           ++G+L ANWYSNGNEIPT E EA+ FW KKL DAERFAST  LL MNGCDHQPVQ DVT+
Sbjct: 63  VIGILLANWYSNGNEIPTDEDEAQTFWVKKLRDAERFASTSQLLFMNGCDHQPVQKDVTQ 122

Query: 243 AIALANQLYPDYEFVHSCFEDYLADLADDLPENLSTVQGEITSQETDGWYTLANTASARI 302
           AI +A  L+PD  F HS F DYL  + ++LP+ L  + GE+ +Q+TDGW TL NTASARI
Sbjct: 123 AIKVAETLFPDVAFKHSNFHDYLTQIKEELPKELQKITGELRNQKTDGWSTLVNTASARI 182

Query: 303 YLKQANTRVSRQLENITEPLAAMAYEVTSTYPHDQLRYAWKTLMQNHPHDSICGCSVDSV 362
           YLKQAN R    L N+ EP+  +      +   D   Y WK LM+NHPHDSICGCS+D+V
Sbjct: 183 YLKQANDRCQTLLTNVLEPMCLLV--ENKSLHRDFSEYYWKLLMENHPHDSICGCSIDAV 240

Query: 363 HREMMTRFEKAYEVGHYLAKEAAKQIADAIDTRDFPMDSQPFVLFNTSGHSKTSVAELSL 422
           HREM TRFEK          E  K+IA  I+T     ++ P V+  T+G S   V    +
Sbjct: 241 HREMKTRFEKVEAGATTFIAEQGKEIAAQINTLHDSEEAIPLVVLKTNGTSGKRVVRHKV 300

Query: 423 TWKKYHFGQRFPKEVYQEAQEYLARLSQSFQIIDTSGQVRPEAEILGTSIAFDYDLPKRS 482
             KK +F +      ++   + L  +      ++   +     E+    + F YDLP+
Sbjct: 301 AMKKIYFDEM----DFRHIPDRLKEIVMPTYRLEFPNKGSVPIEVQDAGVRFGYDLPRDG 356

Query: 483 FREPYFAIKVRLRLPITLPAMSWKTLALKLG------NETTPSETVSLYDDSNQCLENGF 536
           FR PY+A      L +T    S   L  + G       + T +    + D S   LEN
Sbjct: 357 FRRPYYA----RELEVTFSYDSDLYLGYECGFLVPVEEKQTEARKELIGDPSMNTLENEA 412
```

```
Query: 537 LKVMIQTDGRLTITDKQSGLIYQDLLRFEDCGDIGNEYISRQPNHDQPFYADQGTIKLNI 596
           +KVMI  +G  +I DK +G  Y+ L  +ED GDIGNEY+ +  +    +  +     + I
Sbjct: 413 MKVMIHRNGSYSILDKTTGFEYRHLGIYEDVGDIGNEYMFKASSDGVRYTTEACEASIRI 472

Query: 597 ISNTAQVAELEIQQTFAIPISADKLLQAEMEAVIDITERQARRSQEKAELTLTTLIRMEK 656
           I N +  A +EI  QT ++P +AD+ L+ E E ++    +R+A RS+E+ ++TL T + +E+
Sbjct: 473 IENNSLCATVEICQTLSVPAAADERLKEEQERLVWHPDRKAGRSKERTDITLRTELTLEQ 532

Query: 657 NNPRLQFTTRFDNQMTNHRLRVLFPTHLKTDHHLADSIFETVKRPNHPDATFWKNPSNPQ 716
               L+      DN   +HR+R LFP       +H ADSI+E V+RPN PD   W+NP+
Sbjct: 533 GAKGLKVNVNIDNTAKDHRMRALFPVERARGNHYADSIYEIVERPNTPDPK-WQNPAFDH 591

Query: 717 HQECFVSLFDGENGVTIGNYGLNEYEILPDTNTIAITLLRSVGEMGDWGYFPTPEAQCLG 776
           H +  VSL +GE G+TI   GL+EYEI+ D  +IA+TLLRSVGE+GDWG F TPEAQC G
Sbjct: 592 HMQRLVSLDNGEYGLTIATKGLHEYEIVSD--SIAVTLLRSVGELGDWGLFETPEAQCFG 649

Query: 777 KHSLSYSFESITKQTQFAS-YWRAQEGQVPVITTQTNQHEGTLAAEYSYLTGTNDQVALT 835
           ++  +           A+ Y  A +  V     QT Q  G L    +  + + LT
Sbjct: 650 QNEAQFVLLPHKGDVLSANVYVAAYDDPVEPTVIQTEQSMGPLPHATNLFQWSGEGLVLT 709

Query: 836 AFKRRLADNALITRSYN 852
           A K  +    +I R +N
Sbjct: 710 ACKPTMDGRGMILRWFN 726
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2830

A DNA sequence (GASx1589R) was identified in *S.pyogenes* <SEQ ID 8159> which encodes the amino acid sequence <SEQ ID 8160>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-11.30   Transmembrane  203 - 219 ( 195 - 221)
   INTEGRAL    Likelihood = -8.17   Transmembrane   61 -  77 (  59 -  82)
   INTEGRAL    Likelihood = -3.98   Transmembrane  107 - 123 ( 107 - 124)
   INTEGRAL    Likelihood = -3.40   Transmembrane   39 -  55 (  38 -  58)
   INTEGRAL    Likelihood = -2.34   Transmembrane  129 - 145 ( 126 - 145)
   INTEGRAL    Likelihood = -2.07   Transmembrane   89 - 105 (  87 - 105)

----- Final Results -----
              bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC10175 GB:AJ278302 histidine kinase [Streptococcus pneumoniae]
 Identities = 114/432 (26%), Positives = 219/432 (50%), Gaps = 10/432 (2%)

Query: 21  LTLKLFSFVSAIPLRLKNIFYLSLSMVLFQVVFWAFFPDHFILDVVMLAQF---LFFALI 77
           L + +F  V  I L  + IF   L  +L  VVF       +++  V L+ F   L+   +
Sbjct: 16  LKIVIFFKVDGISLTFERIFKAFLFKILLAVVFGML---GYMVGNVYLSYFMEPLYGIGL 72

Query: 78  ALYYGKSIKAKFLMFYAFFPLVSISLVKRFIVFFVMPLFGMPYSVVKHNTLLIYSITCFS 137
           +    + +  K L+FY  FP++ ++L  R + +FV+P  G    V    + +  I  F+
Sbjct: 73  SFLLLRELPKKLLLFYGLFPMILVNLFYRGVSYFVLPFLGQG-QVYDDYSFIWLCIIIFN 131

Query: 138 IFLIYRCIQVFHFDFSTWRQYFQSHRASKLLVFTNSSMALYYLCVQGIDVMSPSLSGLAT 197
            F+    ++   +DF++ R+        K L   N  M  YYL +Q +        G+ +
Sbjct: 132 FFISLAFLKWLDYDFTSLRKGILDKDFQKSLTQINWIMGAYYLVIQNLSYFEYE-QGIQS 190
```

```
Query: 198 TTARSIIVLFYFILFLTLLIHLERYVKQNSIEAIVQQKE--YRELINYSQHLGLLYQDIQ 255
          TT R +I++FY + F+ ++  L+ Y+K    E + Q+++  YRE+  YS+H+  LY++++
Sbjct: 191 TTVRHLILVFYLLFFMGIIKKLDTYLKDKLHERLNQEQDLRYREMERYSRHIEELYKEVR 250

Query: 256 ELRRLLTTVSSRLKIGIEQNDISIVRLTYEGILNAEKNNAKDDRLDLTCLDKLQVEAIRH 315
             R   T + + L++GIE+ D+  ++  Y+ +L       +D++ DL  L  ++  A++
Sbjct: 251 SFRHDYTNLLTSLRLGIEEEDMEQIKEIYDSVLKDSSEKLQDNKYDLGRLVNVRDRALKS 310

Query: 316 IVLAKLIEAKNKKLKVEVSIPNCIATFFLEVVDFTKLLSFLLDNAIEMSLETKQPCLSIA 375
          ++  K I+A++K +   V +P  I    + ++DF  ++S L DNAIE S+E  QP +SIA
Sbjct: 311 LLAGKFIKARDKNIVFNVEVPEEIQVEGVSLLDFLTVVSILCDNAIEASVEACQPHVSIA 370

Query: 376 FLDQNHKLVIVIQSSTKQGQDDSQSVFAIPALKKRDDWQFDLRNVTTILNRYDYLTISSQ 435
          F     +   +I++S K+   D   +F+  A  K ++    L  V  I+  +    ++++
Sbjct: 371 FFKNGAQETFIIENSIKEEGIDISEIFSFGASSKGEERGVGLYTVMKIVESHPNTSLNTT 430

Query: 436 IHDGILTQLIEI 447
             D +  Q++ +
Sbjct: 431 CQDHVFRQVLTV 442
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2831

A DNA sequence (GASx1593R) was identified in *S.pyogenes* <SEQ ID 8161> which encodes the amino acid sequence <SEQ ID 8162>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.28   Transmembrane    2 -  18 (   1 -  18)

----- Final Results -----
               bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2832

A DNA sequence (GASx1594) was identified in *S.pyogenes* <SEQ ID 8163> which encodes the amino acid sequence <SEQ ID 8164>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.93   Transmembrane   76 -  92 (  76 -  92)

----- Final Results -----
               bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF61313 GB:U96166 unknown [Streptococcus cristatus]
 Identities = 31/66 (46%), Positives = 40/66 (59%), Gaps = 2/66 (3%)

Query: 14 LLGRILSKYVGRLTSCIENETTKIRNHSRQNDTIGLNHLLGNLKTVHNPEIILKTINVYS 73
          + G  +SK    +   +  E  K+  ++  ND IG N LLG+LKTVHNPEII +   VYS
Sbjct: 30 VFGMDVSKTSSEVAILVNGE--KVHGYTILNDAIGFNRLLGDLKTVHNPEIIFEATGVYS 87

Query: 74 RRLQVF 79
          RRLQ F
Sbjct: 88 RRLQAF 93
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2833

A DNA sequence (GASx1598) was identified in *S.pyogenes* <SEQ ID 8165> which encodes the amino acid sequence <SEQ ID 8166>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2117(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2834

A DNA sequence (GASx1608) was identified in *S.pyogenes* <SEQ ID 8167> which encodes the amino acid sequence <SEQ ID 8168>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2835

A DNA sequence (GASx1619) was identified in *S.pyogenes* <SEQ ID 8169> which encodes the amino acid sequence <SEQ ID 8170>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2916(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2836

A DNA sequence (GASx1621) was identified in *S.pyogenes* <SEQ ID 8171> which encodes the amino acid sequence <SEQ ID 8172>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1899(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
alpha subunit [Escherichia coli]
 Identities = 110/211 (52%), Positives = 153/211 (72%)

Query: 7   KEITIKEAVAHVKDGDTIMVGGFMTNGTPEKLIDALVEKGVKDLTLICNDAGFPDKGVGK 66
           K +T+++A    +DG TIMVGGFM  GTP +L++AL+E GV+DLTLI ND  F D G+G
Sbjct: 4   KLMTLQDATGFFRDGMTIMVGGFMGIGTPSRLVEALLESGVRDLTLIANDTAFVDTGIGP 63

Query: 67  MVANKQFSTIIASHIGLNREAGRQMTEGETVIDLVPQGTLAERIRSGGFGLGGFLTPTGI 126
           ++ N +   +IASHIG N E GR+M  GE  + LVPQGTL E+IR GG GLGGFLTPTG+
Sbjct: 64  LIVNGRVRKVIASHIGTNPETGRRMISGEMDVVLVPQGTLIEQIRCGGAGLGGFLTPTGV 123

Query: 127 GTEVAKGKEVITIDGKDYLLEKPLKADVALIFANKADKNGNLQYAGSENNFNHVMAANAK 186
           GT V +GK+ +T+DGK +LLE+PL+AD+ALI A++ D  GNL Y  S  NFN ++A  A
Sbjct: 124 GTVVEEGKQTLTLDGKTWLLERPLRADLALIRAHRCDTLGNLTYQLSARNFNPLIALAAD 183

Query: 187 TTIVEAREIVDVGQMDPNFVHTPGIFVNYLV 217
            T+VE  E+V+ G++ P+ + TPG  +++++
Sbjct: 184 ITLVEPDELVETGELQPDHIVTPGAVIDHII 214
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2837

A DNA sequence (GASx1622) was identified in *S.pyogenes* <SEQ ID 8173> which encodes the amino acid sequence <SEQ ID 8174>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4668(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD54948 GB:AF157306 acetoacetate:butyrate/acetate coenzyme A
          transferase [Clostridium beijerinckii]
 Identities = 121/214 (56%), Positives = 161/214 (74%), Gaps = 5/214 (2%)

Query: 7   VLSKEEIQTRIAKRVAQELEHNTLVNLGIGLPTKVANYIPEGVTITLQSENGFVGLTGLT 66
           VL+KE I    AKRVA+EL+   LVNLGIGLPT VANY+P+ + IT +SENG VG+  +
Sbjct: 6   VLAKEII----AKRVAKELKKGQLVNLGIGLPTLVANYVPKEMNITFESENGMVGMAQMA 61

Query: 67  DD-HYDPTIVNAGGQPVSIAPGGAFFDSSTSFGIIRGGHVAATVLGALQVDKEASIANYL 125
               DP I+NAGG+ V++ P GAFFDSSTSF +IRGGHV   VLGAL+VD+E ++AN++
Sbjct: 62  SSGENDPDIINAGGEYVTLLPQGAFFDSSTSFALIRGGHVDVAVLGALEVDEEGNLANWI 121

Query: 126 IPGKMVPGMGGAMDLLVGAKKVIVAMEHTNKGKAKILDKCTLPLTAQNVVNLIITEMGVF 185
           +P K+VPGMGGAMDL +GAKK+IVAM+HT KGK KI+ KCTLPLTA+  V+LI+TE+ V
Sbjct: 122 VPNKIVPGMGGAMDLAIGAKKIIVAMQHTGKGKPKIVKKCTLPLTAKAQVDLIVTELCVI 181

Query: 186 EYQDEGLCALEINPDYTFEDVQNVTEVTLIDKTN 219
           +  ++GL   EI+ D T ++++ +T+  LI   N
Sbjct: 182 DVTNDGLLFREIHKDTTIDEIKFLTDADLIIPDN 215
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2838

A DNA sequence (GASx1628R) was identified in *S.pyogenes* <SEQ ID 8175> which encodes the amino acid sequence <SEQ ID 8176>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1243(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2839

A DNA sequence (GASx1639R) was identified in *S.pyogenes* <SEQ ID 8177> which encodes the amino acid sequence <SEQ ID 8178>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.65   Transmembrane   55 -  71 (  44 -  73)
   INTEGRAL    Likelihood = -7.64   Transmembrane   13 -  29 (   5 -  31)

----- Final Results -----
               bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2840

A DNA sequence (GASx1643) was identified in *S.pyogenes* <SEQ ID 8179> which encodes the amino acid sequence <SEQ ID 8180>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.0766(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2841

A DNA sequence (GASx1645R) was identified in *S.pyogenes* <SEQ ID 8181> which encodes the amino acid sequence <SEQ ID 8182>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2842

A DNA sequence (GASx1649R) was identified in *S.pyogenes* <SEQ ID 8183> which encodes the amino acid sequence <SEQ ID 8184>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0931(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2843

A DNA sequence (GASx1650) was identified in *S.pyogenes* <SEQ ID 8185> which encodes the amino acid sequence <SEQ ID 8186>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5678(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2844

A DNA sequence (GASx1651R) was identified in *S.pyogenes* <SEQ ID 8187> which encodes the amino acid sequence <SEQ ID 8188>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2761(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2845

A DNA sequence (GASx1667R) was identified in *S.pyogenes* <SEQ ID 8189> which encodes the amino acid sequence <SEQ ID 8190>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2967(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2846

A DNA sequence (GASx1672) was identified in *S.pyogenes* <SEQ ID 8191> which encodes the amino acid sequence <SEQ ID 8192>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.82   Transmembrane    3 -  19 (   1 -  20)

----- Final Results -----
             bacterial membrane --- Certainty=0.2529(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2847

A DNA sequence (GASx1673R) was identified in *S.pyogenes* <SEQ ID 8193> which encodes the amino acid sequence <SEQ ID 8194>. Analysis of this protein sequence reveals the following:

```
Possible site: 38

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.86   Transmembrane   51 -  67 (  47 -  75)
   INTEGRAL    Likelihood = -5.20   Transmembrane   27 -  43 (  24 -  45)
   INTEGRAL    Likelihood = -3.66   Transmembrane  112 - 128 ( 112 - 131)

----- Final Results -----
```

```
bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF41294 GB:AE002440 conserved hypothetical protein [Neisseria
        meningitidis MC58]
 Identities = 61/148 (41%), Positives = 96/148 (64%)

Query: 1   LKKSITNEKAILAQGGQEFGAQNTKFLTLLHIMIYVFAVIEALLKQIKFDGISFLGLLLM 60
           L  SI +EKA++A+G +++G  N+  L  +H + Y+   +   L    F+GIS +G L +
Sbjct: 19  LAVSIKHEKALIAKGAKQYGKTNSTLLAAVHTLYYLACFVWVWLSDTAFNGISLIGTLTV 78

Query: 61  LLSVAVLYEVTRILGDIWTVKLMLAKDHKYVDHWLFKTIKHPNYFLNIAPELVGIALLCH 120
           + S  +L  + + LG+IWTVK+ +  +H+    WLFKT +HPNYFLNI PEL+GIALLC
Sbjct: 79  MASFVILSLIIKQLGEIWTVKIYILPNHQINRSWLFKTFRHPNYFLNIIPELIGIALLCQ 138

Query: 121 AKITAMLLFPCYIVVIYLRIREENKLLA 148
           A    ++  P Y++V++ RIR+E + +A
Sbjct: 139 AWYVLLIGLPIYLLVLFKRIRQEEQAMA 166
```

A related GBS gene <SEQ ID 9009> and protein <SEQ ID 9010> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
McG: Discrim Score:      5.86
GvH: Signal Score (-7.5): 0.14
     Possible site: 60
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 2 value:  -8.23 threshold:  0.0
   INTEGRAL    Likelihood = -8.23   Transmembrane   69 -  85 (  64 -  89)
   INTEGRAL    Likelihood = -3.29   Transmembrane  142 - 158 ( 140 - 159)
   PERIPHERAL  Likelihood =  1.70     123
 modified ALOM score:   2.15

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4291(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
42.1/64.0% over 168aa

                                                                    imported
  EGAD|177248| conserved hypothetical protein (Neisseria meningitidis) Insert characterized
  GP|7379797|emb|CAB84365.1||AL162755 putative integral membrane protein (Neisseria
meningitidis) Insert characterized
  GP|7226121|gb|AAF41294.1||AE002440 conserved hypothetical protein (Neisseria meningitidis
MC58) Insert characterized
   PIR|F81147|F81147 probable integral membrane protein NMA1102 - Neisseria meningitidis
(group B strain MD58, group A strain Z2491) Insert
 characterized

ORF00432(301 - 807 of 1140)
EGAD|177248|NMB0883(1   -   169   of   169)   conserved   hypothetical   protein   (Neisseria
meningitidis)GP|7379797|emb|CAB84365.1||AL162755 putative integral membrane protein
(Neisseria meningitidis)GP|7226121|gb|AAF41294.1||AE002440 conserved hypothetical protein
(Neisseria meningitidis MC58)PIR|F81147|F81147 probable integral membrane protein NMA1102
[imported] - Neisseria meningitidis (group B strain MD58, group A strain Z2491)
%Match = 19.0
%Identity = 42.0  %Similarity = 63.9
Matches = 71  Mismatches = 61  Conservative Sub.s = 37
```

```
237       267       297       327       357       387       417       447
SSGEYHLLTSDHSLV*IGKAXX*LIXXXEFTMSIIIGLMAAMFIIRLAYLKLSIANEKALRKNGAKEYGVGVSKAITVLH
                                |::|: :::  ||||| :| :|| :|||| |||:||  |  :  :|
                               MTMILSILSLFFIIRLLFLAVSIKHEKALIAKGAKQYGKTNSTLLAAVH
                                       10        20        30        40

477       507       537       567       597       627       657       687
IIIYFSSVTEAILTKASFNFVSVIGLSLMIFSVFMLHTVTRLLGRIWTVKLMVDKNHQFVDHWLFRVVKHPNYFLNIAPE
 : |::      |:  :|| :|:||   :: |  :|  : : || |||||: :  |||    |||:  :||||||| ||
TLYYLACFVWVWLSDTAFNGISLIGTLTVMASFVILSLIIKQLGEIWTVKIYILPNHQINRSWLFKTFRHPNYFLNIIPE
         60        70        80        90       100       110       120

717       747       777       807       837       867       897       927
LLGVTLLCHAKYTALFVLPIYAFVIYLRIREENLLLKTIIIPNGIKKSRVY*E*DK**T*KSFFVILSQ*EEVFISCFFS
|:|: |||:|  |   |  |||| :|:: |||:|   : |:
LIGIALLCQAWYVLLIGLPIYLLVLFKRIRQEEQAMATLF
        140       150       160
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2848

A DNA sequence (GASx1674R) was identified in *S.pyogenes* <SEQ ID 8195> which encodes the amino acid sequence <SEQ ID 8196>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3098(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2849

A DNA sequence (GASx1677R) was identified in *S.pyogenes* <SEQ ID 8197> which encodes the amino acid sequence <SEQ ID 8198>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.86   Transmembrane  254 - 270 ( 248 - 280)
   INTEGRAL    Likelihood = -7.01   Transmembrane  303 - 319 ( 296 - 322)
   INTEGRAL    Likelihood = -2.39   Transmembrane   74 -  90 (  74 -  91)
   INTEGRAL    Likelihood = -1.91   Transmembrane  201 - 217 ( 199 - 217)
   INTEGRAL    Likelihood = -1.91   Transmembrane  223 - 239 ( 220 - 240)
   INTEGRAL    Likelihood = -1.65   Transmembrane  118 - 134 ( 115 - 135)
   INTEGRAL    Likelihood = -1.49   Transmembrane   56 -  72 (  55 -  72)
   INTEGRAL    Likelihood = -0.32   Transmembrane   13 -  29 (  13 -  30)

----- Final Results -----
               bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05126 GB:AP001511 unknown conserved protein [Bacillus halodurans]
 Identities = 249/534 (46%), Positives = 380/534 (70%)

Query: 12  QDIAFHFFGGLGLFLFSIKYMGDGLQQAAGDKLRYYIDKYTSNPFFGILVGIAMSALIQS 71
           Q + F FFGGLG+FLF IKYMGDGLQ+ AG++LR  +DK+T+NP  G+L GI ++ L+Q+
Sbjct: 6   QTLLFMFFGGLGIFLFGIKYMGDGLQKVAGERLRDLLDKFTTNPLMGVLAGIVVTVLLQT 65

Query: 72  SSGVTVITVGLVSAGLLNLRQAIGIVMGANIGTTITSFLIGFKLGDYALPMIFIGAACLF 131
           S+G TV+T+GLV+AG + L+QAIG++MGANIGTT+T+F+IG K+ +YALP+I +GAA +F
Sbjct: 66  STGTTVLTIGLVNAGFMTLKQAIGVIMGANIGTTVTAFIIGIKISEYALPIIAVGAALIF 125

Query: 132 FTSNKKLNNFGRIIFGVGGIFFSLNLMGDAMDPLKSVSAFQNYLATLGDKPFQGVFIGTA 191
           F  NKK+NN G++IFG G +F+ LN MG+ ++PL+ + AF +   ++ + P  GV IGT
Sbjct: 126 FIKNKKVNNIGQVIFGFGTLFYGLNTMGEGLNPLRELQAFADLTVSMSENPLLGVLIGTI 185

Query: 192 LTMLIQSSAAIIGILQGLFSGGLLTLQGAIPILLGSNIGTCITAVLAAIGSNIAAKRVAA 251
            T  +QSS+A IG+LQ L+  G + L  A+P+L G NIGT ITAVLAAIG+++AAKR A
Sbjct: 186 FTAAVQSSSASIGLLQQLYDQGAMDLFAALPVLFGDNIGTTITAVLAAIGASVAAKRAAL 245

Query: 252 AHVLFNLIGTIIFMIILVPFTSLMLWLQSKLSLTPEMTIAFSHGSFNITNTILLIPFISL 311
            HV+FNLIGTII +II++PFT  + +L    +L   MTIAF+HG FN++NTI+  PFI +
Sbjct: 246 THVIFNLIGTIIVLIIIIPFTHFIAYLAEVFALNRPMTIAFAHGIFNVSNTIIQFPFIGI 305

Query: 312 LAMIVTRLIPGEDEVVKYEALYLDRLLITQAPSIALGNAHKELVHLASYAIQAFEASYSY 371
           LA+IVT+L+PG+D  ++Y+A +LD   +  +P+IALG A +E++ +A ++ +       Y
Sbjct: 306 LAIIVTKLVPGDDFYIEYKAKHLDPRFVGSSPAIALGQAKQEVLRMAEFSEKGLLEVSKY 365

Query: 372 IMTADGKFGEKVKRYERAVDTIDEELTTYLVDISNEALSPSENEVLAGILDSSRDLERIG 431
           +     K  E   ++E A++ +D ++T YL+ IS+ +LS  ++++   ++D+ RD+ERIG
Sbjct: 366 MENGQKKHAEMAVQFEDAINNLDRKITEYLISISSRSLSAQDSKMHGMLMDTVRDIERIG 425

Query: 432 DHSESLGILIEGIISKQIGFSISARQELTEMYQLTHCLTLDAIRAIVDSDTDLAQTIVTR 491
           DH E++  L +   + ++  S  A  +L EM+ LTH     +AI ++   D + A++++ +
Sbjct: 426 DHIENIVELKDYQKANKVKISEKALHDLQEMFDLTHSTLTEAIMSLETGDLEAARSVIEK 485

Query: 492 HKEIEEKERRLRKTHIKRLNCGECTAQAGINFIDIISHYTRITDHALNLAEKVL 545
           + I++ ER+LRK HI R+N G CT  AGI F+DI+S+  RI DH++N+AE V+
Sbjct: 486 EEHIDQMERKLRKQHIIRVNEGNCTGAAGIVFVDIVSNLERIGDHSVNIAEAVI 539
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2850

A DNA sequence (GASx1678R) was identified in *S.pyogenes* <SEQ ID 8199> which encodes the amino acid sequence <SEQ ID 8200>. Analysis of this protein sequence reveals the following:

```
Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2940(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2851

A DNA sequence (GASx1685R) was identified in *S.pyogenes* <SEQ ID 8201> which encodes the amino acid sequence <SEQ ID 8202>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.11   Transmembrane   13 -  29 (   9 -  31)

----- Final Results -----
               bacterial membrane --- Certainty=0.3845(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2852

A DNA sequence (GASx1695R) was identified in *S.pyogenes* <SEQ ID 8203> which encodes the amino acid sequence <SEQ ID 8204>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1357(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2853

A DNA sequence (GASx1698) was identified in *S.pyogenes* <SEQ ID 8205> which encodes the amino acid sequence <SEQ ID 8206>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1970(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2854

A DNA sequence (GASx1713) was identified in *S.pyogenes* <SEQ ID 8207> which encodes the amino acid sequence <SEQ ID 8208>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3092(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2855

A DNA sequence (GASx1737) was identified in *S.pyogenes* <SEQ ID 8209> which encodes the amino acid sequence <SEQ ID 8210>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1878(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2856

A DNA sequence (GASx1748R) was identified in *S.pyogenes* <SEQ ID 8211> which encodes the amino acid sequence <SEQ ID 8212>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2841(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2857

A DNA sequence (GASx1750R) was identified in *S.pyogenes* <SEQ ID 8213> which encodes the amino acid sequence <SEQ ID 8214>. Analysis of this protein sequence reveals the following:

```
Possible site: 59

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.22   Transmembrane   18 -  34 (  18 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2858

A DNA sequence (GASx1754) was identified in *S.pyogenes* <SEQ ID 8215> which encodes the amino acid sequence <SEQ ID 8216>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2859

A DNA sequence (GASx1759) was identified in *S.pyogenes* <SEQ ID 8217> which encodes the amino acid sequence <SEQ ID 8218>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1534(Affirmative) < succ>
```

```
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2860

A DNA sequence (GASx1764R) was identified in *S.pyogenes* <SEQ ID 8219> which encodes the amino acid sequence <SEQ ID 8220>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.74   Transmembrane   90 - 106 (  87 - 121)
   INTEGRAL    Likelihood = -4.57   Transmembrane  210 - 226 ( 205 - 229)
   INTEGRAL    Likelihood = -4.19   Transmembrane   43 -  59 (  42 -  62)
   INTEGRAL    Likelihood = -3.77   Transmembrane  137 - 153 ( 137 - 155)

----- Final Results -----
               bacterial membrane --- Certainty=0.3697(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2861

A DNA sequence (GASx1768R) was identified in *S.pyogenes* <SEQ ID 8221> which encodes the amino acid sequence <SEQ ID 8222>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.37   Transmembrane   26 -  42 (  17 -  47)
   INTEGRAL    Likelihood = -7.54   Transmembrane   53 -  69 (  46 -  73)
   INTEGRAL    Likelihood = -3.29   Transmembrane  209 - 225 ( 209 - 225)
   INTEGRAL    Likelihood = -2.13   Transmembrane   82 -  98 (  82 -  98)
   INTEGRAL    Likelihood = -1.65   Transmembrane    9 -  25 (   9 -  25)
   INTEGRAL    Likelihood = -0.85   Transmembrane  117 - 133 ( 117 - 134)

----- Final Results -----
               bacterial membrane --- Certainty=0.5946(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB84959 GB:AE000829 conserved protein [Methanobacterium
          thermoautotrophicum]
 Identities = 54/192 (28%), Positives = 90/192 (46%), Gaps = 6/192 (3%)
```

```
Query: 7   TKLLLLVLANACFFFRVDGFLEFIIVIFLLLLLSALNKKKLA--FKLAVVYLLMIGLSVI 64
           +KL ++V A     F  D  L  I+ +    L++     + A  F   ++   ++ L++I
Sbjct: 32  SKLTVVVSATLLSTFISDLTLLIIMGVIFTALIAHSGSLRFAAPFLSFIILFWLVSLAII 91

Query: 65  PLSIFPSYLDHLLSFVSIAGRLVFPSLLAGLITIKTTTIYELVHGLRKWRFPEVWLLTLA 124
            +    S   H + F+S+     F    AGL    TT   +L   LR  R P   + TL
Sbjct: 92  MVL---SGNPHTMGFLSLFFARFFIISAAGLSFAFTTEPQKLAESLRSVRIPGEIVFTLT 148

Query: 125 VMCRFIPMIRQECCVIHRSLKIRGIILTKWSILIRPKQYLEYLMVPLLLSLIRSSQELTI 184
           V  R+IP +  E   I  SLK+R   L+  SI+ RP      L++P+++  ++ S E+ I
Sbjct: 149 VALRYIPALAVEASSIWDSLKLR-TSLSGSSIIRRPSLLYRGLIIPMIIRTVKISDEVAI 207

Query: 185 ASLTKGLAVNKG 196
           A+ T+G    +G
Sbjct: 208 AAETRGFNPREG 219
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2862

A DNA sequence (GASx1769R) was identified in *S.pyogenes* <SEQ ID 8223> which encodes the amino acid sequence <SEQ ID 8224>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.32   Transmembrane  164 - 180 ( 158 - 186)
   INTEGRAL    Likelihood = -4.67   Transmembrane   85 - 101 (  84 - 105)
   INTEGRAL    Likelihood = -3.03   Transmembrane   42 -  58 (  42 -  61)
   INTEGRAL    Likelihood = -2.76   Transmembrane  118 - 134 ( 117 - 134)
   INTEGRAL    Likelihood = -2.07   Transmembrane   64 -  80 (  64 -  82)
   INTEGRAL    Likelihood = -1.22   Transmembrane   18 -  34 (  17 -  34)

----- Final Results -----
              bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2863

A DNA sequence (GASx1776R) was identified in *S.pyogenes* <SEQ ID 8225> which encodes the amino acid sequence <SEQ ID 8226>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.37   Transmembrane    4 -  20 (   1 -  22)
   INTEGRAL    Likelihood = -0.43   Transmembrane  261 - 277 ( 261 - 278)

----- Final Results -----
              bacterial membrane --- Certainty=0.3548(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2864

A DNA sequence (GASx1777R) was identified in *S.pyogenes* <SEQ ID 8227> which encodes the amino acid sequence <SEQ ID 8228>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.17   Transmembrane 1217 -1233 (1215 -1235)

----- Final Results -----
               bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF53254 GB:AE003639 CG16974 gene product [Drosophila
          melanogaster]
 Identities = 84/238 (35%), Positives = 133/238 (55%), Gaps = 10/238 (4%)

Query: 516 LRLDHYELTDISLL--KHAKNITELHLDGNQITEIPKELFSQMKQLRFLNLRSNHLTYLD 573
           L +    L++ SLL  ++ K + ELHLD +++T +P+    ++ +LR LNL  N LT L
Sbjct: 232 LEMSGNRLSNCSLLNLQYMKQLQELHLDRSELTYLPQRFLGELSELRMLNLSQNLLTELP 291

Query: 574 KDTFKSNAQLRELYLSSNFIHSLEGGLFQSLHHLEQLDLSKNRIGRLCDNPFEGLSRLTS 633
           +D F    +L  LYLS N +  L   LFQ+   L+ LDLS NR+    DN F    +L
Sbjct: 292 RDIFVGALKLERLYLSGNRLSVLPFMLFQTAADLQVLDLSDNRLLSFPDNFFARNGQLRQ 351

Query: 634 LGFAENSLEEIPEKALEPLTSLNFIDLSQNNLALLP-KTIEKLRALSTIVASRNHITRID 692
           L    N L+ I + +L  L  L  +DLSQN+L+++  K  E L  L  +  S N++T +
Sbjct: 352 LHLQRNQLKSIGKHSLYSLRELRQLDLSQNSLSVIDRKAFESLDHLLALNVSGNNLTLLS 411

Query: 693 NISFKNLPKLSVLDLSTNEISNLPNGIFKQNNQL-------TKLDFFNNLLTQVEESV 743
           +I F++L  L  LDLS N+   LP+G+F++    L       T ++ F+N +++ +ES+
Sbjct: 412 SIIFQSLHALRQLDLSRNQFKQLPSGLFQRQRSLVLLRIDETPIEQFSNWISRYDESL 469
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2865

A DNA sequence (GASx1778R) was identified in *S.pyogenes* <SEQ ID 8229> which encodes the amino acid sequence <SEQ ID 8230>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1067(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2866

A DNA sequence (GASx1779) was identified in *S.pyogenes* <SEQ ID 8231> which encodes the amino acid sequence <SEQ ID 8232>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1885(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2867

A DNA sequence (GASx1786R) was identified in *S.pyogenes* <SEQ ID 8233> which encodes the amino acid sequence <SEQ ID 8234>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0612(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2868

A DNA sequence (GASx1790) was identified in *S.pyogenes* <SEQ ID 8235> which encodes the amino acid sequence <SEQ ID 8236>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2869

A DNA sequence (GASx1791R) was identified in *S.pyogenes* <SEQ ID 8237> which encodes the amino acid sequence <SEQ ID 8238>. Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.90   Transmembrane   28 -  44 (  28 -  44)

----- Final Results -----
               bacterial membrane --- Certainty=0.1362(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9155> which encodes the amino acid sequence <SEQ ID 9156>. Analysis of this protein sequence reveals the following:

```
     Possible site: 25
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty= 0.300(Affirmative) < succ>
              bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA24923 GB:L06331 endoglycosidase [Chryseobacterium
          meningosepticum]
 Identities = 105/322 (32%), Positives = 153/322 (46%), Gaps = 53/322 (16%)

Query: 106 ADKQAQELAKMKIPEKIPMKPLHGSLYGGYFRTWHDKTSDPTEKDKVNSMGELPKEVDLA 165
           A K    ++ + +   I  K     +  GY+RTW D     T    + SM  LP  +D+
Sbjct: 37  AQKSGVTVSAVNLSNLIAYKNSDHQISAGYYRTWRDSA---TASGNLPSMRWLPDSLDMV 93

Query: 166 FIFHDWTKDYSLFWKELATKHVPKLNKQGTRVIRTIPWRFLAGGDNSGIAEDTSKYPNTP 225
            +F D+T   + +W  L T +VP L+K+GT+VI T+      G  NS     T+
Sbjct: 94  MVFPDYTPPENAYWNTLKTNYVPYLHKRGTKVIITL------GDLNSA----TTTGGQDS 143

Query: 226 EGNKALAKAIVDEYVYKYNLDGLDVDVEHDSIPKVDKKEDTAGVERSIQVFEEIGKLIGP 285
            G  + AK I D++V +YNLDG+D+D+E              A + + +   + + K  GP
Sbjct: 144 IGYSSWAKGIYDKWVGEYNLDGIDIDIE--------SSPSGATLTKFVAATKALSKYFGP 195

Query: 286 KGVDKSRLFIMDSTYMADKNP--LIERGAPYINLLLVQVYGSQGEKGGWEPVSNRPEKTM 343
           K     + F+ D+     ++NP     + AP  N + +Q YG              R    +
Sbjct: 196 KS-GTGKTFVYDT----NQNPTNFFIQTAPRYNYVFLQAYG-------------RSTTNL 237

Query: 344 EERWQGYSKYIRPEQYMIGFSFYEENAQEGNLWYDINSRKDEDKANGINTDITGTRAERY 403
                Y+ YI  +Q++ GFSFYEEN   GN W D+   +     NG     TG RA  Y
Sbjct: 238 TTVSGLYAPYISMKQFLPGFSFYEENGYPGNYWNDVRYPQ-----NG-----TG-RAYDY 286

Query: 404 ARWQPKTGGVKGGIFSYAIDRD 425
           ARWQP T G KGG+FSYAI+RD
Sbjct: 287 ARWQPATGKKGGVFSYAIERD 307
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2870

A DNA sequence (GASx1803) was identified in *S.pyogenes* <SEQ ID 8239> which encodes the amino acid sequence <SEQ ID 8240>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2099(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2871

A DNA sequence (GASx1806R) was identified in *S.pyogenes* <SEQ ID 8241> which encodes the amino acid sequence <SEQ ID 8242>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2706(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB16126 GB:Z99124 ribosomal protein S18 [Bacillus subtilis]
 Identities = 51/77 (66%), Positives = 63/77 (81%)

Query: 1  MAQQRRGGFKRRKKVDFIAANKIEYVDYKDTELLSRFVSERGKILPRRVTGTSAKNQRKV 60
          MA  RRGG  +R+KV +  +N I ++DYKD +LL +FVSERGKILPRRVTGT+AK QRK+
Sbjct: 3  MAGGRRGGRAKRRKVCYFTSNGITHIDYKDVDLLKKFVSERGKILPRRVTGTNAKYQRKL 62

Query: 61 TTAIKRARVMALMPYVN 77
          T AIKRAR MAL+PYV+
Sbjct: 63 TAAIKRARQMALLPYVS 79
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2872

A DNA sequence (GASx1809R) was identified in *S.pyogenes* <SEQ ID 8243> which encodes the amino acid sequence <SEQ ID 8244>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
```

```
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.59   Transmembrane   70 -  86 (  66 -  92)
   INTEGRAL    Likelihood = -6.42   Transmembrane   13 -  29 (   8 -  33)
   INTEGRAL    Likelihood = -5.68   Transmembrane   48 -  64 (  43 -  69)

----- Final Results -----
               bacterial membrane --- Certainty=0.4036(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2873

A DNA sequence (GASx1813R) was identified in *S.pyogenes* <SEQ ID 8245> which encodes the amino acid sequence <SEQ ID 8246>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.51   Transmembrane  127 - 143 ( 113 - 147)
   INTEGRAL    Likelihood =-10.46   Transmembrane  151 - 167 ( 149 - 167)
   INTEGRAL    Likelihood = -4.41   Transmembrane   59 -  75 (  57 -  77)

----- Final Results -----
               bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB98363 GB:U67490 lipoprotein B (lppB) [Methanococcus
          jannaschii]
 Identities = 43/143 (30%), Positives = 68/143 (47%), Gaps = 7/143 (4%)

Query: 25  LLNVLLKIITGVMY--ILYPSFLIFTLWQGMTFQLWLRLLIIPAVGFIALSYIRKRFDFP 82
           + + ++ II+   Y  I   S +IF   +    +L   L +   + F +L Y+       P
Sbjct: 181 IFDAIMPIISKTAYPLIAITSLIIFIKNRKFGMKLIFALFLAFMIAF-SLKYLVNE---P 236

Query: 83  RPYEKWNIKPLIDKDTKGRSMPSRHVFSATMISMCLLRYYVYFGIVCLILSALLAICRVI 142
           RPY   +   L+  +    S PS H   A  ++  LL Y    GI+ L  + ++A  RV
Sbjct: 237 RPYLVLDNVHLLCNEGNEPSFPSGHTTLAFTLATSLLFYSKKLGILFLSWAIIVAYSRVY 296

Query: 143 AGIHYPKDVIVGYLIGLMLGLCL 165
            G+HYP DV+ G +IG+  G CL
Sbjct: 297 VGVHYPLDVLAGMIIGIFCG-CL 318
```

A related GBS gene <SEQ ID 9011> and protein <SEQ ID 9012> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:      3.19
GvH: Signal Score (-7.5): -2.18
     Possible site: 55
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 3 value: -11.78 threshold:  0.0
```

```
    INTEGRAL     Likelihood =-11.78   Transmembrane  126 - 142 ( 112 - 147)
    INTEGRAL     Likelihood =-11.30   Transmembrane  150 - 166 ( 147 - 166)
    INTEGRAL     Likelihood = -4.41   Transmembrane   58 -  74 (  56 -  76)
    PERIPHERAL   Likelihood =  3.29     107
modified ALOM score:   2.86

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5713(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01020(472 - 792 of 1098)
EGAD|44548|MJ0374(213  -  318  of  330)  conserved  hypothetical  protein  {Methanococcus
jannaschii} OMNI|MJ0374 conserved hypothetical protein SP|Q57819|Y374_METJA HYPOTHETICAL
PROTEIN  MJ0374.  GP|1591081|gb|AAB98363.1||U67490  lipoprotein  B  (lppB)  {Methanococcus
jannaschii} PIR|F64346|F64346 hypothetical protein MJ0374 - Methanococcus jannaschii
%Match = 6.8
%Identity = 30.8  %Similarity = 53.3
Matches = 33  Mismatches = 49  Conservative Sub.s = 24

222       252       282       312       342       372       402       432
EGVTKYLRRNKHVKHFAYAPQNAGGSGATIVTLG*IMESYEQFYAKLSQPFRKSPQLIILLNFLLKIVTGMMYILYPSFL

VIAWLSGIFEMHKLLFTVGTIIGRLPRFLAVAYFGDVLGNINRLSDINIYLFYLINSHYNYIFDAIMPIISKTAYPLIAI
        130       140       150       160       170       180       190

462       492       522       552       582       612       642       672
IFTLWQGMTFQLWLRLLIIPAVGFIALSYIRKRLDFPRPYEKWNIKPLIYKDTEGRSMPSRHVFSATMISMCLLRYYVYF
              ::|:    : |:    ::  :: ||||   :   |: :    | || |   | ::  || |   :
TSLIIFIKNRKFGMKLIFALFLAFMIAFSLKYLVNEPRPYLVLDNVHLLCNEGNEPSFPSGHTTLAFTLATSLLFYSKKL
        210       220       230       240       250       260       270

702       732       762       792       822       852       882       912
GIVCLILSVLLAICRVIAGIHYPKDVIVGYLIGLILGLCLFI*RVRSK*FQKQLDSCTIGLSLR*NGEKRWH*K*QMLHL
||: |  ::::|  ||  |:||| ||: | :||:  | ||
GILFLSWAIIVAYSRVYVGVHYPLDVLAGMIIGIFCG-CLTRIDIYKLIDNI
        290       300       310       320       330
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2874

A DNA sequence (GASx1815R) was identified in *S.pyogenes* <SEQ ID 8247> which encodes the amino acid sequence <SEQ ID 8248>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0888(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2875

A DNA sequence (GASx1825R) was identified in *S.pyogenes* <SEQ ID 8249> which encodes the amino acid sequence <SEQ ID 8250>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16   Transmembrane    7 -  23 (   7 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2876

A DNA sequence (GASx1832) was identified in *S.pyogenes* <SEQ ID 8251> which encodes the amino acid sequence <SEQ ID 8252>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0918(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2877

A DNA sequence (GASx1836R) was identified in *S.pyogenes* <SEQ ID 8253> which encodes the amino acid sequence <SEQ ID 8254>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4084(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2878

A DNA sequence (GASx1864R) was identified in *S.pyogenes* <SEQ ID 8255> which encodes the amino acid sequence <SEQ ID 8256>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5280(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC36810 GB:L12244 ribosomal protein L28 [Bacillus subtilis]
 Identities = 45/62 (72%), Positives = 52/62 (83%)

Query: 1  MAKVCYFTGRKTVSGNNRSHAMNQTKRTVKPNLQKVTILVDGKPKKVWASARALKSGKVE 60
          MA+ C  TG+KT +GNNRSHAMN +KRT   NLQKV ILV+GKPKKV+ SARALKSGKVE
Sbjct: 1  MARKCVITGKKTTAGNNRSHAMNASKRTWGANLQKVRILVNGKPKKVYVSARALKSGKVE 60

Query: 61 RI 62
          R+
Sbjct: 61 RV 62
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2879

A DNA sequence (GASx1869) was identified in *S.pyogenes* <SEQ ID 8257> which encodes the amino acid sequence <SEQ ID 8258>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1858(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2880

A DNA sequence (GASx1881) was identified in *S.pyogenes* <SEQ ID 8259> which encodes the amino acid sequence <SEQ ID 8260>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2752(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif 136-138
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF04356 GB:AF177167 type IC restriction subunit [Streptococcus thermophilus]
 Identities = 358/1047 (34%), Positives = 571/1047 (54%), Gaps = 91/1047 (8%)

Query: 7    TELELEKELIHLLETGESQWTYRKELKTEDALWDNFFKILAQNNTQYLNEEPLTASEKEQ 66
             +E  +E + I +L   E+QWTYR +LK+E+ALW NF   L + N   L E+PLT  E +Q
Sbjct: 4    SEQMIENQFIQILSEKENQWTYRPDLKSEEALWQNFRSHLNRINLAVLGEQPLTDKEFKQ 63

Query: 67   IKNQLNFVNY--YEAAKWLAGENGIAKVQVQREDAKLGTIRLEVVKADNVAGGTSVYEIA 124
             +K + + +    + A++WL GENG+A++ ++RED K   + LE  +  +++GGTS YE+
Sbjct: 64   VKVEFSRLTGTPFLASQWLRGENGVAQILLEREDGK--RVTLEAFRNKDISGGTSSYEVV 121

Query: 125  NQVAFSGSRDRRGDVTLLINGLPMIQIELKSQNHQ--CIEAFNQVKKYDKEGQFRGIFST 182
             +QV    SR  RGDV+LLINGLP+I IELK ++ +   ++A+ Q+++Y ++G F+GI++T
Sbjct: 122  HQVVPDSSRVDRGDVSLLINGLPIIHIELKKESAKDGFMQAYYQIQRYAEDGFFKGIYAT 181

Query: 183  LQMFVVSNKTDTRYIAAAKENKLNP-----NFLTQWVDQNNKPQKDLFAFAKEVLSIPRA 237
              Q+ V+SNK DTRY A   E+          FL  W  ++N+   DLF F + VL IP A
Sbjct: 182  TQIMVISNKVDTRYFARPSEDTAEAYARMKKFLFNWRTEDNQTVSDLFDFTRTVLRIPDA 241

Query: 238  HQMVMTYSVIDDDKKA---LILLRPYQIHAIEAVAEASRHRKSGYIWHTTGSGKTLTSYK 294
             H+++  Y+++ DD+K    L+ LRPYQIHAI  + + +   + G+IWH TGSGKT+TS+
Sbjct: 242  HELISQYTILVDDQKNQKFLMALRPYQIHAIRKIRQKAAQHEGGFIWHATGSGKTITSFV 301

Query: 295  VARNILQIP-AVEKSIFVIDRKDLDNQTASAFQSYA---------QNDIFD--VDETEDT 342
               + + Q    V++++ V+DR DLD QT   F  +A          +N + +  +   ++
Sbjct: 302  ATKLLAQNAIGVDRTVMVVDRTDLDAQTQDEFTKFASEYHTGQTTENSVANTLIVGIKNQ 361

Query: 343  RQLIKNLESS--DRRVVVTTIQKLNAMISQMESYDTPKFKKLKERLAHLNVVFVVDECHR 400
             +QL +NL SS  +  ++VTTIQKL+A +   +     K     E+L   ++VF+VDE HR
Sbjct: 362  KQLAQNLLSSKNNNTILVTTIQKLSAAMRSAQQESEEKGSNQFEKLRQEHIVFIVDEAHR 421

Query: 401  AVTPERQRYLTNTFRNSRWYGFTGTPIFVENKRAQLGDLAQITEQQYGKCLHQYTVKEAI 460
             AV+ E  + +      NS W+G TGTPIF ENK+ + G  A+TT QQYG  LH YT+K A+
Sbjct: 422  AVSDEEMKRIKKILPNSTWFGLTGTPIFEENKKQENGTFARTTSQQYGPLLHSYTIKNAM 481

Query: 461  HDKAVLGFQVEYKTTIPD--------------MPEDS------IPEEAYDHEEHMLAVLD 500
              D AVLGFQVEY + I +                +P+D+       +P E Y+ +EH+  +L
Sbjct: 482  DDGAVLGFQVEYHSLISEEDQEVIVTQLNKGKLPDDALQQEKLLPTELYETDEHIRTMLQ 541

Query: 501  SIINQSR--KKLGFNNGIGQTFEGLLTVKSIARAQAYYDLMKKVKAGETDLVISKKVKEK 558
              I N+    KK    NG   T   +LT  SIA+A+  Y ++K++K    T L+  ++  E+
Sbjct: 542  KIFNRRSVVKKFKVKNGF-PTMSAILTTHSIAQAKHIYRILKEMKDNGT-LLNGRQFDER 599

Query: 559  L----PDFPKVAITYSITENDNASISRQDKMTKNLEDYNHLFGTNFTIDNLQGYNRDLND 614
                  DFP+VAIT+S   +     + D++ + +++Y   F  +   D  + YN+++N
Sbjct: 600  HQLIDKDFPRVAITFSTNPDQLEKNEQDDELVEIMKEYEKQFDASPYQDE-KLYNQNINK 658

Query: 615  RLARKKDKFKDRHEQLDLVIVVDRLLTGFDAPCLSTIFIDRQPMKPQHIIQAFSRTNRIF 674
```

```
            RLARK+ +++    + LD VIVVDRLLTGFD+P + T++IDR+ M  Q ++QAFSRTNRI+
Sbjct: 659  RLARKEKQYQSDGQWLDFVIVVDRLLTGFDSPTIQTLYIDRE-MNYQKLLQAFSRTNRIY 717

Query: 675  ESRKHYGQVVTFQTPLRFKEAVDKALSLYSNGGEN-DVLAP-SWEEEKARFFEKVTVLKN 732
             + K  G +V+F+ P   +E V     L+SN  +N D L P  +EE K  F E  T+ K
Sbjct: 718  -TGKDSGLIVSFRKPFTMRENVRNTFRLFSNEKQNFDQLIPKEYEEVKKEFIECSTLYKQ 776

Query: 733  IVPDPDAFPTIESAQTAFLKQYAKAFQAFDKLFASVQVYSDFNETLLSEVGLSDEVIDTY 792
               D    P       A +  Y K  +++  L +  Q   DF E   SEV    E +  Y
Sbjct: 777  SEADLSDNPNDLKTMIAQVSAYQKLEKSYKALRSYDQYEEDFEE--FSEV---VEQLPQY 831

Query: 793  KGTYQNVIAEIRKRRED--------DEAIPEINIDYELESVQMDDINYHYILTLIQAFVD 844
             +G  +N+  +I++  ED         ++ + EI    +L +   D ++  YI  L++A
Sbjct: 832  QGKTENIKTKIKEMIEDEGHPEEDFEKLLQEIAFSSQLNATHKDVVDSFYINQLLKAIQL 891

Query: 845  QEQEALQERLNDNPMDQYIQDLAKSNPAMADSLAELWQDIQKEPKAYEGKSIVYELDNLI 904
              E  A+++   +  + Q    + K    + D L    ++I           +   +  I
Sbjct: 892  NEAGAVEK--FEKEIQQKDPQIQKMYHTLKDQLVNTTEEI----------DVAQLKETSI 939

Query: 905  GDKIQRAIKHFADQWKADPDKLAFVATNYHSANSTKQVGMSTLKE-SLDYQAYKEKQGDS 963
              ++IQR ++  A+++    D L      Y S   T       L   +L  + ++ K G+
Sbjct: 940  QNEIQRQLQKEAEEFGLSFDFLQSAMNEYQSDKKTIPYLTHLLDSMTLSKEEFEAKTGE- 998

Query: 964  AMNKLKYKSQFERELVQFIRDQIQPLK 990
               K + +++    E +Q    +Q+Q  K
Sbjct: 999  ---KYRRRTKVLEERLQQNFEQLQKWK 1022
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2881

A DNA sequence (GASx1882) was identified in *S.pyogenes* <SEQ ID 8261> which encodes the amino acid sequence <SEQ ID 8262>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3653(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB53491 GB:U35629 unknown [Lactococcus lactis subsp. lactis]
 Identities = 141/241 (58%), Positives = 178/241 (73%)

Query: 3    KSKQPQYRFDGFEGEWEEKELGDIVQITMGQSPSSQNYTTNPSDYILVQGNADIKNGYVF 62
             K K P+ RF GF  EWE ++LGD V+I MGQSP+S+NYT +P+DYILVQGNAD+KNG V
Sbjct: 13   KKKVPELRFKGFTDEWELRKLGDEVRIVMGQSPNSENYTDDPNDYILVQGNADMKNGRVL 72

Query: 63   PRVWTTQITKQADKGDIILSVRAPVGDVGKTNYHVIIGRGVAAIKGNEFIFQILKYLKEI 122
             PRVWTTQ+TKQA+K D+ILSVRAPVGD+GKT Y V+IGRGVAAIKGNEFIFQ L  +K
Sbjct: 73   PRVWTTQVTKQAEKDDLILSVRAPVGDIGKTAYDVVIGRGVAAIKGNEFIFQNLGKMKSD 132

Query: 123  GYWKRISTGSTFDSISSSDIKYAKIQIPSLPEQEAIGELFQMVDQLIQLQDQKLATLKEQ 182
             GYW R STGSTF+SI+S+DIK A I +P++ EQ+ IG  F+ +D  I L  +KL  LKEQ
Sbjct: 133  GYWTRYSTGSTFESINSTDIKEAIISVPAIEEQDKIGSFFKQLDNTIALHQRKLDLLKEQ 192

Query: 183  KQTFLRKMFPAQGQKVPEIRLQGFKGEWEEKKLREVSTHRSGTAIEKYFDSEGEFKVISIG 243
             K+ FL+KMFP  G KVPE+R  GF  +WEE+KL +++    +G          G++   + G
Sbjct: 193  KKGFLQKMFPKNGAKVPELRFAGFADDWEERKLGDITKISTGKLDANAMVENGKYDFYTSG 253
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2882

A DNA sequence (GASx1883) was identified in *S.pyogenes* <SEQ ID 8263> which encodes the amino acid sequence <SEQ ID 8264>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4318(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF04357 GB:AF177167 type IC modification subunit [Streptococcus thermophilus]
 Identities = 293/523 (56%), Positives = 377/523 (72%), Gaps = 6/523 (1%)

Query: 6   TSLRQALWHSADQLRGQMDANDYKNYLLGLIFYKHLSDKLLLAVCDNLEKHFNTFTEAQK 65
           TSL Q LW SAD LRG+MDA++YKNYLLGLIFYK+LSDK L  V +       +TF E
Sbjct: 3   TSLNQQLWASADILRGKMDASEYKNYLLGLIFYKYLSDKQLREVYEQENGKTDTFPERST 62

Query: 66  I---FEDAYQDEGLKDDLISVVTGDLGYFIEPTLTFEKLIQDVYHNTFQLESLAQGFRDI 122
           +   F + Y+++  KDDLI  +    GYFI+P   F        +  F L  L  GF ++
Sbjct: 63  LYAGFMEWYEED--KDDLIENIQPRQGYFIQPDRLFYHYRIKADNYEFNLTDLQAGFNEL 120

Query: 123 EQSGEDFENLFEDIDLYSKKLGSTPQKQNQTISNVMKTLNEIDFEAVDGDTLGDAYEYLI 182
           E+ GE+F  LF DIDL S KLGS  Q++N TI+ V++ L+EID    +GD +GDAYEYLI
Sbjct: 121 ERQGEEFSGLFSDIDLNSTKLGSNAQQRNVTITEVLRALDEIDLFEHNGDVIGDAYEYLI 180

Query: 183 GEFASESGKKAGEFYTPQAVSHLMTQIVFLGREDQKGMTLYDPAMGSGSLLLNAKKYSNQ 242
           G FA+ +GKKAGEFYTPQAVS +M++I  +G+E +    +YDPAMGSGSL+LN ++Y
Sbjct: 181 GMFAAGAGKKAGEFYTPQAVSRIMSEITSIGQESRVPFHIYDPAMGSGSLMLNIRRYLIH 240

Query: 243 SDTVSYYGQEINTSTYNLARMNMMLHGVAIENQHLSNADTLDADWPTDEPINFDGVLMNP 302
            + V Y+GQE+NT+T+NLARMN++LHGV  E  +L+N DTLDADWP++EP  FD V+MNP
Sbjct: 241 PNQVHYHGQELNTTTFNLARMNLILHGVDKERMNLNNGDTLDADWPSEEPYQFDSVVMNP 300

Query: 303 PYSLKWSATAGFLTDPRFSSYGVLAPKSKADFAFLLHGFYHLKNTGTMAIVLPHGVLFRG 362
           PYS KWSA   FL+DPRF  +G LAPKSKADFAFLLHGFYHLK +GTM IVLPHGVLFRG
Sbjct: 301 PYSAKWSAADKFLSDPRFERFGKLAPKSKADFAFLLHGFYHLKESGTMGIVLPHGVLFRG 360

Query: 363 AAEGKIRQKLLEQGAIDTIIGLPSNIFYNTSIPTTIIILKKNRTNKDVFFIDASKEFDKG 422
            AEG IRQ LLE GAID +IGLP+NIF+ TSIPTT+IILKKNR+ +DV FIDAS++F+K
Sbjct: 361 GAEGTIRQALLEMGAIDAVIGLPANIFFGTSIPTTVIILKKNRSRRDVLFIDASQDFEKQ 420

Query: 423 KNQNTMTDNHIKKILDAYKSRDNSDKFSYLASFDEIIENDYNLNIPRYVDTFEEVPVKPL 482
           KNQN + D HI KI+  YK R++ ++++++ASFDEI END+NLNIPRYVDTFEE     L
Sbjct: 421 KNQNVLLDEHIDKIVSTYKKREDIERYAHVASFDEIQENDFNLNIPRYVDTFEEEEPVDL 480

Query: 483 PELAKQLSDIDQEIAKTNAKLDQLMKQLVGTTKEAQDELDTFR 525
            E+   L  I++E+ +    L  L+      ++E Q  +++ R
Sbjct: 481 VEVNTNLLKINEELVQQEQTLLSLINDF-SESEENQAMIESMR 522
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2883

A DNA sequence (GASx1886R) was identified in *S.pyogenes* <SEQ ID 8265> which encodes the amino acid sequence <SEQ ID 8266>. Analysis of this protein sequence reveals the following:

```
Possible site: 59

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.17   Transmembrane  155 - 171 ( 147 - 173)
   INTEGRAL    Likelihood = -7.22   Transmembrane   14 -  30 (  11 -  33)
   INTEGRAL    Likelihood = -7.17   Transmembrane  182 - 198 ( 179 - 205)
   INTEGRAL    Likelihood = -5.68   Transmembrane  132 - 148 ( 128 - 152)
   INTEGRAL    Likelihood = -4.14   Transmembrane   46 -  62 (  43 -  62)
   INTEGRAL    Likelihood = -3.50   Transmembrane   73 -  89 (  73 -  90)
   INTEGRAL    Likelihood = -0.96   Transmembrane   95 - 111 (  95 - 111)

----- Final Results -----
               bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2884

A DNA sequence (GASx1890R) was identified in *S.pyogenes* <SEQ ID 8267> which encodes the amino acid sequence <SEQ ID 8268>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4757(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif 339-341
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA62650 GB:L37110 clyM [Plasmid pAD1]
 Identities = 127/492 (25%), Positives = 230/492 (45%), Gaps = 30/492 (6%)

Query: 46  KLFYSEFENQLFETIMFLSMKTLVLDINHFSKEIENK----SEAYEQYIQQ-IREENGIN 100
           K F       L + ++ L+ KTLVLD++ F K    K    S+ +  Y+++    +  I
Sbjct: 135 KEFIINLLENLTQELIHLTSKTLVLDLHTFKKNEPLKGNDSSKRFIYYLKKRFNSKKDII 194

Query: 101 HFFDRYPYLLKQINKEVGLIEESYSLLFDRFLEDLSEIKSCFNI-SEPLSNVAFSLGDSH 159
            F+  YP L++     +    ++   + R  EDL  I++CFNI S  L++++ S GDSH
Sbjct: 195 AFYTCYPELMRITVVRMRYFLDNTKQMLIRVTEDLPSIQNCFNIQSSELNSISESQGDSH 254

Query: 160 SKKQTVVKIAFKE-KSVYYKPKSYHSHSILLELTSLLKSSNIPSFSLPKSLVKADYCWQL 218
           S+ +TV  + F + K + YKPK  +S + L +    L          + K + +  Y ++
Sbjct: 255 SRGKTVSTLTFSDGKKIVYKPK-INSENKLRDFFEFLNKELEADIYIVKKVTRNTYFYEE 313

Query: 219 GVAYTSSNK-DEVAKIYFKYGVLAAFSEIFSITDLHMENVIVSGGDLYLIDVETFFQRKL 277
            +     N  +EV K Y +YG L   + +F++TDLH EN+I  G     +ID ETFFQ+ +
Sbjct: 314 YIDNIEINNIEEVKKYYERYGKLIGIAFLFNVTDLHYENIIAHGEYPVIIDNETFFQQNI 373
```

```
Query: 278 NVQNQNFEGITVDTYQRIYETSLSNGLFP---VQFEKNSAPNVSGISRKGGKRQKGKYEL 334
           ++  N    TVD   +  ++ +  GL P   ++ + +S      +S    K Q   +++
Sbjct: 374 PIEFGN--SATVDAKYKYLDSIMVTGLVPYLAMKDKSDSKDEGVNLSALNFKEQSVPFKI 431

Query: 335 I---NKNRGDLKLVKVDYFQEDRFNIPTLNGKVVEPLDYANEIISGFRECYIFLLSQRSK 391
           +   N    +++     +  +   N P +N + +  + Y   I++G +   +     + K
Sbjct: 432 LKIKNTFTDEMRFEYQTHIMDTAKNTPIMNNEKISFISYEKYIVTGMKSILMKAKDSKKK 491

Query: 392 IKEIV-EGFPELKSRVPFRNTSDYGKFLQASTNPKYLFS----EKKRKNLFSILYETKHI 446
           I   +      L  R   R T  Y   L+ S +P    +    EK   N+++  Y+ K +
Sbjct: 492 ILAYINNNLQNLIVRNVIRPTQRYADMLEFSYHPNCFSNAIEREKVLHNMWAYPYKNKKV 551

Query: 447 EHFIVDNEIKDLMNGDIP-YFSMDTRGNVYNSVGTLIGNLGDTTSL---FDSITILNDER 502
            H+    E  DL++GDIP +++   ++ ++  S G L+ +     ++L    + I  L DE
Sbjct: 552 VHY----EFSDLIDGDIPIFYNNISKTSLIASDGCLVEDFYQESALNRCLNKINDLCDED 607

Query: 503 LKFTCELLEIVL 514
           +      LEI L
Sbjct: 608 ISIQTVWLEIAL 619
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2885

A DNA sequence (GASx1891R) was identified in *S.pyogenes* <SEQ ID 8269> which encodes the amino acid sequence <SEQ ID 8270>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3487(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA02867 GB:L07740 salivaricin A [Streptococcus salivarius]
 Identities = 46/51 (90%), Positives = 48/51 (93%)

Query: 1  MSFMKNSKDILTNAIEEVSEKELMEVAGGKKGSGWFATITDDCPNSVFVCC 51
          M+ MKNSKDIL NAIEEVSEKELMEVAGGK+GSGW ATITDDCPNSVFVCC
Sbjct: 1  MNAMKNSKDILNNAIEEVSEKELMEVAGGKRGSGWIATITDDCPNSVFVCC 51
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2886

A DNA sequence (GASx1901R) was identified in *S.pyogenes* <SEQ ID 8271> which encodes the amino acid sequence <SEQ ID 8272>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.59   Transmembrane    3 -  19 (   1 -  20)

----- Final Results -----
              bacterial membrane --- Certainty=0.1638(Affirmative) < succ>
```

```
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2887

A DNA sequence (GASx1905R) was identified in *S.pyogenes* <SEQ ID 8273> which encodes the amino acid sequence <SEQ ID 8274>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.48   Transmembrane   38 -  54 (  37 -  54)

----- Final Results -----
              bacterial membrane --- Certainty=0.1192(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2888

A DNA sequence (GASx1911R) was identified in *S.pyogenes* <SEQ ID 8275> which encodes the amino acid sequence <SEQ ID 8276>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.40   Transmembrane   27 -  43 (  22 -  48)
   INTEGRAL    Likelihood = -9.82   Transmembrane   52 -  68 (  50 -  74)
   INTEGRAL    Likelihood = -7.27   Transmembrane  113 - 129 ( 111 - 134)
   INTEGRAL    Likelihood = -1.97   Transmembrane  137 - 153 ( 135 - 153)

----- Final Results -----
              bacterial membrane --- Certainty=0.5161(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2889

A DNA sequence (GASx1915R) was identified in *S.pyogenes* <SEQ ID 8277> which encodes the amino acid sequence <SEQ ID 8278>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.77   Transmembrane  242 - 258 ( 238 - 262)

----- Final Results -----
               bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2890

A DNA sequence (GASx1918R) was identified in *S.pyogenes* <SEQ ID 8279> which encodes the amino acid sequence <SEQ ID 8280>. Analysis of this protein sequence reveals the following:

```
Possible site: 38

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.32   Transmembrane   40 -  56 (  39 -  60)

----- Final Results -----
               bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2891

A DNA sequence (GASx1923R) was identified in *S.pyogenes* <SEQ ID 8281> which encodes the amino acid sequence <SEQ ID 8282>. Analysis of this protein sequence reveals the following:

```
Possible site: 42

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.26   Transmembrane   20 -  36 (  13 -  42)

----- Final Results -----
               bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2892

A DNA sequence (GASx1926) was identified in *S.pyogenes* <SEQ ID 8283> which encodes the amino acid sequence <SEQ ID 8284>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2322(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2893

A DNA sequence (GASx1928R) was identified in *S.pyogenes* <SEQ ID 8285> which encodes the amino acid sequence <SEQ ID 8286>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3395(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2894

A DNA sequence (GASx1929R) was identified in *S.pyogenes* <SEQ ID 8287> which encodes the amino acid sequence <SEQ ID 8288>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.86   Transmembrane   17 -  33 (  15 -  33)

----- Final Results -----
               bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2895

A DNA sequence (GASx1931R) was identified in *S.pyogenes* <SEQ ID 8289> which encodes the amino acid sequence <SEQ ID 8290>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0551(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2896

A DNA sequence (GASx1941R) was identified in *S.pyogenes* <SEQ ID 8291> which encodes the amino acid sequence <SEQ ID 8292>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2377(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2897

A DNA sequence (GASx1949) was identified in *S.pyogenes* <SEQ ID 8293> which encodes the amino acid sequence <SEQ ID 8294>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0262(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2898

A DNA sequence (GASx1951R) was identified in *S.pyogenes* <SEQ ID 8295> which encodes the amino acid sequence <SEQ ID 8296>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1330(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2899

A DNA sequence (GASx1953) was identified in *S.pyogenes* <SEQ ID 8297> which encodes the amino acid sequence <SEQ ID 8298>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2900

A DNA sequence (GASx1957) was identified in *S.pyogenes* <SEQ ID 8299> which encodes the amino acid sequence <SEQ ID 8300>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2409(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2901

A DNA sequence (GASx1969) was identified in *S.pyogenes* <SEQ ID 8301> which encodes the amino acid sequence <SEQ ID 8302>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.28   Transmembrane    7 -  23 (   7 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2902

A DNA sequence (GASx1971R) was identified in *S.pyogenes* <SEQ ID 8303> which encodes the amino acid sequence <SEQ ID 8304>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1545(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2903

A DNA sequence (GASx1973) was identified in *S.pyogenes* <SEQ ID 8305> which encodes the amino acid sequence <SEQ ID 8306>. Analysis of this protein sequence reveals the following:

```
Possible site: 49

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.44   Transmembrane   31 -  47 (  31 -  48)

----- Final Results -----
               bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB51744 GB:AJ245405 speX [Streptococcus pyogenes]
 Identities = 236/256 (92%), Positives = 243/256 (94%)

Query: 3   MIISFESVILKHNKIITPEKRLFMKKTKLIFSFTSIFIAIISRPVFGLEVDNNSLLRNIY 62
           MIISFESVILKHNKIITPEKRLFMKKTKLIFSFTSIFIAIISRPVFGLEVDNNSLLRNIY
Sbjct: 1   MIISFESVILKHNKIITPEKRLFMKKTKLIFSFTSIFIAIISRPVFGLEVDNNSLLRNIY 60

Query: 63  STIVYEYSDTVIDFKTSHNLVTKKLDVRDARDFFINSEMDEYAANDFKDGDKIAMFSVPF 122
           STIVYEYSD VIDFKTSHNLVTKKLDVRDARDFFINSEMDEYAANDFK GDKIA+FSVPF
Sbjct: 61  STIVYEYSDIVIDFKTSHNLVTKKLDVRDARDFFINSEMDEYAANDFKTGDKIAVFSVPF 120

Query: 123 DWNYLSEGKVIAYTYGGMTPYQEEPMSKNIPVNLWINRKQIPVPYNQISTNKTTVTAQEI 182
           DWNYLS+GKV AYTYGG+TPYQ+    K   VNLWIN KQI VPYN+ISTNKTTVTAQEI
Sbjct: 121 DWNYLSKGKVTAYTYGGITPYQKLQYLKISLVNLWINGKQISVPYNEISTNKTTVTAQEI 180

Query: 183 DLKVRKFLISQHQLYSSGSSYKSGKLVFHTNDNSDKYSLDLFYVGYRDKESIFKVYKDNK 242
           DLKVRKFLI+QHQLYSSGSSYKSG+LVFHTNDNSDKYS DLFYVGYRDKESIFKVYKDNK
Sbjct: 181 DLKVRKFLIAQHQLYSSGSSYKSGRLVFHTNDNSDKYSFDLFYVGYRDKESIFKVYKDNK 240

Query: 243 SFNIDKIGHLDIEIDS 258
           SFNIDKIGHLDIEIDS
Sbjct: 241 SFNIDKIGHLDIEIDS 256
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2904

A DNA sequence (GASx1974R) was identified in *S.pyogenes* <SEQ ID 8307> which encodes the amino acid sequence <SEQ ID 8308>. Analysis of this protein sequence reveals the following:

```
Possible site: 53

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2022(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2905

A DNA sequence (GASx1983) was identified in *S.pyogenes* <SEQ ID 8309> which encodes the amino acid sequence <SEQ ID 8310>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
bacterial cytoplasm --- Certainty=0.0989(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2906

A DNA sequence (GASx1987) was identified in *S.pyogenes* <SEQ ID 8311> which encodes the amino acid sequence <SEQ ID 8312>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2389(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2907

A DNA sequence (GASx1988) was identified in *S.pyogenes* <SEQ ID 8313> which encodes the amino acid sequence <SEQ ID 8314>. Analysis of this protein sequence reveals the following:

```
Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5904(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB16031 GB:AB030747 transposase [Streptococcus pyogenes]
 Identities = 22/24 (91%), Positives = 23/24 (95%)

Query: 1   LERLFGTAKEYHNLCYTREKGKSK 24
           +ERLFGTAKEYHNL YTREKGKSK
Sbjct: 399 IERLFGTAKEYHNLRYTREKGKSK 422
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2908

A DNA sequence (GASx1990R) was identified in *S.pyogenes* <SEQ ID 8315> which encodes the amino acid sequence <SEQ ID 8316>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2909

A DNA sequence (GASx1991) was identified in *S.pyogenes* <SEQ ID 8317> which encodes the amino acid sequence <SEQ ID 8318>. Analysis of this protein sequence reveals the following:

```
Possible site: 53

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.16   Transmembrane    2 -  18 (   1 -  18)

----- Final Results -----
             bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2910

A DNA sequence (GASx1994) was identified in *S.pyogenes* <SEQ ID 8319> which encodes the amino acid sequence <SEQ ID 8320>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.44   Transmembrane   28 -  44 (  28 -  44)

----- Final Results -----
             bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2911

A DNA sequence (GASx1996) was identified in *S.pyogenes* <SEQ ID 8321> which encodes the amino acid sequence <SEQ ID 8322>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1076(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2912

A DNA sequence (GASx1997R) was identified in *S.pyogenes* <SEQ ID 8323> which encodes the amino acid sequence <SEQ ID 8324>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.96   Transmembrane   53 -  69 (  49 -  75)
   INTEGRAL    Likelihood = -2.34   Transmembrane   24 -  40 (  24 -  43)

----- Final Results -----
               bacterial membrane --- Certainty=0.4185(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2913

A DNA sequence (GASx2007R) was identified in *S.pyogenes* <SEQ ID 8325> which encodes the amino acid sequence <SEQ ID 8326>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.64   Transmembrane   46 -  62 (  43 -  65)

----- Final Results -----
               bacterial membrane --- Certainty=0.3654(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB97959 GB:U96166 ATP-binding cassette lipoprotein
          [Streptococcus cristatus]
 Identities = 37/60 (61%), Positives = 42/60 (69%), Gaps = 1/60 (1%)

Query: 59  FLTACGTKKDSKKEEVKEIKMSDIKDDAVSKKTKVVDGEEVTEYTTKDGNVIQIPAGNEE 118
           FL ACG+K    KE + + K  D K DAV +KTK VDG+EVTEYT  DGNVIQIPA  EE
Sbjct: 12  FLAACGSKNADNKE-ISDGKKVDFKKDAVDQKTKTVDGKEVTEYTMPDGNVIQIPADGEE 70
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2914

A DNA sequence (GASx2009) was identified in *S.pyogenes* <SEQ ID 8327> which encodes the amino acid sequence <SEQ ID 8328>. Analysis of this protein sequence reveals the following:

```
Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1246(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2915

A DNA sequence (GASx2010) was identified in *S.pyogenes* <SEQ ID 8329> which encodes the amino acid sequence <SEQ ID 8330>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2549(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2916

A DNA sequence (GASx2012R) was identified in *S.pyogenes* <SEQ ID 8331> which encodes the amino acid sequence <SEQ ID 8332>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3307(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA27007 GB:L26141 pyrogenic exotoxin B [Streptococcus pyogenes]
 Identities = 40/102 (39%), Positives = 57/102 (55%), Gaps = 7/102 (6%)

Query: 2   EMHFVRTEPEARRIAETFCAENTQTKTPMRVQQLSYPSDTDHSGGEL-----YIYALSPA 56
           + +F R E EA+  A TF  ++    K   R  +     D  + GGEL     YIY +S
Sbjct: 28  DQNFARNEKEAKDSAITFIQKSAAIKAGARSAE-DIKLDKVNLGGELSGSNMYIYNISTG 86

Query: 57  GFIIVSGDTRAHTILGYSFDNNLDLN-HDNVRSMIEAYQKQI 97
           GF+IVSGD R+  ILGYS   + D+N  +N+ S +E+Y +QI
Sbjct: 87  GFVIVSGDKRSPEILGYSTSGSFDVNGKENIASFMESYVEQI 128
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2917

A DNA sequence (GASx2013R) was identified in *S.pyogenes* <SEQ ID 8333> which encodes the amino acid sequence <SEQ ID 8334>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2918

A DNA sequence (GASx2014R) was identified in *S.pyogenes* <SEQ ID 8335> which encodes the amino acid sequence <SEQ ID 8336>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1392(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2919

A DNA sequence (GASx2015) was identified in *S.pyogenes* <SEQ ID 8337> which encodes the amino acid sequence <SEQ ID 8338>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.75   Transmembrane   18 -  34 (  17 -  37)

----- Final Results -----
               bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2920

A DNA sequence (GASx2018) was identified in *S.pyogenes* <SEQ ID 8339> which encodes the amino acid sequence <SEQ ID 8340>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.84   Transmembrane   23 -  39 (  22 -  40)

----- Final Results -----
               bacterial membrane --- Certainty=0.3336(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2921

A DNA sequence (GASx2019) was identified in *S.pyogenes* <SEQ ID 8341> which encodes the amino acid sequence <SEQ ID 8342>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0669(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC98898 GB:AF023179 low temperature requirement C protein
           [Listeria monocytogenes]
 Identities = 95/144 (65%), Positives = 117/144 (80%)

Query: 15  LAERGVSLEAIAELVLFLQNDYIPNLTMAECLESVEAVLAKREVQNAIITGVELDKLAEA 74
           L ERGV ++ IAELVLFLQ  Y P L +  C ++VE VL KREVQNA++TG++LD +AE
Sbjct: 16  LIERGVEIDDIAELVLFLQQKYHPGLELDICRQNVEHVLRKREVQNAVLTGIQLDVMAEK 75

Query: 75  NQLSEPLLSILKTDQGLYGIDEILALSIVNLYGSIGFTNYGYLDKTKPGIVDKLNHKDGY 134
            +L +PL +I+  D+GLYG+DEILALSIVN+YGSIGFTNYGY+DK KPGI+ KLN  DG
Sbjct: 76  GELVQPLQNIISADEGLYGVDEILALSIVNVYGSIGFTNYGYIDKVKPGILAKLNEHDGI 135

Query: 135 SCHTFLDDIVSAIAAAAASRIAHN 158
           + HTFLDDIV AIAAAAASR+AH+
Sbjct: 136 AVHTFLDDIVGAIAAAAASRLAHS 159
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2922

A DNA sequence (GASx2030) was identified in *S.pyogenes* <SEQ ID 8343> which encodes the amino acid sequence <SEQ ID 8344>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0320(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2923

A DNA sequence (GASx2031) was identified in *S.pyogenes* <SEQ ID 8345> which encodes the amino acid sequence <SEQ ID 8346>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0583(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2924

A DNA sequence (GASx2032R) was identified in *S.pyogenes* <SEQ ID 8347> which encodes the amino acid sequence <SEQ ID 8348>. Analysis of this protein sequence reveals the following:

```
Possible site: 53

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.76   Transmembrane   27 -  43 (  26 -  43)

----- Final Results -----
               bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8467> and protein <SEQ ID 8468> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:    -11.19
GvH: Signal Score (-7.5): -4.94
     Possible site: 49
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -4.19 threshold:  0.0
   INTEGRAL    Likelihood = -4.19   Transmembrane   25 -  41 (  25 -  42)
   PERIPHERAL  Likelihood = 13.26       41
 modified ALOM score:   1.34

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2678(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01616(304 - 429 of 771)
SP|O06442|SECE_STAAU(7    -    48    of    60)    PREPROTEIN    TRANSLOCASE    SECE    SUBUNIT.
GP|2078376|gb|AAB54017.1||U96619 SecE (Staphylococcus aureus)
%Match = 5.4
%Identity = 26.2  %Similarity = 57.1
Matches = 11  Mismatches = 18  Conservative Sub.s = 13

99        129        159        189        219        249        279        309
RIIQIMLK*HLWRRYGTKESKPSVYRMRKPKLLNRSK*HPQANTTRSK*IL*IL*EVYNTQRNALI*RNKLQKGELIMFV
                                                                              |
                                                                            MAKKESFF

339       369        399        429        459        489        519        549
KGIFQVLRDTTWPNRKQRWKDFISILEYTVFFTIVIYIFDKLLAAGVMDLINRF***IILDRNNPNP*ILLRVFCVENNI
||:   :  |:|| ::: :|  : ::   :|| :  | :|
KGVKSEMEKTSWPTKEELFKYTVIVVSTVIFFLVFFYALDLGITALKNLLFG
          20         30         40         50         60
```

SEQ ID 8468 (GBS396) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 9; MW 35kDa).

GBS396-GST was purified as shown in Figure 217, lane 8.

Example 2925

A DNA sequence (GASx2034R) was identified in *S.pyogenes* <SEQ ID 8349> which encodes the amino acid sequence <SEQ ID 8350>. Analysis of the protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.59   Transmembrane   53 -  69 (  53 -  70)

----- Final Results -----
               bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2926

A DNA sequence (GASx2035) was identified in *S.pyogenes* <SEQ ID 8351> which encodes the amino acid sequence <SEQ ID 8352>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2928(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2927

A DNA sequence (GASx2042R) was identified in *S.pyogenes* <SEQ ID 8353> which encodes the amino acid sequence <SEQ ID 8354>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2547(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2928

A DNA sequence (GASx2043) was identified in *S.pyogenes* <SEQ ID 8355> which encodes the amino acid sequence <SEQ ID 8356>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3289(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2929

A DNA sequence (GASx2049) was identified in *S.pyogenes* <SEQ ID 8357> which encodes the amino acid sequence <SEQ ID 8358>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4014(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2930

A DNA sequence (GASx2052) was identified in *S.pyogenes* <SEQ ID 8359> which encodes the amino acid sequence <SEQ ID 8360>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
```

```
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2931

A DNA sequence (GASx2055R) was identified in *S.pyogenes* <SEQ ID 8361> which encodes the amino acid sequence <SEQ ID 8362>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3048(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05703 GB:AP001513 imidazolonepropionase
        (imidazolone-5-propionate hydrolase) [Bacillus halodurans]
 Identities = 203/416 (48%), Positives = 278/416 (66%), Gaps = 4/416 (0%)

Query: 11  DVLLTHFNQLFCLNDPGHPLTGQEMKKATIVEDGYIAIKDGLIVALGSGEPDAELVGTQT 70
           D LL +  QL  +   G P  G+EM +  ++E   + I+DG +  +G+          Q
Sbjct: 6   DTLLVNIGQLLPMESKG-PKRGKEMSELQLLEHAALGIRDGKVAFIGTMVEADTFTANQM 64

Query: 71  IMRSYKGKIATPGIIDCHTHLVYGGSREHEFAKKLAGVSYLDILAQGGGILSTVRATRSA 130
           I    +GK+ TPG++D HTHL++GGSREHE A K  GV YL+IL  GGGIL+TV ATR+A
Sbjct: 65  I--DCQGKLVTPGLVDPHTHLIFGGSREHEMALKQQGVPYLEILKNGGGILATVEATRAA 122

Query: 131 SFDNLYQKSKRLLDYMLLHGVTTVEAKSGYGLDWETEKRQLDVVAALEKDHPIDLVSTFM 190
           S + L  K+   L+ ML +GVTT+EAKSGYGLD ETE +QL    A+ + HPID+VSTF+
Sbjct: 123 SEEELITKAICHLNRMLSYGVTTIEAKSGYGLDRETEWKQLRAAKAVGEQHPIDIVSTFL 182

Query: 191 AAHAIPEEYKGNPKAYLDVIIKDMLPVVKEENLAEFCDIFCEKNVFTADESRYLLSKAKE 250
            AHAIP  ++ +P  +LD +  DML  +KE+NLAEF DIF E  VFT ++SR  L KAKE
Sbjct: 183 GAHAIPTSHRNDPDRFLDEMA-DMLGEIKEQNLAEFVDIFTETGVFTVEQSRTFLQKAKE 241

Query: 251 MGFKLRIHADEIASIGGVDVAAELSAVSAEHLMMITDDGIAKLIGAGVIGNLLPATTFSL 310
           GF L++HADEI  +GG ++A EL A+SA+HL+  +D GI K+  AG I  LLP TTF L
Sbjct: 242 RGFGLKLHADEIDPLGGAELAGELGAISADHLVGASDQGIQKMAAAGTIACLLPGTTFYL 301

Query: 311 MEDTYAPARKMIDAGMAITLSTDSNPGSCPTANMQFVMQLGCFMLRLTPIEVLNAVTINA 370
           +DTYA AR MID G+A+T+STD NPGS PT N+Q +M +    L++TP E+ +AVT+N
Sbjct: 302 GKDTYARARDMIDQGLAVTISTDFNPGSSPTENLQLIMSIAALRLKMTPEEIWHAVTVNG 361

Query: 371 AYSVNRQERVGSLTVGKEADIAIFDAPNIDYPFYFFATNLIHQVYKKGQLTVDRGR 426
           A+++ R +  G L VG+ AD+ ++DA N  Y  Y +  N +H V+KKG++  +R R
Sbjct: 362 AHAIGRGDTAGQLAVGRAADVVVWDAKNYYYVPYHYGVNHVHSVWKKGEVVYERRR 417
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2932

A DNA sequence (GASx2056) was identified in *S.pyogenes* <SEQ ID 8363> which encodes the amino acid sequence <SEQ ID 8364>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1847(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB61139 GB:AL132952 predicted using Genefinder~cDNA EST
          yk155e6.3 comes from this gene~cDNA EST yk155e6.5 comes
          from this gene~cDNA EST yk156d6.5 comes from this
          gene~cDNA EST yk259b10.3 comes fr
 Identities = 302/649 (46%), Positives = 419/649 (64%), Gaps = 17/649 (2%)

Query: 29  EGIRRAPDRGFRLTQAQTEIALKNALRYVPTKFHEEVIPEFLEELKTRGRIYGYRFRPKD 88
           + +  AP R   LTQ +  +A++NALRY+P + H  +  EF EEL T G IYGYRF P
Sbjct: 85  KNVAHAPKRPCNLTQTEKMLAVRNALRYIPKEHHVLLATEFAEELNTYGHIYGYRFMPNF 144

Query: 89  RIYGKPIDEYKGNCTAAKAMQVMIDNNLSFEIALYPYELVTYGETGSVCANWMQYCLIKK 148
            ++  P+ E   +C  A A+ +MI NNL   +A +P ELVTYG  G V +NW+Q+ L+ +
Sbjct: 145 DLFAPPVSEIGAHCEQASAIILMILNNLDKRVAQFPQELVTYGGNGQVFSNWIQFRLVLR 204

Query: 149 YLEVMTDEQTLVVESGHPVGLFKSKPEAPRVIITNGLLVGEYDNMKDWEIAEEMGVTNYG 208
           YL  MTD QTLV+ SGHP+GLF S P++PR+ +TNG+++  Y   + ++    +GVT YG
Sbjct: 205 YLYTMTDHQTLVLYSGHPLGLFPSTPDSPRMTVTNGMMIPSYSTKELYDKYFALGVTQYG 264

Query: 209 QMTAGGWMYIGPQGIVHGTFNTLLNAGRLKLGVADDGDLTGKLFISSGLGGMSGAQGKAA 268
           QMTAG + YIGPQGIVHGT  T+LNAGR ++G+     L GK+F+++GLGGMSGAQ KAA
Sbjct: 265 QMTAGSFCYIGPQGIVHGTTITVLNAGR-RMGL---DSLAGKVFVTAGLGGMSGAQPKAA 320

Query: 269 EIAKAVAIIAEVDQSRIKTRHSQGWISQIAESPEEALQLAQKAIDAKESTSIAYHGNIVD 328
           +IA  + +IAE+  + +  RH QGW+   ++  EE +   ++  + KE+ SI Y GN+VD
Sbjct: 321 KIAGCIGVIAEISDTALLKRHQQGWLDVYSKDLEEIVNWIKEYREKKEAISIGYLGNVVD 380

Query: 329 LLE-YVNDKQIHVDLLSDQTSCHNVYDGGYCPVGISFDERTRLLAEDKDTFHQMVDDTLA 387
           L E    + +  V+L SDQTS HN + GG+ P G++F++  +++  +DK  F +V ++L
Sbjct: 381 LWERLAEEPECLVELGSDQTSLHNPFLGGFYPAGLTFEQSNQMMTSDPVKFKKLVQNSLI 440

Query: 388 RHFEAIKTLTENGTYFFDYGNAFMKSVYDSGITEISKNGRNDKDGFIWPSYVEDIMGPML 447
           R  AI  +  G YF+DYGNAF+     +G   + ++ ++DK  F +PSY++DIMG  +
Sbjct: 441 RQIAAIDKIAAKGMYFWDYGNAFLLECQRAGANLLREDAQDDK-SFRYPSYMQDIMGD-I 498

Query: 448 FDYGYGPFRWVCLSGNHDDLVATDKAAMEAIDPDR--------RYQDRDNYNWIRDAEKN 499
           F  G+GPFRWVC SG  +DL  TD+ A + ID  +         + Q  DN  WI +AEKN
Sbjct: 499 FSMGFGPFRWVCTSGKPEDLRLTDQTACKIIDELKDTDVPEYVKQQYLDNKKWIEEAEKN 558

Query: 500 QLVVGTQARILYQDCIGRVTIALKFNELVRKGKI-GPVMIGRDHHDVSGTDSPFRETSNI 558
           +LVVG+QARILY D  GRV +A  FNELV+ GK+   ++I RDHHDVSGTDSPFRETSN+
Sbjct: 559 KLVVGSQARILYSDRAGRVALASAFNELVKSGKVSAAIVISRDHHDVSGTDSPFRETSNV 618

Query: 559 KDGSNVTCDMAVQCYAGNAARGMSLVALHNGGGTGIGKAINGGFGLVLDGSERIDEIIKS 618
            DGS  T DMAVQ   G++ RG + VALHNGGG G G  INGGFG+VLDGS       +
Sbjct: 619 YDGSAFTADMAVQNCIGDSFRGATWVALHNGGGVGWGDVINGGFGIVLDGSSDAARRAEG 678

Query: 619 AIAWDTMGGVARRNWARNEHAIETAIEYNRLHAGTDHITIPYLADDDLV 667
           + WD   GV RR+W+ N  A E AI+          +T+P  AD++L+
Sbjct: 679 MLNWDVPNGVTRRSWSGNAKAQE-AIQRAEKQVDGLRVTLPVEADEELL 726
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2933

A DNA sequence (GASx2057) was identified in *S.pyogenes* <SEQ ID 8365> which encodes the amino acid sequence <SEQ ID 8366>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1887(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD35925 GB:AE001751
           formiminotransferase-
           cyclodeaminase/formiminotetrahydrofolate cyclodeaminase,
           putative [Thermotoga maritima]
 Identities = 160/296 (54%), Positives = 214/296 (72%), Gaps = 2/296 (0%)

Query: 3   KIVECIPNFSEGQNQAVIDGLVATAKSIPGVTLLDYSSDASHNRSVFTLVGDDQSIQEAA 62
           K++E +PNFSEG+ + V++ +VA AK     V +LD+S DA HNRSV TLVG+ +++  A
Sbjct: 2   KLIESVPNFSEGRRKEVVEKIVAEAKKYDRVWVLDWSMDADHNRSVITLVGEPENLINAL 61

Query: 63  FQLVKYASENIDMTKHHGEHPRMGATDVCPFVPIKDITTQECVEISKQVAERINRELGIP 122
           F + K A+E ID+  H G+HPRMGA DV P VP+ + T +ECVE SK +  RI  ELGIP
Sbjct: 62  FDMTKKAAELIDLRNHTGQHPRMGAADVIPLVPLYNTTMEECVEYSKILGRRIGEELGIP 121

Query: 123 IFLYEDSATRPERQNLAKVRKGQFEGMPEKLLEEDWAPDYGDRKIHPTAGVTAVGARMPL 182
           ++LYE SATRPERQNLA +RKG+FEG  EK+ +  W PD+G  ++HPTAGVTAVGAR  L
Sbjct: 122 VYLYEKSATRPERQNLADIRKGEFEGFFEKIKDPLWKPDFGPDRVHPTAGVTAVGAREFL 181

Query: 183 VAFNVNLDTDNIDIAHKIAKIIRGSGGGYKYCKAIGVMLEDRHIAQVSMNMVNFEKCSLY 242
           +AFNVNL T ++ IA KIA+ IR S GG +Y KAIGV L+ R + QVS+N+ N +K  LY
Sbjct: 182 IAFNVNLGTRDVKIAEKIARAIRFSSGGLRYVKAIGVDLKGRGVVQVSINITNHKKTPLY 241

Query: 243 RTFETIKFEARRYGVNVIGSEVIGLAPAKALIDVAEYYLQVEDFDYHKQILENHLL 298
           R FE IK EA RYGV V+GSE++GL P ++L+     YYL+ +      K+++E++LL
Sbjct: 242 RVFELIKMEAERYGVPVLGSEIVGLFPLESLLKTVSYYLRTD--LNAKKVIESNLL 295
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2934

A DNA sequence (GASx2058) was identified in *S.pyogenes* <SEQ ID 8367> which encodes the amino acid sequence <SEQ ID 8368>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2776(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA62653 GB:L33465 methenyl tetrahydrofolate cyclohydrolase
           [Methylobacterium extorquens]
 Identities = 79/198 (39%), Positives = 112/198 (55%)

Query: 7   SLTDFAKVLGSDAPAPGGGSAAALSGANGISLTKMVCELTLGKKKYADYQDIITEIHAKS 66
           ++  F   L S AP PGGG AAA+SGA G +L  MVC LT+GKKKY + +  + ++  KS
Sbjct: 6   TIETFLDGLASSAPTPGGGGAAAISGAMGAALVSMVCNLTIGKKKYVEVEADLMQVLEKS 65

Query: 67  TALQASLLAAIDKDTEAFNLVSAVFDMPKETDEDKAARRTAMQKALKTAAQSPFEMMTLM 126
             L+ +L   I  D EAF+ V   + +PK TDE+KAAR   +Q+ALKTA   P     +
Sbjct: 66  EGLRRTLTGMIADDVEAFDAVMGAYGLPKNTDEEKAARAAKIQEALKTATDVPLACCRVC 125

Query: 127 VEALEITATAVGKSNTNAASDLGVAALNLKAGLQGAWLNVLINLSGIKDEDFVTDYRQKG 186
            E +++       K N N  SD GVA L+  AGL+ A LNV +N  G+ D  F  +  ++
Sbjct: 126 REVIDLAEIVAEKGNLNVISDAGVAVLSAYAGLRSAALNVYVNAKGLDDRAFAEERLKEL 185

Query: 187 QALLDKGCHLADDIYTKI 204
           + LL +   L + IY  +
Sbjct: 186 EGLLAEAGALNERIYETV 203
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2935

A DNA sequence (GASx2061) was identified in *S.pyogenes* <SEQ ID 8369> which encodes the amino acid sequence <SEQ ID 8370>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3924(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2936

A DNA sequence (GASx2063) was identified in *S.pyogenes* <SEQ ID 8371> which encodes the amino acid sequence <SEQ ID 8372>. Analysis of this protein sequence reveals the following:

```
Possible site: 57

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.06   Transmembrane  231 - 247 ( 231 - 247)
   INTEGRAL    Likelihood = -0.53   Transmembrane    2 -  18 (   1 -  18)

----- Final Results -----
             bacterial membrane --- Certainty=0.1426(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15971 GB:Z99124 histidase [Bacillus subtilis]
 Identities = 236/477 (49%), Positives = 321/477 (66%), Gaps = 2/477 (0%)

Query: 42  VINLDGESLTIEDVIAIARQGVACHIDDSAIEAVNASRKIVDDIVSEKRVVYGVTTGFGS 101
           ++ LDG SLT  DV  +          + ++E V  SR  V+ IV +++ +YG+ TGFG
Sbjct: 1   MVTLDGSSLTTADVARVLFDFEEAAASEESMERVKKSRAAVERIVRDEKTIYGINTGFGK 60

Query: 102 LCNVSISPEDTVQLQENLIRTHASGFGDPLPEDAVRAIMLIRINSLVKGYSGIRLSTIEK 161
             +V I  ED+  LQ NLI +HA G GDP PE   RA++L+R N+L+KG+SG+R   IE+
Sbjct: 61  FSDVLIQKEDSAALQLNLILSHACGVGDPFPECVSRAMLLLRANALLKGFSGVRAELIEQ 120

Query: 162 LLELLNKGVHPYIPEKGSLGASGDLAPLAHMVLPMLGLGKAYYKGELLSGQEALDKAGID 221
           LL  LNK VHP IP++GSLGASGDLAPL+H+ L ++G G+ +++GE +     L KAGI
Sbjct: 121 LLAFLNKRVHPVIPQQGSLGASGDLAPLSHLALALIGQGEVFFEGERMPAMTGLKKAGIQ 180

Query: 222 KISLAAKEGLALINGTTVLTAVGALATYDAIQLLKLSDLAGALSLEVHNGITSPFEENLH 281
            ++L +KEGLALINGT  +TA+G +A  +A +L    ++  +L++E   GI   F+E++H
Sbjct: 181 PVTLTSKEGLALINGTQAMTAMGVVAYIEAEKLAYQTERIASLTIEGLQGIIDAFDEDIH 240

Query: 282 TIRPQSGQLATARNIRNLLEGSQNTTVATQSRVQDPYTLRCMPQIHGASKDSIAYVKSKV 341
             R    Q+  A  IR  L  S  TT   + RVQD Y+LRC+PQ+HGA+  ++ YVK K+
Sbjct: 241 LARGYQEQIDVAERIRFYLSDSGLTTSQGELRVQDAYSLRCIPQVHGATWQTLGYVKEKL 300

Query: 342 DIEINSVTDNPIICKDG-HVISGGNFHGEPMAQPFDFLGIAISEIGNVSERRVERLVNSQ 400
           +IE+N+ TDNP+I  DG  VISGGNFHG+P+A   DFL IAISE+ N++ERR+ERLVN Q
Sbjct: 301 EIEMNAATDNPLIFNDGDKVISGGNFHGQPIAFAMDFLKIAISELANIAERRIERLVNPQ 360

Query: 401 LSKLPSFLVKYPGLNSGFMITQYACASLASENKVLAHPASVDSIPSCENQEDFVSMGTTA 460
           L+ LP FL  +PGL SG MI QYA ASL SENK LAHPASVDSIPS  NQED VSMGT A
Sbjct: 361 LNDLPPFLSPHPGLQSGAMIMQYAAASLVSENKTLAHPASVDSIPSSANQEDHVSMGTIA 420

Query: 461 ARKAFEILKNSRRIVATEIMAACQALDLKPENHELGKGTKVAYDLFRKEVNFIEHDK 517
           AR A++++ N+RR++A E + A QA++ +   H     TK  +  RK V  I+ D+
Sbjct: 421 ARHAYQVIANTRRVIAIEAICALQAVEYRGIEH-AASYTKQLFQEMRKVVPSIQQDR 476
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2937

A DNA sequence (GASx2064) was identified in *S.pyogenes* <SEQ ID 8373> which encodes the amino acid sequence <SEQ ID 8374>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4483(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG06563 GB:AE004741 probable arginase family protein
        [Pseudomonas aeruginosa]
 Identities = 99/275 (36%), Positives = 147/275 (53%), Gaps = 9/275 (3%)
```

```
Query: 53  LIGFKSDKGVYINNGRVGAVESPAAIRTQLAKFPWHLGNQVMVYDVGNIDGPNRSLEQLQ 112
           L+GF SD+GV  N GR GA   P A+R  LA   WH G Q  +YD G+I   +  LE  Q
Sbjct: 42  LLGFASDEGVRRNQGRQGARHGPPALRRALANLAWH-GEQA-IYDAGDIVAGD-DLEAAQ 98

Query: 113 NSLSKAIKRMCDLNLKPIVLGGGHETAYGHYLGLRQSLSPSDDL---AVINMDAHFDLRP 169
              ++ +  +     + + LGGGHE AY  + GL + LS  + L    ++N DAHFDLR
Sbjct: 99  ECYAQRVADLLACGHRVVGLGGGHEIAYASFAGLARHLSRHERLPRIGILNFDAHFDLRH 158

Query: 170 YDQTGPNSGTGFRQMFDDAVADKRLFKYFVLGIQEHNNNLFLFDFVAKSKGIQFLTGQDI 229
            ++   +SGT FRQ+ +   A    F Y  LGI   +N   LFD  A+  G+++L  + +
Sbjct: 159 AERA--SSGTPFRQIAELCQASDWPFAYCCLGISRLSNTAALFD-QAQRLGVRYLLDRQL 215

Query: 230 YQMGHQKVCRAIDRFLEGQERVYLTIDMDCFSVGAAPGVSAIQSLGVDPNLAVLVLQHIA 289
               ++    +D FL+  + +YLT+ +D      APGVSA  + GV+  +   +++
Sbjct: 216 QPWNLERSEAFLDGFLQSVDHLYLTVCLDVLPAAQAPGVSAPSAHGVEMPVVEHLVRRAK 275

Query: 290 ASGKLVGFDVVEVSPPHDIDNHTANLAATFIFYLV 324
           ASGKL   D+ E++P  D D  TA +AA  +  LV
Sbjct: 276 ASGKLRLADIAELNPQLDSDQRTARIAARLVDSLV 310
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2938

A DNA sequence (GASx2065R) was identified in *S.pyogenes* <SEQ ID 8375> which encodes the amino acid sequence <SEQ ID 8376>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane  375 - 391 ( 375 - 392)

----- Final Results -----
               bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB37582 GB:AL035569 putative regulatory protein [Streptomyces
           coelicolor A3(2)]
 Identities = 95/437 (21%), Positives = 177/437 (39%), Gaps = 28/437 (6%)

Query: 271 EVGALLLIGDTGIGKRTLARQVLANQTQTFQIVTAKCFREEAMDSL--LPWRNILDGLGD 328
           E  ALLL G+ G+GK  L  +  A   +   +V      E   D L   P+   L  L
Sbjct: 95  EPQALLLGGEAGVGKTRLVEEFAAAADRRGAVVALGGCVEIGADGLPFAPFSTALRALRR 154

Query: 329 LVIQNRLLTTKAWKAALKRCFP-VATIFQEDNNQPFIKDHTSLLVSFIVDILQHLAEIKA 387
            + +         +  L R  P +A            ++  + L      +L+ +A
Sbjct: 155 HLPEELAAAAAGQEEELARLLPELAEGTPVTGGGRHDEESMARLFELTARLLERVAARHT 214

Query: 388 LVILIEDCHWMDEDSLTLLQRVMNQLVHYPIAFVLT-------KHLGTTPELGLCLNALM 440
           +V+++ED HW D  +  L+  ++  L    +  + T       +     P L   L+ L
Sbjct: 215 VVLVLEDLHWADASTRHLIAYLLRTLRTGRLVVLATYRSDDIHRRHPLRPLLAE-LDRLR 273

Query: 441 SQGRLESICLEPFNRQESLVYINSQLGSQPVTAEEMEHLYQASQGNPFFLSEYTQALLRH 500
           +  RLE   L  F R E    I   L  +P    +++ +++ S GN FF+ E   A  R
Sbjct: 274 TVRRLE---LGRFTRDEVGRQIAGILAHEP-DQLQVDEIFERSDGNAFFVEELAVA-ARV 328

Query: 501 EKFVPLTPAIKAKLGLKLANLSSRDDALLNYLSCCRRPIPLNTLAQLMLLPLEEVIEMVD 560
               LT +++  L +++  L      +   ++      +    LA +  L  +++IE +
Sbjct: 329 GSCTGLTDSLRDLLLVRVEALPESAQRVARIVAEGGSTVEYRLLAAVARLAEDDLIEALR 388
```

```
Query: 561 NLGHYYILVEESVGEEVLISFRQRIIQLYSYDRLSLSKRRLLHGQIAKRLEDLLPILTPS 620
           +  +  IL+    G+     FR  +++    D L   +R  L+ + A+ L D  P L P+
Sbjct: 389 SAVNANILLPAPDGDG--YRFRHSLVREAVGDDLLPGERSRLNRRYAEAL-DADPTLVPA 445

Query: 621 PHLLDDIAYHYQESRQVIKALEYNLNYLDATLPFQHELFPIYSKSIGSLEKSDRDHQRLM 680
              +  +A ++  +    KAL      LDA++  +      YS+ +  LE++       L
Sbjct: 446 AERVMRLASYWYHAHAPAKALP---AVLDASVEARRR--HAYSEQLRLLERA----MELW 496

Query: 681 EEQFDKIRQSIADLELT 697
           +   D +R ++  ++ T
Sbjct: 497 DSAPDDVRATLRPVDCT 513
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2939

A DNA sequence (GASx2072) was identified in *S.pyogenes* <SEQ ID 8377> which encodes the amino acid sequence <SEQ ID 8378>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3702(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2940

A DNA sequence (GASx2074R) was identified in *S.pyogenes* <SEQ ID 8379> which encodes the amino acid sequence <SEQ ID 8380>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.90   Transmembrane   21 -  37 (  21 -  38)

----- Final Results -----
               bacterial membrane --- Certainty=0.1362(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2941

A DNA sequence (GASx2075R) was identified in *S.pyogenes* <SEQ ID 8381> which encodes the amino acid sequence <SEQ ID 8382>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3545(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2942

A DNA sequence (GASx2076R) was identified in *S.pyogenes* <SEQ ID 8383> which encodes the amino acid sequence <SEQ ID 8384>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2340(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC44494 GB:U44893 orf108; unknown function [Butyrivibrio
          fibrisolvens]
 Identities = 42/75 (56%), Positives = 55/75 (73%)

Query: 1   LLKGTLRFGQLKSSIGSVSQKVLTAQLRAMEADGLVHREVYAEVPPRVEYSLTETGLSLA 60
           LL    RF +LK+++  +SQKVLT  LR+ME DG++ R VY EVPPRVEYSL+E G S+
Sbjct: 31  LLVRPWRFNELKNNLEGISQKVLTDSLRSMEEDGIITRTVYPEVPPRVEYSLSELGESMR 90

Query: 61  PVIEAMSDWGQTYQE 75
           P+I+AM  WG  Y+E
Sbjct: 91  PIIKAMEQWGTEYKE 105
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2943

A DNA sequence (GASx2097) was identified in *S.pyogenes* <SEQ ID 8385> which encodes the amino acid sequence <SEQ ID 8386>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.40   Transmembrane   26 -  42 (  23 -  44)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.2359(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2944

A DNA sequence (GASx2098) was identified in *S.pyogenes* <SEQ ID 8387> which encodes the amino acid sequence <SEQ ID 8388>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1385(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2945

A DNA sequence (GASx2100) was identified in *S.pyogenes* <SEQ ID 8389> which encodes the amino acid sequence <SEQ ID 8390>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2138(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA98589 GB:L44593 ORF79; putative [Lactococcus lactis phage
        BK5-T]
 Identities = 34/62 (54%), Positives = 44/62 (70%)

Query: 3  QITLKAARINAGYTLKQVAGAVGKNPQTISKYEKDSSDISLGLLQKLSSLYGVTIDNLFL 62
          +I LKAAR NA ++ K+VA  VGKN QTI  YEKDS++I + L  KL+ +Y   ID +FL
Sbjct: 8  KIKLKAARTNADFSAKEVAEIVGKNYQTILSYEKDSTEIPMSLAIKLAEIYDYPIDFIFL 67

Query: 63 GK 64
          GK
Sbjct: 68 GK 69
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2946

A DNA sequence (GASx2103) was identified in *S.pyogenes* <SEQ ID 8391> which encodes the amino acid sequence <SEQ ID 8392>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3316(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2947

A DNA sequence (GASx2104) was identified in *S.pyogenes* <SEQ ID 8393> which encodes the amino acid sequence <SEQ ID 8394>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4371(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2948

A DNA sequence (GASx2105) was identified in *S.pyogenes* <SEQ ID 8395> which encodes the amino acid sequence <SEQ ID 8396>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2263(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2949

A DNA sequence (GASx2106) was identified in *S.pyogenes* <SEQ ID 8397> which encodes the amino acid sequence <SEQ ID 8398>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.42   Transmembrane    9 -  25 (   6 -  29)

----- Final Results -----
               bacterial membrane --- Certainty=0.3569(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2950

A DNA sequence (GASx2107) was identified in *S.pyogenes* <SEQ ID 8399> which encodes the amino acid sequence <SEQ ID 8400>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1355(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2951

A DNA sequence (GASx2108) was identified in *S.pyogenes* <SEQ ID 8401> which encodes the amino acid sequence <SEQ ID 8402>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3050(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2952

A DNA sequence (GASx2109) was identified in *S.pyogenes* <SEQ ID 8403> which encodes the amino acid sequence <SEQ ID 8404>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3628(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB46557 GB:AJ242479 putative replication protein [Streptococcus thermophilus]
 Identities = 143/242 (59%), Positives = 180/242 (74%), Gaps = 2/242 (0%)

Query: 1   MAIYEARGFSSYLY--PYKGPLEPFDYIAQFRPLKPPEDIDIEEYKRTQAPYCLSGKVTA 58
           MAIYE+RGF + L+         +PF ++A FRP+K P+  DI ++KR  APYC+SG+V
Sbjct: 1   MAIYESRGFGNILHLNNSNASKDPFKFVATFRPMKVPQGEDIADFKRYHAPYCISGEVKQ 60

Query: 59  EKNGSYKRNNASLVYRDLIFLDYDEIETGVNLPKIVSQTLWEYSYIIYPTIKHTPEKPRY 118
           +++G+YKRNNASL+YRDLIFLDYD++E   + P+ VS  L  YSY+IYPTIKHT EKPRY
Sbjct: 61  DEDGNYKRNNASLLYRDLIFLDYDKLEASTDFPRAVSNALNGYSYVIYPTIKHTAEKPRY 120

Query: 119 RLVMKPSDVMTEATYKQVVKEIADKIGLPFDLASLTWSQLQGLPVTTGDPEDYQRYVNHG 178
           RLV+KP+D M E TYK   +EIADKIGLPFD +SLTWSQLQGLPVTTGDPE Y+R VN G
Sbjct: 121 RLVVKPTDKMDEQTYKATAQEIADKIGLPFDDSSLTWSQLQGLPVTTGDPEKYERIVNRG 180

Query: 179 LDYPVPKNGSTPNRQVVTTYTPRPRSQRSITMRVIDTLFNGFGNEGGRNVALTKFVGLLF 238
             YPV    +          +TPR    +S+TMRV+DTL NGFG+EGGRN+ +T+FVGLL
Sbjct: 181 RCYPVANPNTVKANHSPNYHTPRQSGDKSLTMRVVDTLLNGFGDEGGRNIEVTRFVGLLL 240

Query: 239 NK 240
           +K
Sbjct: 241 SK 242
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2953

A DNA sequence (GASx2110) was identified in *S.pyogenes* <SEQ ID 8405> which encodes the amino acid sequence <SEQ ID 8406>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5215(Affirmative) < succ>
```

```
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB46558 GB:AJ242479 putative DNA primase [Streptococcus thermophilus]
 Identities = 274/548 (50%), Positives = 363/548 (66%), Gaps = 17/548 (3%)

Query: 17  DLKNLENEITEARE------NEDKYFSTFKGVRGQLIKECQEMKDEAFKIAYDGVMADSK 70
           DL  LE E  E+++       +ED Y  TFK +R Q I   ++ K+ A++  YD  M + K
Sbjct: 8   DLTKLEEEYNESKKEASTLFDEDGYLKTFKDIRKQFINILEQKKEIAYQKGYDLYMNNPK 67

Query: 71  HLENVKAGRLTEVQHE-------ELAKEKGQEASEKALPKTPLGVAIMLKHYLRFIRVKP 123
            L  +      E   E       E AK++G++A + A PKTPL  A  LK Y+RFIR++P
Sbjct: 68  VLLKLAKAEKDEENGELIRKTVIEDAKKEGEKAKKNATPKTPLECAEFLKKYIRFIRIRP 127

Query: 124 EAQGQKAPLYFFHPDHGVWLEDNEFLQDLISVIFPNATEKQAFDTLYKIARQSQLKEIQR 183
           + +G++    F     G++LED+EFL DL+  I PN TE+   D LYKIA    LK+ Q
Sbjct: 128 KGKGRERLYTFTRQILGIYLEDDEFLHDLMVTIHPNNTERLGNDALYKIAHSVPLKDKQE 187

Query: 184 EYTVIGNQLYNYKTGQFEELTPDITVTRKIKTGYNKKAKEPTIKGWKPTAWLLELFDGDA 243
            Y V+G +LYN +TG+F +  P I VTRK++ GYN  A EP I GWKPT WL  LF+GD
Sbjct: 188 NYVVVGGELYNNETGEFTQFDPRIIVTRKVRMGYNPDATEPIIDGWKPTVWLKGLFNGDR 247

Query: 244 ELYNLAIQIIKASITGQSLQKIFWLFGEGGTGKGTFQQLLINLVGMDNVASLKITELAKS 303
           + Y+LAIQII+A+ITG++L+ IFWL+GEGGTGKGTFQ LL NLVG +NVAS KI + A
Sbjct: 248 DSYDLAIQIIRATITGKTLENIFWLYGEGGTGKGTFQTLLENLVGSENVASFKI-DGASG 306

Query: 304 RFTTSILLGKSIVIGDDIQKDAVIKDTSDIFSLATGDIMTIEDKGKRPYSIRLNMTVVQS 363
           +F TSIL+GK++VIGDDIQKD VIKDTS +FSLATGD + IEDKGKRPY+ R  MTVVQS
Sbjct: 307 KFDTSILIGKTVVIGDDIQKDVVIKDTSVVFSLATGDPIRIEDKGKRPYTTRKRMTVVQS 366

Query: 364 SNGLPRMNGDKSAIDRRFRILPFTKVFKGKPNKAIRNDYINRKEVLEYLLKLAIETPITD 423
           SNG PRMN D+ AI+RRFR+L F+++ KGK +K I+NDY+ RKEVLEY +KLAIETP  D
Sbjct: 367 SNGFPRMNADQKAINRRFRVLTFSEL-KGKADKRIKNDYVGRKEVLEYFVKLAIETPFRD 425

Query: 424 INPKASIEILEEHHKEMNPVIDFVSKFFTDE-LTSEFIPNSFVYHVWKGFLEYYDIKQ-I 481
           +NP+ SIE L+E +KEMNPV DFV +FF DE +   ++PN +V+  +K + E     +
Sbjct: 426 VNPQKSIEFLDEAYKEMNPVADFVDRFFNDEVIKCNYVPNGYVFECFKAYCEKNQNRNYF 485

Query: 482 KSERGLHKEIKSNLPEGFEAGQKVIPVGRQLHTGFYPKEDLPLFASASYANGRASPEKRK 541
            + R LHK+IK  LP+ F   +  I  G++ +  F P      +Y NGR   E ++
Sbjct: 486 LNSRTLHKQIKKILPKTFRPKEVTIKKGQKFYEEFNPHLVSNPWHFDAYDNGRNKKEDQQ 545

Query: 542 KPKNERGY 549
             K ERGY
Sbjct: 546 DAKKERGY 553
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2954

A DNA sequence (GASx2111) was identified in *S.pyogenes* <SEQ ID 8407> which encodes the amino acid sequence <SEQ ID 8408>. Analysis of this protein sequence reveals the following:

```
Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0994(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2955

A DNA sequence (GASx2112) was identified in *S.pyogenes* <SEQ ID 8409> which encodes the amino acid sequence <SEQ ID 8410>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3058(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2956

A DNA sequence (GASx2114) was identified in *S.pyogenes* <SEQ ID 8411> which encodes the amino acid sequence <SEQ ID 8412>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2815(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2957

A DNA sequence (GASx2115R) was identified in *S.pyogenes* <SEQ ID 8413> which encodes the amino acid sequence <SEQ ID 8414>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2958

A DNA sequence (GASx2116) was identified in *S.pyogenes* <SEQ ID 8415> which encodes the amino acid sequence <SEQ ID 8416>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4213(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2959

A DNA sequence (GASx2117) was identified in *S.pyogenes* <SEQ ID 8417> which encodes the amino acid sequence <SEQ ID 8418>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3091(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2960

A DNA sequence (GASx2118) was identified in *S.pyogenes* <SEQ ID 8419> which encodes the amino acid sequence <SEQ ID 8420>. Analysis of this protein sequence reveals the following:

```
Possible site: 41

>>> Seems to have an uncleavable N-term signal seq
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2961

A DNA sequence (GASx2119) was identified in *S.pyogenes* <SEQ ID 8421> which encodes the amino acid sequence <SEQ ID 8422>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2531(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF63071 GB:AF158600 gp137 [Streptococcus thermophilus
          bacteriophage Sfi11]
 Identities = 41/121 (33%), Positives = 65/121 (52%), Gaps = 3/121 (2%)

Query: 4   KNAIRKLKEFHRWQRIAN-SLDLTYTELYQFDIEYHPTRR--KHLEISRECALEELDAIR 60
           K   RKL+E+ RW+ IA+ S +   T+ + F        +  +++ + R  AL EL+AI
Sbjct: 13  KRCKRKLREYPRWREIAHDSAEQKITQEFTFMPRGGGVNKPVENIAVRRVDALNELEAIE 72

Query: 61  YAINQLSKVEYRQILIECYLISEEKTQQDIMEELNGSQSWYYESKKRALLEFVEFYRDGAL 121
            A+N L + +YR+ILIE YL    K    I + +   ++ + E    ++L F E YRDG L
Sbjct: 73  QAVNGLYRPDYRRILIEKYLAYPPKPNWQIAQSIGFERTAFQELLNNSILAFAELYRDGRL 133
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2962

A DNA sequence (GASx2120) was identified in *S.pyogenes* <SEQ ID 8423> which encodes the amino acid sequence <SEQ ID 8424>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2666(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2963

A DNA sequence (GASx2121) was identified in *S.pyogenes* <SEQ ID 8425> which encodes the amino acid sequence <SEQ ID 8426>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2964

A DNA sequence (GASx2123R) was identified in *S.pyogenes* <SEQ ID 8427> which encodes the amino acid sequence <SEQ ID 8428>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3441(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2965

A DNA sequence (GASx2132) was identified in *S.pyogenes* <SEQ ID 8429> which encodes the amino acid sequence <SEQ ID 8430>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2966

A DNA sequence (GASx2136) was identified in *S.pyogenes* <SEQ ID 8431> which encodes the amino acid sequence <SEQ ID 8432>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.19   Transmembrane   57 -  73 (  54 -  78)

----- Final Results -----
               bacterial membrane --- Certainty=0.2275(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB18271 GB:U74623 CadX [Staphylococcus lugdunensis]
 Identities = 50/110 (45%), Positives = 76/110 (68%)

Query: 11  MKKDSICQVGVINQQNVTTATNYLEKEKVQKSLRILSKFTDNKQINIIFYLLAVEELCVC 70
           M  ++ C V  +++  V  A ++LE +K +K L IL K  D K++ II  L+  +ELCVC
Sbjct: 1   MSYENACDVICVHEDKVNNALSFLEDDKSKKLLNILEKICDEKKLKIILSLIKEDELCVC 60

Query: 71  DIACLLNLSMASASHHLRKLANQNILDTRREGKIIYYFIKDEEIRDFFNQ 120
           DI+ +L +S+AS SHHLR L   ++LD  ++GK+ YYFIKD+EIR+FF++
Sbjct: 61  DISLILKMSVASTSHHLRLLYKNDVLDFYKKGKMAYYFIKDDEIREFFSK 110
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2967

A DNA sequence (GASx2137) was identified in *S.pyogenes* <SEQ ID 8433> which encodes the amino acid sequence <SEQ ID 8434>. Analysis of this protein sequence reveals the following:

```
Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4582(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2968

A DNA sequence (GASx2139) was identified in *S.pyogenes* <SEQ ID 8435> which encodes the amino acid sequence <SEQ ID 8436>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -5.89   Transmembrane   63 -  79 (  54 -  80)

----- Final Results -----
               bacterial membrane --- Certainty=0.3357(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2969

A DNA sequence (GASx2141R) was identified in *S.pyogenes* <SEQ ID 8437> which encodes the amino acid sequence <SEQ ID 8438>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4663(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2970

A DNA sequence (GASx2142) was identified in *S.pyogenes* <SEQ ID 8439> which encodes the amino acid sequence <SEQ ID 8440>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.08   Transmembrane  143 - 159 ( 135 - 165)
   INTEGRAL    Likelihood = -7.64   Transmembrane   53 -  69 (  49 -  79)
   INTEGRAL    Likelihood = -7.17   Transmembrane  252 - 268 ( 248 - 275)
   INTEGRAL    Likelihood = -6.74   Transmembrane  186 - 202 ( 183 - 208)
   INTEGRAL    Likelihood = -5.63   Transmembrane  220 - 236 ( 218 - 240)
   INTEGRAL    Likelihood = -5.26   Transmembrane  116 - 132 ( 115 - 136)
   INTEGRAL    Likelihood = -2.02   Transmembrane   85 - 101 (  85 - 101)
   INTEGRAL    Likelihood = -0.64   Transmembrane  165 - 181 ( 165 - 181)

----- Final Results -----
               bacterial membrane --- Certainty=0.5034(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD35257 GB:AE001701 conserved hypothetical protein [Thermotoga maritima]
 Identities = 81/275 (29%), Positives = 137/275 (49%), Gaps = 29/275 (10%)

Query: 9   FKGMIIALGFILPGVSGGVLAAILGIYERMISFLAHMRDNFIENVLFFLPVGIG---GIL 65
           F G+++ +  ++PGVSGG +A ++G+YE++I  +          ++   +PVG G   G+
Sbjct: 7   FSGVLMGIANVVPGVSGGTIAVLMGVYEKLIESVNSFFHGNSRSLKVLIPVGAGVLVGVF 66

Query: 66  GIALFSFPVEFLLKHYQVSVLWGFAGAIVGTIPSLIKESTKQSQRDKADWLWLVLTFVIS 125
           GIA F   +E  L  Y V   + F G I   I S +K  TK+    K     + + FV+
Sbjct: 67  GIARF---LEIFLSKYPVPTHFFFLGLI---IVSFVK--TKEYFSIKP----VNIFFVLL 114

Query: 126 GLGLYFLNDLIG--TLPANFLTFILAGALIALGVLVPGLSPSNLLLILGLYGPMLIGFKS 183
           G+ L F+    G  T   +    +L G + A  ++VPG+S S +LLI G+Y  +L
Sbjct: 115 GMFLIFMLHFSGETTAKESMFLLVLGGFVAATAMVVPGISGSLILLIFGVYDHVLYLVSH 174

Query: 184 LDLLGTFLPIAIGGVLAILAFSKSMDYALQHHHSKVYHFIIGIVLSSTLLILIPNSSSPE 243
           L ++G  L  +IG V  IL   K M++ L+    + Y FI G++L+S L  ++P   +
Sbjct: 175 L-IIGELLIFSIGVVAGILVSVKIMNFLLKRFREETYSFIGGMILAS-LYEVLPKKMNTN 232

Query: 244 SISYSHAGILTWLMAFVLFALGIWLGLWMSQLEEK 278
            +         L + +   L + LG ++  +E+K
Sbjct: 233 VV----------LPSVLSLVLSLTLGFFLLYIEKK 257
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2971

A DNA sequence (GASx2143R) was identified in *S.pyogenes* <SEQ ID 8441> which encodes the amino acid sequence <SEQ ID 8442>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3964(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05000 GB:AP001511 unknown conserved protein in others
           [Bacillus halodurans]
 Identities = 28/78 (35%), Positives = 37/78 (46%)

Query: 44  EVDKVFIVPLRQLLFTDPVYYRLEVTPIETTDFPFDRIRNGKYYQFSQEYRSIPFYENLE 103
           EVD VF VP+   +   P  YR+ V       FP +RI N   YQ S    +  FY
Sbjct: 127 EVDHVFTVPIDHFISHPPEQYRINVHFEPGAGFPIERIANQSAYQKSTRQITESFYYYQS 186

Query: 104 ETIWGMTAQFTKCLTDIL 121
             IWG+TA+  + +  IL
Sbjct: 187 YVIWGLTAKILRHVITIL 204
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2972

A DNA sequence (GASx2144R) was identified in *S.pyogenes* <SEQ ID 8443> which encodes the amino acid sequence <SEQ ID 8444>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4761(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2973

A DNA sequence (GASx2145) was identified in *S.pyogenes* <SEQ ID 8445> which encodes the amino acid sequence <SEQ ID 8446>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -4.09   Transmembrane    2 -  18 (   1 -  19)

----- Final Results -----
               bacterial membrane --- Certainty=0.2635(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA49519 GB:X69895 X [Bacillus sphaericus]
 Identities = 40/97 (41%), Positives = 57/97 (58%), Gaps = 5/97 (5%)

Query: 10  IEFLILAIVEKNDSYGYDISQTIKLVAN----IKESTLYPILKKLEKAGFLTTYSQE-HQ 64
           ++ +IL ++ + D YGY+ISQ I    N    IKE+TLY + ++LEK   +  Y  +
Sbjct: 11  LDSIILRLILEKDRYGYEISQEISNRTNNSFQIKEATLYAVFQRLEKKEVIEAYYGDVSD 70

Query: 65  GRKRKYYAVTSSGRAQLIFLKKEWQSYKFALDGIIEG 101
           G KRKYY +TS G+A L  L KEW   K  +D  +EG
Sbjct: 71  GGKRKYYRITSLGKAYLSELVKEWAEVKEIIDLFMEG 107
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2974

A DNA sequence (GASx2146) was identified in *S.pyogenes* <SEQ ID 8447> which encodes the amino acid sequence <SEQ ID 8448>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.75   Transmembrane   97 - 113 (  77 - 143)
```

```
   INTEGRAL    Likelihood = -6.85   Transmembrane  116 - 132 ( 114 - 143)
   INTEGRAL    Likelihood = -5.68   Transmembrane  156 - 172 ( 149 - 175)
   INTEGRAL    Likelihood = -5.47   Transmembrane   79 -  95 (  77 -  96)

----- Final Results -----
              bacterial membrane --- Certainty=0.6901(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2975

A DNA sequence (GASx2147) was identified in *S.pyogenes* <SEQ ID 8449> which encodes the amino acid sequence <SEQ ID 8450>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.11   Transmembrane    8 -  24 (   6 -  30)

----- Final Results -----
              bacterial membrane --- Certainty=0.3845(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF04457 GB:AF078161 lacunin [Manduca sexta]
 Identities = 68/310 (21%), Positives = 117/310 (36%), Gaps = 12/310 (3%)

Query: 55   DIDSSASTITVETGPVQRPTVTYYTHPKLIDPIVTTVTGKTLSLSQTPKDVVITGGIEIL 114
            DI+ + ++ + E+      T++  T    +    TT T  T  +S T +   I      +
Sbjct: 1004 DIEGTTASGSTESTFTDETTMSKVTEESSVAEEETTKTTITEEVSGTSESASINSDKTTM 1063

Query: 115  GFTLNNSRQEKNYRSIT--ITVPEKTSLNEVKASNVPHTTLSNLT--VQDMQFDGNLTLL 170
                 ++ +       IT  +TV E+TS           TT+S ++   +         T
Sbjct: 1064 TTLSEDTGKTSVSEEITTEMTVTEETSETSPTEGTSDKTTMSTVSEETESSSVTEETTTE 1123

Query: 171  HTKVKKATITGMLEATKSQLTNLELKADYSFSNLTDSSVE-NGTISLGNGQLTTKDTTLK 229
             T V+ AT     E T S  T +   ++ S     +++ E   T  +     T   T+ K
Sbjct: 1124 TTVVENATDISSTEVTASDKTTMTTMSEESEKTTEEATTEITVTKEVTESSSTETATSDK 1183

Query: 230  AVNIQSLHPGGIE-AERTTLENVTFTVSKSKEEEENDYYDNDAIFTAHALTLKGTNTITG 288
             ++  S   G    AE +T E VT T   +  EE           T+  +T+K T T
Sbjct: 1184 TISTLSEETGKTSVAEESTTEKVTETTVTTMPEETGK------TITSEEITIKTTVTEEP 1237

Query: 289  GDIDVDITLTKAKAIAYRARTENGKVSLGSQLTPAKIGKESTSDVISYVAENKAATGNLT 348
             D+     +T  K     A  E GK S+  + T      E++++  S  A     T   T
Sbjct: 1238 TDVGSSEAITSDKTTVSTASEETGKYSVSEEETVKTTVAEASTEPSSTEAITSDKTKMST 1297

Query: 349  VNLNKGDITI 358
            ++   G  ++
Sbjct: 1298 ISEETGKTSV 1307
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2976

A DNA sequence (GASx2148R) was identified in *S.pyogenes* <SEQ ID 8451> which encodes the amino acid sequence <SEQ ID 8452>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2977

A DNA sequence (GASx2160) was identified in *S.pyogenes* <SEQ ID 8453> which encodes the amino acid sequence <SEQ ID 8454>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1630(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2978

A DNA sequence (GASx2170R) was identified in *S.pyogenes* <SEQ ID 8455> which encodes the amino acid sequence <SEQ ID 8456>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.32   Transmembrane  181 - 197 ( 175 - 203)

----- Final Results -----
               bacterial membrane --- Certainty=0.6328(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2979

A DNA sequence (GASx2174) was identified in *S.pyogenes* <SEQ ID 8457> which encodes the amino acid sequence <SEQ ID 8458>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.39   Transmembrane    3 -  19 (   3 -  19)

----- Final Results -----
               bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2980

A DNA sequence (GASx2181R) was identified in *S.pyogenes* <SEQ ID 8459> which encodes the amino acid sequence <SEQ ID 8460>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3751(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2981

A DNA sequence (GASx2185R) was identified in *S.pyogenes* <SEQ ID 8461> which encodes the amino acid sequence <SEQ ID 8462>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.90   Transmembrane   18 -  34 (  18 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.1362(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2982

A DNA sequence (GASx2186R) was identified in *S.pyogenes* <SEQ ID 8463> which encodes the amino acid sequence <SEQ ID 8464>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4803(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA78948 GB:Z17279 transposase [Streptococcus salivarius]
 Identities = 48/77 (62%), Positives = 55/77 (71%), Gaps = 1/77 (1%)

Query: 1   VSMKPIDLSKMVSIRKRSKKVMKTNKKTLGKSIEERPEYINDRSEFGHWEIDLALGKKTK 60
           + +K IDL + V IRK+  K   T KK LGKSIEERPE IN+RS FG WEID  LG KT
Sbjct: 150 LEIKVIDLPRAVRIRKKFTKRPST-KKHLGKSIEERPEEINNRSRFGDWEIDSVLGGKTI 208

Query: 61  SEAVMLTLVERQTRYAL 77
            E  +LTLVERQTRYA+
Sbjct: 209 GEPSILTLVERQTRYAV 225
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2983

A DNA sequence (GASx2187R) was identified in *S.pyogenes* <SEQ ID 8465> which encodes the amino acid sequence <SEQ ID 8466>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3287(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA78948 GB:Z17279 transposase [Streptococcus salivarius]
 Identities = 48/87 (55%), Positives = 57/87 (65%)

Query: 1  MNMSNINSTRKSSYSHLSATERGEIAAYLKMGKKPVEIARLLGSHRSTICREIKRGSVDQ 60
          MNMS   ST   SY HLS  ERGEI AYL +G KP EIAR LG +RSTI REI RGS+ Q
Sbjct: 1  MNMSTNYSTTNQSYKHLSEAERGEIEAYLSVGLKPAEIARRLGRNRSTITREINRGSITQ 60

Query: 61 VKDKNGKQTFFNAYFADSRQRVYETNR 87
          VK  NG++ ++  Y+AD+    Y   R
Sbjct: 61 VKKVNGQKVYYQHYYADAAHNRYRHAR 87
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2984

A DNA sequence <SEQ ID 9013> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9014>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:     10.50
GvH: Signal Score (-7.5): -5.2
     Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 4 value: -12.26 threshold:  0.0
   INTEGRAL    Likelihood =-12.26   Transmembrane   98 - 114 (  94 - 116)
   INTEGRAL    Likelihood = -8.17   Transmembrane    5 -  21 (   1 -  27)
   INTEGRAL    Likelihood = -6.95   Transmembrane   62 -  78 (  57 -  80)
   INTEGRAL    Likelihood = -5.84   Transmembrane   37 -  53 (  30 -  55)
   PERIPHERAL  Likelihood = 17.35      81
 modified ALOM score:   2.95

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01345(292 - 636 of 951)
PIR|G64646|G64646(56 - 168 of 205) hypothetical protein HP1015 - Helicobacter pylori
(strain 26695)
%Match = 4.4
%Identity = 30.6  %Similarity = 54.1
Matches = 34  Mismatches = 46  Conservative Sub.s = 26

87        117       147       177       207       237       267       297
LSGMGATFVPQTLIHRYLDKECNVYHFHKNKLFSEYIMIYKKDVELSGIALLLYKAFLTK*FR*FY*KSVYFLPKSV*NR
                                                                           |
                          RYFLQNIIHIHQNKELQFIKKCLLGYFFAPLCGAILLVLFIVSSGAKSFQISNLFNN
                                   10        20        30        40        50

327       357       381       411       441       471       501
PMIYKIIASLFLVLIPIFSQVL--VKIFKLKKFNIMFPDVAFPIFVFLIPLISSSLLKQNLLPYYLILISLLAIGITI--
 : | :: ||||  :   :  :   :: |: :    | : ||  :  ::  | | |:     |     :||:  |::|
QLAYIVLLSLFLCALGFIAGAIGFYRLSKITRHLSFFENFAFSFLAVILCAILSYLV-----PNASNALSLIGNGVSIFY
          70        80        90       100       110            120       130

549       579       606       636       666       696       726       756
--KLLRTKTLFSYKRFLKLFWRSGF-ILTFLFYLGLLVIIFIKVQ*KELDKLNCTPKVRQKI*RLGCFSDEIKL*R*TRN
  || |  :|:: :||    : ||| :| | | | || |:
LHKLYRELSLYTQERF----FLSGFRLLLFSFMLALLGILVQALVIIFLTTAVVLMCVALGFLARAFLNFSQVFLKA
                  150       160       170       180       190       200
```

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2985

A DNA sequence <SEQ ID 9015> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9016>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
```

```
McG: Discrim Score:      13.20
GvH: Signal Score (-7.5): -2.08
     Possible site: 34
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:  10.45 threshold:  0.0
   PERIPHERAL  Likelihood = 10.45      36
 modified ALOM score:  -2.59

*** Reasoning Step: 3

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
43.9/72.0% over 56aa

                                                                    Streptococcus
pneumoniae
  EGAD|7626| epua protein Insert characterized
  SP|Q03159|EPUA_STRPN EPUA PROTEIN. Insert characterized
  GP|47373|emb|CAA38133.1||X54225 7 kDa protein Insert characterized
  PIR|S10640|S10640 epuA protein - Insert characterized

ORF01809(331 - 501 of 801)
EGAD|7626|7426(8 - 64 of 64) epua protein {Streptococcus pneumoniae}SP|Q03159|EPUA_STRPN
EPUA     PROTEIN.GP|47373|emb|CAA38133.1||X54225     7     kDa     protein     {Streptococcus
pneumoniae}PIR|S10640|S10640 epuA protein - Streptococcus pneumoniae
%Match = 10.0
%Identity = 43.9  %Similarity = 71.9
Matches = 25  Mismatches = 16  Conservative Sub.s = 16

171       201       231       261       291       321       351       381
RSCLLTYELVQL*SWQEWLRKGKQ*LAN*PI*TVVIINSMKN*RLLVLILNV*VRRNMASSGWKYVLKQIGLIVLVILLA
                                                        :    ||:|:: |:::|::|
                                                    MKMNKKSSYVVKRLLLVIIVLILG
                                                             10        20

411       441       471       501       531       561       591       621
LLFLAVGLMLGYSVFGDGEHAYSILSLDKWQNIIGKFLGK*KEPL*VI*CL*WFPLRVNFSSRIIQ*QKNKNK*QLRL*L
 | | :|||:|| ::| |:  ::||| |||:| || |
TLALGIGLMVGYGILGKGQDPWAILSPAKWQELIHKFTGN
              40        50        60
```

A related DNA sequence <SEQ ID 10507> was identified in GBS which encodes amino acid sequence <SEQ ID 10508>.

SEQ ID 9016 (GBS168) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 9; MW 7.6kDa) and in Figure 34 (lane 5; MW 7.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 42 (lane 2; MW 32.6kDa).

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vacc

Example 2986

A DNA sequence <SEQ ID 9017> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9018>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      -2.85
```

```
GvH: Signal Score (-7.5): -5.7
     Possible site: 21
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 0 value:   5.25 threshold:  0.0
   PERIPHERAL  Likelihood =  5.25     103
 modified ALOM score:  -1.55

*** Reasoning Step: 3

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1210(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
56.1/72.0% over 131aa
                                        Escherichia coli
  EGAD|40237| arsenate reductase Insert characterized
  SP|P52147|ARC2_ECOLI ARSENATE REDUCTASE (ARSENICAL PUMP MODIFIER). Edit characterized
  GP|1061418|gb|AAB09628.1||U38947 ArsC (Plasmid R46) Insert characterized

ORF00095(304 - 699 of 1008)
EGAD|40237|42398(1 - 132 of 141) arsenate reductase (Escherichia coli) SP|P52147|ARC2_ECOLI
ARSENATE REDUCTASE (ARSENICAL PUMP MODIFIER). GP|1061418|gb|AAB09628.1||U38947 ArsC
(Plasmid R46)
%Match = 22.0
%Identity = 56.1  %Similarity = 72.0
Matches = 74  Mismatches = 37  Conservative Sub.s = 21

129       159       189       219       249       279       309       339
RIHSSLSL*PIFHRKRPYPSRAFRRYFSNSCG*LWC*YCDDWRELLAGLGINFYFLKTLVALKIERKMMEKIRIYHNPNC
                                                                 |  | |||||:|
                                                                MSNITIYHNPHC
                                                                        10

369       399       429       459       489       519       549       579
GTSRNVLAIIRHCGIEPEIIYYLKTPPSRMELVELLLEMKLSARELLRTDVPAYEKFNLESSSVTDEEMIDAMIQDPILI
|||||  | :||: |||| :| ||:||||| ||::|: :| :| | ||| :|  ||:: |     ||:::|| |:| ||||
GTSRNTLEMIRNSGIEPTVILYLETPPSRDELLKLIADMGISVRALLRKNVEPYEELGLAEDKFTDDQLIDFMLQHPILI
                  30        40        50        60        70        80        90

609       639       669       699       729       759       789       819
NRPIVVTSKGAKLCRPCEAILTILPVKMEKDFVKEDGQIIQSL*HIV**IMV*EVSK*HY*KKLMRLETFCKQKASQHQN
|||||||  | |||||  | :| |||   :  | ||||: :
NRPIVVTPLGTKLCRPSEVVLDILPDAQKAAFTKEDGEKVVDDSGKRLK
                  110       120       130       140
```

SEQ ID 9018 (GBS45) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 8 (lane 4; MW 18.6kDa).

The GBS45-His fusion product was purified (Figure 97A; see also Figure 191, lane 5) and used to immunise mice (lane 1 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 97B), FACS (Figure 97C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2987

A DNA sequence <SEQ ID 9019> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9020>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:      6.84
GvH: Signal Score (-7.5): 2.98
     Possible site: 25
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:  13.69 threshold:  0.0
   PERIPHERAL  Likelihood = 13.69      77
 modified ALOM score:  -3.24

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear)
```

A DNA sequence <SEQ ID 10337> was identified in GBS which encodes amino acid sequence <SEQ ID 10338>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

SEQ ID 9020 (GBS55) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 7; MW 11.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 5; MW 36.3kDa).

GBS55-GST was purified as shown in Figure 197, lane 5.

GBS671 was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 161 (lane 2-4; MW 12kDa) and in Figure 188 (lane 2; MW 12kDa). Purified protein is shown in Figure 242, lane 3.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2988

A DNA sequence <SEQ ID 9021> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9022>. Analysis of the amino acid sequence reveals the following:

```
Lipop Possible site: -1   Crend: 3
McG: Discrim Score:    -14.35
GvH: Signal Score (-7.5): -2.12
     Possible site: 44
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 4 value: -13.90 threshold:  0.0
   INTEGRAL    Likelihood =-13.90   Transmembrane  101 - 117 (  92 - 126)
   INTEGRAL    Likelihood = -7.64   Transmembrane  130 - 146 ( 125 - 148)
   INTEGRAL    Likelihood = -6.64   Transmembrane   24 -  40 (  20 -  45)
   INTEGRAL    Likelihood = -2.44   Transmembrane   55 -  71 (  55 -  75)
   PERIPHERAL  Likelihood = 17.40       2
 modified ALOM score:   3.28

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6562(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 9022 (GBS215) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 175 (lane 10; MW 45kDa).

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2989

A DNA sequence <SEQ ID 9023> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9024>. Analysis of the amino acid sequence reveals the following:

```
Lipop Possible site: -1   Crend: 0
McG: Discrim Score:      11.66
GvH: Signal Score (-7.5): -5.3
     Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 2 value: -14.12 threshold:  0.0
   INTEGRAL    Likelihood =-14.12   Transmembrane   13 -  29 (   5 -  35)
   INTEGRAL    Likelihood = -8.17   Transmembrane   44 -  60 (  39 -  65)
   PERIPHERAL  Likelihood = 39.00       29
 modified ALOM score:   3.32

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6647(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 9024 (GBS217) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 2; MW 36.1kDa) and in Figure 156 (lane 1 & 3; MW 36kDa).

GBS217-GST was purified as shown in Figure 224, lane 5-6.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2990

A DNA sequence <SEQ ID 9025> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9026>. Analysis of the amino acid sequence reveals the following:

```
Lipop Possible site: -1   Crend: 10
McG: Discrim Score:      8.20
GvH: Signal Score (-7.5): -3.7
     Possible site: 33
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 4 value:  -9.98 threshold:  0.0
   INTEGRAL    Likelihood = -9.98   Transmembrane   22 -  38 (  12 -  43)
   INTEGRAL    Likelihood = -7.80   Transmembrane   61 -  77 (  56 -  85)
   INTEGRAL    Likelihood = -5.20   Transmembrane  121 - 137 ( 117 - 148)
   INTEGRAL    Likelihood = -2.97   Transmembrane   99 - 115 (  98 - 119)
   PERIPHERAL  Likelihood = 10.77       5
 modified ALOM score:   2.50

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4991(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence <SEQ ID 10701> was identified in GBS which encodes amino acid sequence <SEQ ID 10702>.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2991

A DNA sequence <SEQ ID 9027> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9028>. Analysis of the amino acid sequence reveals the following:

```
Lipop Possible site: -1   Crend: 7
McG: Discrim Score:     10.61
GvH: Signal Score (-7.5): -4.21
     Possible site: 51
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 3 value: -10.99 threshold:  0.0
   INTEGRAL    Likelihood =-10.99   Transmembrane   38 -  54 (  33 -  61)
   INTEGRAL    Likelihood = -8.01   Transmembrane    5 -  21 (   1 -  26)
   INTEGRAL    Likelihood = -7.01   Transmembrane   65 -  81 (  60 -  87)
   PERIPHERAL  Likelihood = 13.85      99
 modified ALOM score:   2.70

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2992

A DNA sequence <SEQ ID 9029> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9030>. Analysis of the amino acid sequence reveals the following:

```
Lipop Possible site: -1   Crend: 10
McG: Discrim Score:    -21.39
GvH: Signal Score (-7.5): -1.85
     Possible site: 57
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -8.44 threshold:  0.0
   INTEGRAL    Likelihood = -8.44   Transmembrane   38 -  54 (  36 -  59)
   PERIPHERAL  Likelihood = 19.10      18
 modified ALOM score:   2.19

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2993

A DNA sequence <SEQ ID 9031> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9032>. Analysis of the amino acid sequence reveals the following:

```
Lipop Possible site: -1   Crend: 5
McG: Discrim Score:     12.87
GvH: Signal Score (-7.5): -3.57
     Possible site: 41
```

```
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 4 value: -10.30 threshold:  0.0
   INTEGRAL    Likelihood =-10.30   Transmembrane   69 -  85 (  63 -  98)
   INTEGRAL    Likelihood = -8.65   Transmembrane    4 -  20 (   1 -  29)
   INTEGRAL    Likelihood = -2.07   Transmembrane   96 - 112 (  95 - 118)
   PERIPHERAL  Likelihood =  9.71     113
 modified ALOM score:   2.56

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5118(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
20.1/50.5% over 114aa
                                                                  Streptococcus pneumoniae
  GP|9798572| BlpX protein Insert characterized

ORF02100(316 - 660 of 999)
GP|9798572|emb|CAC03527.1||AJ276410(9 - 123 of 132) BlpX protein {Streptococcus pneumoniae}
%Match = 5.0
%Identity = 20.0  %Similarity = 50.4
Matches = 23  Mismatches = 57  Conservative Sub.s = 35

90       120       150       180       210       240       270       300
LMSLF*DPQVSGEELDKFTVRLDSHRKSNSRG*NQLVIILRLYSQIN*REPNMLVGPFLNKGEHMTQDYICYL*SRGGED

                                                                        MEV

330       360       390       420       450       480       510       540
MHNILRFLGIVIISAVILFSIGSFYDLTLMKNILLICWSFLFDLLVFVFKQRQTTEVLTWYQVVKQFWLFIKCTILIPIL
 :   |:: :: :|:|:   :|:|  ::::    |: :| ||:  ||::  :     :    :  :| :    |:  ||
FNMKYRLFFVIFLSSVLDILLGTFLQISIVSIGWLVLYSGLFEAGVFLLANKGVAVKIKEVDIRNRFKFIFGKTLWFQIL
                 20        30        40        50        60        70        80

570       600       630       660       690       720       750       780
VAFIIMKGCLTSISDILIYFYLHLVVVYYTIGMILSLGRIISPEHSMFNKLRK*NELYLKFVFNRADLTICCLPCLS*FF
:   ::      :   || || |: :|:  :  :||  ::   :: : :
LLIFLIIKLYLGLDARLILFYGHIFIVFNALMYLLSSSQVSLKKNKLSS
               100       110       120       130
```

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2994

A DNA sequence <SEQ ID 9033> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9034>. Analysis of the amino acid sequence reveals the following:

```
Lipop Possible site: -1   Crend: 9
McG: Discrim Score:      3.25
GvH: Signal Score (-7.5): -3.39
     Possible site: 59
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 4 value:  -6.64 threshold:  0.0
   INTEGRAL    Likelihood = -6.64   Transmembrane   46 -  62 (  43 -  64)
   INTEGRAL    Likelihood = -5.15   Transmembrane   17 -  33 (  15 -  34)
   PERIPHERAL  Likelihood = 11.03     100
 modified ALOM score:   1.83

*** Reasoning Step: 3

----- Final Results -----
```

```
bacterial membrane --- Certainty=0.3654(Affirmative) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
35.5/63.8% over 127aa

  OMNI|NT01BS4455| wall teichoic acid glycosylation protein GtcA Insert characterized

ORF01715(343 - 750 of 1053)
OMNI|NT01BS4455(58 - 185 of 187) wall teichoic acid glycosylation protein GtcA
%Match = 8.0
%Identity = 35.5  %Similarity = 63.7
Matches = 44  Mismatches = 39  Conservative Sub.s = 35

210       240       270       300       330       360       390       411
GN*ASRVV*NNLLSISQTKSKAKLMGDFLITLKHP*YNKNMVKLKSLLKKSIQNEVSLYLLFGLLTSLLYLV---IRQGI
     : :     :         : :| |    :          ::          |: :|:: |::|::: :    |   |
PRRNHQTIICIGPASHLPQLFRRTLGIFYFRQRAREAKNFEKFFRKRGTSVKYREIIMYIIMGVFTTIVNIASFYILVEI
                20        30        40        50        60        70        80

441       471       501       531       549       579       609       639
FNFSQDAPFSAIVANIIAILFAFFTNDRFVFKQTKIEQLQRL----QTFVIARLGTLGLDLILAVIFVDQFPSIIGQFVQ
 |    |  : : | |:::|||: ||  :||:| |   || |      |   |: :||:|| : :|:| ||
MNVDYKA--ATVAAWILSVLFAYITNKLYVFQQ-KTHDLQSLLKELTAFFSVRVLSLGIDLGMMIILVGQF---------
                100       110       120       130       140       150

669       690       720       750       780       810       840       870
HNLNKINTIESL---VSQILIILLNYILSKFVIFKDKKRQL*QELSFLIFLLWIFG*ET*YLHALIQFFLSQFLERWHSV
      || |:|   :   :|:::||: ||:::||  | :
------NTNETLAKILDNAVIVVVNYVASKWLVFKKTKEEGV
           160       170       180
```

SEQ ID 9034 (GBS283) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 63 (lane 8; MW 67.6kDa).

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2995

A DNA sequence <SEQ ID 9035> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9036>. Analysis of the amino acid sequence reveals the following:

```
Lipop Possible site: -1   Crend: 2
SRCFLG: 0
McG: Length of UR:   22
     Peak Value of UR:   3.86
     Net Charge of CR: 2
McG: Discrim Score:     16.84
GvH: Signal Score (-7.5): -4.38
     Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 1 value: -12.37 threshold:  0.0
   INTEGRAL    Likelihood =-12.37   Transmembrane    7 -  23 (   1 -  26)
   PERIPHERAL  Likelihood = 12.84       64
 modified ALOM score:   2.97
icm1 HYPID: 7  CFP: 0.595

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5946(Affirmative) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 9036 (GBS286) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 11; MW 16.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 2; MW 41.3kDa) and in Figure 63 (lane 9; MW 41.4kDa).

The GBS286-GST fusion product was purified (Figure 210, lane 9; Figure 225, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 274), which confirmed that the protein is immunoaccessible on GBS bacteria.

GBS668 was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 139 (lane 2-4; MW 43.5kDa) and in Figure 187 (lane 6; MW 43kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 139 (lane 6 & 7; MW 18.6kDa) and in Figure 179 (lane 12; MW 19kDa).

GBS668-GST was purified as shown in Figure 237 (lane 10). GBS668-His was purified as shown in Figure 231 (lanes 5 & 6).

GBS673 was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 161 (lane 8-10; MW 17kDa) and in Figure 188 (lane 4; MW 17kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 162 (lane 8; MW 41.5kDa) and in Figure 239 (lane 7; MW 41kDa). Purified GBS673-His is shown in Figure 242, lane 5. Purified GBS673-GST is shown in Figure 246, lane 2.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2996

A DNA sequence <SEQ ID 9037> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9038>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:    -18.42
GvH: Signal Score (-7.5): -6.16
     Possible site: 57
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value:  -8.49 threshold:  0.0
   INTEGRAL    Likelihood = -8.49   Transmembrane   51 -  67 (  44 -  95)
   INTEGRAL    Likelihood = -3.08   Transmembrane   70 -  86 (  68 -  95)
   PERIPHERAL  Likelihood = 12.89      32
 modified ALOM score:   2.20

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4397(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 9038 (GBS386) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 70 (lane 2; MW 14kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 8; MW 39.5kDa).

GBS386-GST was purified as shown in Figure 213, lane 8.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2997

A DNA sequence <SEQ ID 9039> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9040>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:    -15.47
GvH: Signal Score (-7.5): -6.21
     Possible site: 14
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value:  -3.61 threshold:  0.0
   INTEGRAL    Likelihood = -3.61   Transmembrane   94 - 110 (  94 - 111)
   INTEGRAL    Likelihood = -1.70   Transmembrane   75 -  91 (  75 -  91)
   PERIPHERAL  Likelihood =  5.94     139
 modified ALOM score:   1.22

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01481(394 - 720 of 1065)
GP|9657521|gb|AAF96047.1||AE004354(16 - 121 of 243) uridine phosphorylase (Vibrio cholerae)
%Match = 5.3
%Identity = 28.0  %Similarity = 48.6
Matches = 30  Mismatches = 54  Conservative Sub.s = 22

150       180       210       240       270       300       330       360
V*KHMV*AI*YGNLP*KW*IVPLSIFIFANLTLPFKFH*VKIEKIFLTR**NIVN*GLKEMLMIINSFDNSRKAIINPED

                                                                        MSIQ

390       420       450       480       510       540       570       600
INSPIKGFPKTVITCFARETFNRILEELPHREIARTSVANLEIPIYELEFKGQKIGFFNAYVGASACVAILEDIIVFGME
          |: |       |||   | : |       | |  ::  ||: | |   :  :|| : :  :|::   | :
PHIHVAQVAPRVVVCGEPNRANRIASLLNNAE---LVAENREYRLFSGEFEEQPITVCSTGIGAPSMIIAVEELARSGAK
              20        30          40        50        60        70        80

630       660       690       720       750       780       810       840
SLIVFGTCGVLDSSIEETSIIIPRSAIRDEGTSFHYSEASSEIAVNTNSIFLLCG*FRCRSMGSKIFRK*RGFRKER*NC
:::  |: | : | |    :|:   |:|||| |  |  |:
AIVRVGSAGAMQSEIGLGELILVEGAVRDEGGSKAYIGAAYPAYSSFELVVEMQRFLAEQSVPIHRGIVRSHDSFYTDEE
                100       110       120       130       140       150       160
```

SEQ ID 9040 (GBS388) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 70 (lane 3; MW 21kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 9; MW 45.6kDa).

The GBS388-GST fusion product was purified (Figure 213, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 311), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2998

A DNA sequence <SEQ ID 9041> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9042>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:    -11.81
GvH: Signal Score (-7.5): -7.49
     Possible site: 25
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -5.68 threshold:  0.0
   INTEGRAL    Likelihood = -5.68   Transmembrane   78 -  94 (  77 -  95)
   PERIPHERAL  Likelihood =  4.61     134
 modified ALOM score:   1.64

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3272(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01912(307 - 720 of 1056)
GP|3845252|gb|AAC71927.1||AE001412(81 - 242 of 244) hypothetical protein (Plasmodium
falciparum) PIR|D71608|D71608 hypothetical protein PFB0690w - malaria parasite (Plasmodium
falciparum)
%Match = 4.0
%Identity = 31.2  %Similarity = 53.5
Matches = 45  Mismatches = 58  Conservative Sub.s = 32

231       261       291                 348       378       405
KKGRFLIDLYCNVMNFKNSKIA*NQCFDV**RVVNHLLN-LSKENIAKIDFDFLNEALNA-NIRLKELVDELKISK----
                                  || | : | :: : |: |  |  :    :|| | |  || |
KYNELQSLLSKEEEKYDFVKNELGDLQKQKDLLKWHLCNNIKKLSMKRSDYKFKTETKSKLESKLKSLKDMNKIHKFEHD
             60        70        80        90       100       110       120

450       480       501       531       558       588       618
----------ELDSKGWSKKDSRTIKILYDGLINK---HIVSLDRADYNII-QVIPFANVHVLLFLIPERENSKNYRIY
           ||::| : | |   |: :::  |||   ::  : :   ::  :     |    |:|| ::||:||| |
TLEELVHKMEQELETKMYIKND---IENIFNECINKKDEYLKDITQERISVFKERKKRQNQLQKLLLIMKQENNKNYNIN
            140          150       160       170       180       190       200

648       672       693       720       750       780       810       840
NYSDYEMELINE--DRQQFSKYET---VDL-DQLILVDIFNIDDYISSYLTI*DIENLDLGLLKLINYADNKSDRHILQT
    ||  |:||    : :  :||   :|| |   | |::
YLKKYESNLMNEINSYKNYKDFETKIAMDLIDDHSLNDLYVT
               220       230       240
```

A related DNA sequence <SEQ ID 10589> was identified in GBS which encodes amino acid sequence <SEQ ID 10590>.

SEQ ID 9042 (GBS408) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 6; MW 20.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 5; MW 45.3kDa).

GBS408-GST was purified as shown in Figure 218, lane 9.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2999

A DNA sequence <SEQ ID 9043> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9044>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:      -9.62
GvH: Signal Score (-7.5): -4.84
     Possible site: 61
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value: -11.09 threshold:  0.0
   INTEGRAL    Likelihood =-11.09   Transmembrane   45 -  61 (  37 -  72)
   INTEGRAL    Likelihood = -8.60   Transmembrane   76 -  92 (  70 -  97)
   PERIPHERAL  Likelihood = 11.62       95
 modified ALOM score:   2.72

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5437(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01977(442 - 627 of 948)
EGAD|88220|96064(204 - 583 of 751) hypothetical 848 kDa protein f23f125 in chromosome iii
{Caenorhabditis elegans} SP|P46501|YLX5_CAEEL HYPOTHETICAL 84.8 KDA PROTEIN F23F12.5 IN
CHROMOSOME III. GP|529214|gb|AAA20607.1||U12965 F23F12.5 gene product {Caenorhabditis
elegans}
%Match = 4.6
%Identity = 35.9  %Similarity = 59.4
Matches = 23  Mismatches = 24  Conservative Sub.s = 15

192       222       252       282       312       342       372       402
DFVSSFFIS*SQTNYNRISFLLKLAKHQLECLNNVAQGLSV**YSSMKDYINRILHFIKEHMTYHVNFIDDFLDIKWEKV

VTLSAYFPFTITVERYYAMNKSEKYEKMPIILGPLFVLFIVKLELKIKDKVTLFQVIVNFGVIFQIYKNETFSHGDVAFS
      120       130       140       150       160       170       180       190

432       462       492       522       552
SNIHLRFWTTIIAYLVIFILSISTVILNLVLLFQGFLTQNPIIYLLFFITLVCAFY----------------------~~~~
            |:|:|:::: |::|||||  || |  :| ||:    |:
LYPPGTAEKVFTFYVVLFLINLLDVMFNLVLLQMSFLNTNRFHWLCFFLWQFALFFCCQQIFSIFYNFSPGLSCDD~~~~
      200       210       220       230       240       250       260

600       627       657       687       717       747
--------------------FAYKFITYTPTIVKNAL-QYIKKLKNV*NNKVICTLTLYQLCFRVFFHTKITKKDSYLTI
                     | : : || |:  :  | ||
AGNFYLSQFVSGAVTAFAKIFVFLLDTYVPSFDRRRLHQYPQIAMILCYCVIMVLMILPESDCGSQGSRDLAIIIINIIG
               560       570       580       590       600       610       620
```

SEQ ID 9044 (GBS411) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 78 (lane 2; MW 16kDa).

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3000

A DNA sequence <SEQ ID 9045> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9046>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:    -17.94
GvH: Signal Score (-7.5): -4.63
     Possible site: 45
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -6.10 threshold:  0.0
   INTEGRAL    Likelihood = -6.10   Transmembrane   31 -  47 (  26 -  49)
   PERIPHERAL  Likelihood = 15.33       3
 modified ALOM score:   1.72

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3442(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01982(313 - 501 of 801)
GP|2444082|gb|AAC79518.1||U88974(93  -  156  of  156)  ORF2  (Streptococcus  thermophilus
temperate bacteriophage O1205) PIR|T13290|T13290 hypothetical protein 2 - Streptococcus
phage phi-O1205
%Match = 11.5
%Identity = 48.4  %Similarity = 59.4
Matches = 31  Mismatches = 25  Conservative Sub.s = 7

174       204       234       264       294       324       354       384
DVDQNIESHKLFKRHFV*RAILPQSKRK*EN**LCVISEPR*KLKSKLGELKMGFFAQRCPYCQSTKVQFMNQDRKGFNG
                                                          | :||||:|| |:|| | || |:
LLMFVGVALLFARLFWEIKHPMTKEQKEQLKIERAKAKEEFRKSRNEFKKAMAEARAVKCPYCKSTDVEFMVQQRKSFSI
            50        60        70        80        90        100       110

414       441       471       501       531       561       591       621
CVGCIGFLIAWPF-LLLGLVGKKGKNNWHCTNCGRTFKTK*KSPTLKFCPRRA*GKF*YSKNLLFGRGFYHTYFNRK*GI
     | ::      | |: |||||  ||| |||  | ||
GKAAAGTIMTGGVGALAGFAGKKGKKEWHCKNCGAVFTTK
           130       140       150
```

SEQ ID 9046 (GBS412) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 6; MW 36kDa). Purified GBS412-GST is shown in Figure 218, lane 10-11.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3001

A DNA sequence <SEQ ID 9047> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9048>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
McG: Discrim Score:      3.67
GvH: Signal Score (-7.5): -3.62
     Possible site: 41
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 5 value:  -7.27 threshold:  0.0
   INTEGRAL    Likelihood = -7.27   Transmembrane   48 -  64 (  32 -  68)
   INTEGRAL    Likelihood = -6.26   Transmembrane   87 - 103 (  85 - 105)
```

```
   INTEGRAL    Likelihood = -6.21   Transmembrane   29 -  45 (  26 -  46)
   INTEGRAL    Likelihood = -3.29   Transmembrane  110 - 126 ( 109 - 130)
   INTEGRAL    Likelihood = -2.87   Transmembrane    2 -  18 (   1 -  18)
   PERIPHERAL  Likelihood =  4.24      66
 modified ALOM score:   1.95

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3909(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01286(304 - 672 of 993)
GP|8272442|dbj|BAA96471.1||AB036428(90 - 212 of 218) type IV prepilin peptidase homologue
{Streptococcus mutans}
%Match = 16.8
%Identity = 46.3  %Similarity = 72.4
Matches = 57  Mismatches = 34  Conservative Sub.s = 32

102       132       162       192       222       252       282       312
*RRLPI*T*IPNFFKRFCTSNKTFIYEF*QKTIQFSRKSATAC*LSL*R*TDYL**KS*SLFYHFSNININYKKDFMIMS
                                                                          :::
LGSFFGLVVDRYPQKSIIFPRSHCNKCYNCLTMRDLIPIFSRIINKNSCRFCGYPIPLRYSLVELLCGLISTGFALDLLT
                  30        40        50        60        70        80        90

342       372       402       432       462       492       522       552
TIYFISLCMSFILSYYDIKYQEYPIFLWILFTISTIILTPITKVSIVLCLFGILAEVVDINIGSGDFLYLATIGLSLPLH
|     | |  :|| ||:: | ||: ||| ||   : : |:  :|::| ||||:| : :|||||||:||||: ||| |:
TSQVCLLFMGVLLSLYDLQDQSYPLTLWIGFTFLLMFIYPLNLISLILFLFGIFAALKNINIGSGDFFYLATLALSLNLQ
                  110       120       130       140       150       160       170

582       612       642       672       702       732       762       792
QMLFIIQIGAWLGIIYCLVMRKMKKTIAFLPFLSIAYIIVTSYSLLF*SL*IIRKVIKLWVLVAFWIFRMTNCTTKINHG
|:::|||| : |||:| |: :| |:  ||:||| : ::|:
QIIWIIQIASLLGILYSLLFQKHKEPFAFVPFLFLGHLIIIFSHLI
                  190       200       210
```

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3002

A DNA sequence <SEQ ID 9049> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9050>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 2
McG: Discrim Score:     10.43
GvH: Signal Score (-7.5): -4.39
     Possible site: 54
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 4 value: -10.30 threshold:  0.0
   INTEGRAL    Likelihood =-10.30   Transmembrane   62 -  78 (  59 -  84)
   INTEGRAL    Likelihood = -6.10   Transmembrane    4 -  20 (   1 -  22)
   INTEGRAL    Likelihood = -4.25   Transmembrane  128 - 144 ( 123 - 145)
   INTEGRAL    Likelihood = -3.13   Transmembrane   88 - 104 (  87 - 104)
   PERIPHERAL  Likelihood =  2.01     109
 modified ALOM score:   2.56

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5118(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
SP|Q48729|LSPA_LACLC(1 - 149 of 150) LIPOPROTEIN SIGNAL PEPTIDASE (EC 3.4.23.36)
(PROLIPOPROTEIN SIGNAL PEPTIDASE) (SIGNAL PEPTIDASE II) (SPASE II).
%Match = 16.3
%Identity = 40.7  %Similarity = 66.0
Matches = 61  Mismatches = 50  Conservative Sub.s = 38

180       210       240       270       300       330       360       390
EWYHCYSIRRSSR*PNDLYQTKRS*FISDGFKICCCYGRVF*GI*FIGEVMRKIIIPIITILLIALDQLSKLWIVKHIEL
                                                  |:|::  :| :: |  ||: | |:| :|:|
                                                  MKKLLSLVIIVVGIIADQVFKNWVVANIQL
                                                          10        20        30

420       450       480       510       540       570       600       630
NQIKEFIPNIVSLTYLRNYGAAFSILQNQQWLFTLITIFVVGVAIIYLMKHINGSYWLLISLTLIISGGLGNFIDRLRLG
   |:  |:::||||::| |||:| :  |||:| ::|  |: ||: :| |   |  | :  |||||:| |||:: |:| |
GDTKKIWPDVLSLTYIKNDGAAWSSFSGQQWFFLVLTPIVLIVALWFLWKK-MGQNWYFAGLTLIIAGALGNLLTRVRQG
        40        50        60        70        80        90        100

660       690       720       750       780       810       840       870
YVVDMVHLDFINFAIFNVADSYLTIGIICLMIALWKEESNGNHN*NSRS*AR*SFSG*F*TVA*SS**RN*KRDCVSKWT
:||||      :| |||:||  |::| : | ||:  ::
FVVDMFXNRIYDFPIFNIADILLSVGFVVLFIAILTDKETK
        120       130       140       150
```

There is also homology to SEQ ID 7750.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vacc

Example 3003

A DNA sequence <SEQ ID 9051> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9052>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:     13.24
GvH: Signal Score (-7.5): -2.18
     Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   2.01 threshold:  0.0
   PERIPHERAL  Likelihood =  2.01      21
 modified ALOM score:  -0.90

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no homology with any sequences in the databases.

SEQ ID 9052 (GBS138) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 2; MW 15kDa)

GBS672 was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 161 (lane 5-7; MW 15kDa) and in Figure 188 (lane 3; MW 15kDa). Purified protein is shown in Figure 242, lane 4.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3004

A DNA sequence <SEQ ID 9053> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9054>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     18.01
GvH: Signal Score (-7.5): -2.35
     Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:  14.80 threshold:  0.0
   PERIPHERAL  Likelihood = 14.80      51
 modified ALOM score:  -3.46

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 9054 (GBS143) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 2; MW 33.5kDa).

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3005

A DNA sequence <SEQ ID 9055> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9056>. Analysis of the amino acid sequence reveals the following:

```
Lipop Possible site: -1   Crend: 0
McG: Discrim Score:      7.43
GvH: Signal Score (-7.5): -6.25
     Possible site: 41
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value: -10.77 threshold:  0.0
   INTEGRAL    Likelihood =-10.77   Transmembrane    2 -  18 (   1 -  20)
   PERIPHERAL  Likelihood =  5.14      29
 modified ALOM score:   2.65

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 9056 (GBS229) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 3; MW 35.9kDa).

GBS229-GST was purified as shown in Figure 206, lane 5.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3006

A DNA sequence <SEQ ID 9183> was identified in GAS which encodes amino acid sequence <SEQ ID 9184>. Analysis of the amino acid sequence reveals the following:

```
        Possible site: 29

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3007

A DNA sequence <SEQ ID 9185> was identified in GAS which encodes amino acid sequence <SEQ ID 9186>. Analysis of the amino acid sequence reveals the following:

```
        Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3008

A DNA sequence <SEQ ID 9187> was identified in GAS which encodes amino acid sequence <SEQ ID 9188>. Analysis of the amino acid sequence reveals the following:

```
        Possible site: 36

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.70   Transmembrane  850 - 866 ( 850 - 866)
   INTEGRAL    Likelihood = -1.22   Transmembrane   15 -  31 (  15 -  31)

----- Final Results -----
              bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3009

A DNA sequence <SEQ ID 9189> was identified in GAS which encodes amino acid sequence <SEQ ID 9190>. Analysis of the amino acid sequence reveals the following:

```
LPXTG motif: 259-263

        Possible site: 13
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.93   Transmembrane  270 - 286 ( 268 - 288)

----- Final Results -----
               bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3010

A DNA sequence <SEQ ID 9191> was identified in GAS which encodes amino acid sequence <SEQ ID 9192>. Analysis of the amino acid sequence reveals the following:

```
Possible site: 21

>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3011

A DNA sequence <SEQ ID 9193> was identified in GAS which encodes amino acid sequence <SEQ ID 9194>. Analysis of the amino acid sequence reveals the following:

```
Possible site: 29

>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3012

A DNA sequence <SEQ ID 9195> was identified in GAS which encodes amino acid sequence <SEQ ID 9196>. Analysis of the amino acid sequence reveals the following:

```
Possible site: 34

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3013

A DNA sequence <SEQ ID 9197> was identified in GAS which encodes amino acid sequence <SEQ ID 9198>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 13
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.50   Transmembrane  346 - 362 ( 343 - 366)
   INTEGRAL    Likelihood = -2.97   Transmembrane  177 - 193 ( 176 - 195)

----- Final Results -----
               bacterial membrane --- Certainty=0.2402(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3014

A DNA sequence <SEQ ID 9199> was identified in GAS which encodes amino acid sequence <SEQ ID 9200>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 19
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.33   Transmembrane   24 -  40 (  24 -  40)

----- Final Results -----
               bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3015

A DNA sequence <SEQ ID 9201> was identified in GAS which encodes amino acid sequence <SEQ ID 9202>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.00   Transmembrane  194 - 210 ( 192 - 214)

----- Final Results -----
               bacterial membrane --- Certainty=0.3399(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 183-187
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3016

A DNA sequence <SEQ ID 9203> was identified in GAS which encodes amino acid sequence <SEQ ID 9204>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.25   Transmembrane    9 -  25 (   4 -  28)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3017

A DNA sequence <SEQ ID 9205> was identified in GAS which encodes amino acid sequence <SEQ ID 9206>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.03   Transmembrane  462 - 478 ( 460 - 479)
   INTEGRAL    Likelihood = -0.90   Transmembrane   18 -  34 (  18 -  34)

----- Final Results -----
              bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 450-454
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3018

A DNA sequence <SEQ ID 9207> was identified in GAS which encodes amino acid sequence <SEQ ID 9208>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.60   Transmembrane   15 -  31 (  12 -  32)

----- Final Results -----
              bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3019

A DNA sequence <SEQ ID 9209> was identified in GAS which encodes amino acid sequence <SEQ ID 9210>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 28
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38   Transmembrane   16 -  32 (  16 -  32)

----- Final Results -----
              bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3020

A DNA sequence <SEQ ID 9211> was identified in GAS which encodes amino acid sequence <SEQ ID 9212>. Analysis of the amino acid sequence reveals the following:

```
      Possible cleavage site: 24
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty= 0.300(Affirmative) < succ>
             bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3021

A DNA sequence <SEQ ID 9213> was identified in GAS which encodes amino acid sequence <SEQ ID 9214>. Analysis of the amino acid sequence reveals the following:

```
      Possible cleavage site: 23
>>> May be a lipoprotein

----- Final Results -----
             bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
              bacterial outside --- Certainty= 0.000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3022

A DNA sequence <SEQ ID 9215> was identified in GAS which encodes amino acid sequence <SEQ ID 9216>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 19

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.76   Transmembrane    3 -  19 (   2 -  20)

----- Final Results -----
             bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 396-398
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3023

A DNA sequence <SEQ ID 9217> was identified in GAS which encodes amino acid sequence <SEQ ID 9218>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 18

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.80   Transmembrane  251 - 267 ( 251 - 267)
   INTEGRAL    Likelihood = -0.75   Transmembrane  179 - 195 ( 179 - 195)

----- Final Results -----
               bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3024

A DNA sequence <SEQ ID 9219> was identified in GAS which encodes amino acid sequence <SEQ ID 9220>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 39

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.22   Transmembrane   52 -  68 (  51 -  68)

----- Final Results -----
               bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3025

A DNA sequence <SEQ ID 9221> was identified in GAS which encodes amino acid sequence <SEQ ID 9222>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 52

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.58   Transmembrane   39 -  55 (  32 -  86)
   INTEGRAL    Likelihood = -9.55   Transmembrane   60 -  76 (  56 -  86)

----- Final Results -----
               bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3026

A DNA sequence <SEQ ID 9223> was identified in GAS which encodes amino acid sequence <SEQ ID 9224>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 18
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3027

A DNA sequence <SEQ ID 9225> was identified in GAS which encodes amino acid sequence <SEQ ID 9226>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3028

A DNA sequence <SEQ ID 9227> was identified in GAS which encodes amino acid sequence <SEQ ID 9228>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 33
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.44   Transmembrane   18 -  34 (  13 -  40)
   INTEGRAL    Likelihood = -7.86   Transmembrane   59 -  75 (  54 -  79)

----- Final Results -----
              bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3029

A DNA sequence <SEQ ID 9229> was identified in GAS which encodes amino acid sequence <SEQ ID 9230>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 27
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3030

A DNA sequence <SEQ ID 9231> was identified in GAS which encodes amino acid sequence <SEQ ID 9232>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 24
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3031

A DNA sequence <SEQ ID 9233> was identified in GAS which encodes amino acid sequence <SEQ ID 9234>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 49
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.87   Transmembrane   58 -  74 (  53 -  81)

----- Final Results -----
               bacterial membrane --- Certainty=0.4949(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3032

A DNA sequence <SEQ ID 9235> was identified in GAS which encodes amino acid sequence <SEQ ID 9236>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.06   Transmembrane   92 - 108 (  92 - 108)

----- Final Results -----
               bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3033

A DNA sequence <SEQ ID 9237> was identified in GAS which encodes amino acid sequence <SEQ ID 9238>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38   Transmembrane   18 -  34 (  18 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3034

A DNA sequence <SEQ ID 9239> was identified in GAS which encodes amino acid sequence <SEQ ID 9240>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 19
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3035

A DNA sequence <SEQ ID 9241> was identified in GAS which encodes amino acid sequence <SEQ ID 9242>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.01   Transmembrane  155 - 171 ( 154 - 171)

----- Final Results -----
               bacterial membrane --- Certainty=0.1404(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3036

A DNA sequence <SEQ ID 9243> was identified in GAS which encodes amino acid sequence <SEQ ID 9244>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -4.25   Transmembrane  113 - 129 ( 111 - 131)

----- Final Results -----
               bacterial membrane --- Certainty=0.2699(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3037

A DNA sequence <SEQ ID 9245> was identified in GAS which encodes amino acid sequence <SEQ ID 9246>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.69   Transmembrane  110 - 126 ( 110 - 126)

----- Final Results -----
               bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3038

A DNA sequence <SEQ ID 9247> was identified in GAS which encodes amino acid sequence <SEQ ID 9248>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.28   Transmembrane  130 - 146 ( 128 - 146)

----- Final Results -----
               bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3039

A DNA sequence <SEQ ID 9249> was identified in GAS which encodes amino acid sequence <SEQ ID 9250>. Analysis of the amino acid sequence reveals the following:

```
     Possible site: 39
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.57   Transmembrane   74 -  90 (  72 -  92)
   INTEGRAL    Likelihood = -3.13   Transmembrane  169 - 185 ( 166 - 185)
   INTEGRAL    Likelihood = -3.13   Transmembrane   28 -  44 (  27 -  44)

----- Final Results -----
               bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3040

A DNA sequence <SEQ ID 9251> was identified in GAS which encodes amino acid sequence <SEQ ID 9252>. Analysis of the amino acid sequence reveals the following:

```
     Possible cleavage site: 56
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.21   Transmembrane    93 - 109 (  87 - 114)
   INTEGRAL    Likelihood = -8.65   Transmembrane   227 - 243 ( 226 - 243)

----- Final Results -----
               bacterial membrane --- Certainty=0.588(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3041

A DNA sequence <SEQ ID 9253> was identified in GAS which encodes amino acid sequence <SEQ ID 9254>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 45
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.53   Transmembrane   73 -  89 (  70 -  94)
   INTEGRAL    Likelihood = -4.41   Transmembrane   32 -  48 (  30 -  51)
   INTEGRAL    Likelihood = -2.55   Transmembrane   10 -  26 (  10 -  26)
   INTEGRAL    Likelihood = -2.39   Transmembrane  106 - 122 ( 104 - 123)
   INTEGRAL    Likelihood = -1.75   Transmembrane  153 - 169 ( 152 - 169)

----- Final Results -----
               bacterial membrane --- Certainty=0.3612(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3042

A DNA sequence <SEQ ID 9255> was identified in GAS which encodes amino acid sequence <SEQ ID 9256>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.68   Transmembrane   25 -  41 (  15 -  46)
   INTEGRAL    Likelihood = -9.24   Transmembrane  255 - 271 ( 248 - 276)
   INTEGRAL    Likelihood = -7.59   Transmembrane   82 -  98 (  79 - 100)
   INTEGRAL    Likelihood = -4.30   Transmembrane  115 - 131 ( 113 - 135)
   INTEGRAL    Likelihood = -0.11   Transmembrane  148 - 164 ( 148 - 164)

----- Final Results -----
               bacterial membrane --- Certainty=0.5670(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3043

A DNA sequence <SEQ ID 9257> was identified in GAS which encodes amino acid sequence <SEQ ID 9258>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.04   Transmembrane  137 - 153 ( 126 - 160)
   INTEGRAL    Likelihood =-10.56   Transmembrane   36 -  52 (  29 -  58)
   INTEGRAL    Likelihood =-10.08   Transmembrane  407 - 423 ( 399 - 426)
   INTEGRAL    Likelihood = -4.94   Transmembrane  230 - 246 ( 228 - 250)
   INTEGRAL    Likelihood = -4.83   Transmembrane   79 -  95 (  77 -  98)
   INTEGRAL    Likelihood = -4.35   Transmembrane  202 - 218 ( 201 - 220)
   INTEGRAL    Likelihood = -1.12   Transmembrane  293 - 309 ( 293 - 309)

----- Final Results -----
               bacterial membrane --- Certainty=0.5416(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3044

A DNA sequence <SEQ ID 9259> was identified in GAS which encodes amino acid sequence <SEQ ID 9260>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 31
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.76   Transmembrane  137 - 153 ( 137 - 154)

----- Final Results -----
               bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3045

A DNA sequence <SEQ ID 9261> was identified in GAS which encodes amino acid sequence <SEQ ID 9262>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.91   Transmembrane  238 - 254 ( 236 - 264)
   INTEGRAL    Likelihood = -6.16   Transmembrane   69 -  85 (  65 -  89)
   INTEGRAL    Likelihood = -6.00   Transmembrane  136 - 152 ( 134 - 155)
   INTEGRAL    Likelihood = -4.73   Transmembrane   29 -  45 (  21 -  48)
   INTEGRAL    Likelihood = -2.97   Transmembrane  194 - 210 ( 193 - 220)

----- Final Results -----
               bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3046

A DNA sequence <SEQ ID 9263> was identified in GAS which encodes amino acid sequence <SEQ ID 9264>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 39
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.87   Transmembrane  574 - 590 ( 568 - 601)
   INTEGRAL    Likelihood = -9.18   Transmembrane  243 - 259 ( 238 - 262)
   INTEGRAL    Likelihood = -7.11   Transmembrane   66 -  82 (  65 -  87)
   INTEGRAL    Likelihood = -1.28   Transmembrane  270 - 286 ( 270 - 287)

----- Final Results -----
               bacterial membrane --- Certainty=0.4949(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3047

A DNA sequence <SEQ ID 9265> was identified in GAS which encodes amino acid sequence <SEQ ID 9266>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 33
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.91   Transmembrane   98 - 114 (  92 - 124)
   INTEGRAL    Likelihood = -6.21   Transmembrane   19 -  35 (  14 -  37)
   INTEGRAL    Likelihood = -5.36   Transmembrane  170 - 186 ( 169 - 189)
   INTEGRAL    Likelihood = -5.15   Transmembrane  147 - 163 ( 136 - 167)
   INTEGRAL    Likelihood = -1.12   Transmembrane   77 -  93 (  77 -  93)

----- Final Results -----
               bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3048

A DNA sequence <SEQ ID 9267> was identified in GAS which encodes amino acid sequence <SEQ ID 9268>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.94   Transmembrane   27 -  43 (  19 -  51)
   INTEGRAL    Likelihood = -4.83   Transmembrane  152 - 168 ( 151 - 171)
   INTEGRAL    Likelihood = -4.09   Transmembrane  277 - 293 ( 276 - 294)
   INTEGRAL    Likelihood = -3.82   Transmembrane  195 - 211 ( 193 - 217)
   INTEGRAL    Likelihood = -2.50   Transmembrane  120 - 136 ( 120 - 137)
   INTEGRAL    Likelihood = -0.64   Transmembrane   81 -  97 (  81 -  98)

----- Final Results -----
               bacterial membrane --- Certainty=0.5776(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3049

A DNA sequence <SEQ ID 9269> was identified in GAS which encodes amino acid sequence <SEQ ID 9270>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.49   Transmembrane   27 -  43 (  14 -  50)
   INTEGRAL    Likelihood = -8.17   Transmembrane   58 -  74 (  52 -  79)
   INTEGRAL    Likelihood = -7.38   Transmembrane  165 - 181 ( 161 - 193)
   INTEGRAL    Likelihood = -3.66   Transmembrane  247 - 263 ( 246 - 270)
   INTEGRAL    Likelihood = -1.54   Transmembrane  134 - 150 ( 134 - 150)

----- Final Results -----
               bacterial membrane --- Certainty=0.440(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3050

A DNA sequence <SEQ ID 9271> was identified in GAS which encodes amino acid sequence <SEQ ID 9272>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.75   Transmembrane  389 - 405 ( 377 - 413)
   INTEGRAL    Likelihood = -8.44   Transmembrane   31 -  47 (  29 -  54)
   INTEGRAL    Likelihood = -7.17   Transmembrane  181 - 197 ( 179 - 205)
   INTEGRAL    Likelihood = -7.01   Transmembrane  339 - 355 ( 326 - 360)
   INTEGRAL    Likelihood = -6.58   Transmembrane  105 - 121 ( 102 - 124)
   INTEGRAL    Likelihood = -5.36   Transmembrane  225 - 241 ( 222 - 244)
   INTEGRAL    Likelihood = -0.43   Transmembrane  139 - 155 ( 139 - 155)
   INTEGRAL    Likelihood = -0.16   Transmembrane  283 - 299 ( 282 - 300)

----- Final Results -----
               bacterial membrane --- Certainty=0.6901(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3051

A DNA sequence <SEQ ID 9273> was identified in GAS which encodes amino acid sequence <SEQ ID 9274>. Analysis of the amino acid sequence reveals the following:

```
      Possible cleavage site: 25
>>> Seems to have a cleavable N-term signal seq.
     INTEGRAL    Likelihood = -5.31   Transmembrane  155 - 171 ( 154 - 174)
     INTEGRAL    Likelihood = -3.50   Transmembrane  111 - 127 ( 110 - 128)
     INTEGRAL    Likelihood = -2.07   Transmembrane   80 -  96 (  78 -  96)
     INTEGRAL    Likelihood = -0.90   Transmembrane   57 -  73 (  57 -  74)

----- Final Results -----
               bacterial membrane --- Certainty=0.312(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3052

A DNA sequence <SEQ ID 9275> was identified in GAS which encodes amino acid sequence <SEQ ID 9276>. Analysis of the amino acid sequence reveals the following:

```
      Possible site: 27
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.93   Transmembrane  463 - 479 ( 461 - 480)

----- Final Results -----
               bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3053

A DNA sequence <SEQ ID 8741> was identified in GBS which encodes amino acid sequence <SEQ ID 8742>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3054

A DNA sequence <SEQ ID 8685> was identified in GBS which encodes amino acid sequence <SEQ ID 8686>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3055

A DNA sequence <SEQ ID 10303> was identified in GBS which encodes amino acid sequence <SEQ ID 10304>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3056

A DNA sequence <SEQ ID 10305> was identified in GBS which encodes amino acid sequence <SEQ ID 10306>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3057

A DNA sequence <SEQ ID 10307> was identified in GBS which encodes amino acid sequence <SEQ ID 10308>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3058

A DNA sequence <SEQ ID 10309> was identified in GBS which encodes amino acid sequence <SEQ ID 10310>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3059

A DNA sequence <SEQ ID 10311> was identified in GBS which encodes amino acid sequence <SEQ ID 10312>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3060

A DNA sequence <SEQ ID 10313> was identified in GBS which encodes amino acid sequence <SEQ ID 10314>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3061

A DNA sequence <SEQ ID 10315> was identified in GBS which encodes amino acid sequence <SEQ ID 10316>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3062

A DNA sequence <SEQ ID 10317> was identified in GBS which encodes amino acid sequence <SEQ ID 10318>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3063

A repeated DNA sequence <SEQ ID 10319> was identified in GBS which encodes amino acid sequence <SEQ ID 10320>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3064

A DNA sequence <SEQ ID 10321> was identified in GBS which encodes amino acid sequence <SEQ ID 10322>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3065

A DNA sequence <SEQ ID 10323> was identified in GBS which encodes amino acid sequence <SEQ ID 10324>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3066

A DNA sequence <SEQ ID 10325> was identified in GBS which encodes amino acid sequence <SEQ ID 10326>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3067

A DNA sequence <SEQ ID 10327> was identified in GBS which encodes amino acid sequence <SEQ ID 10328>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3068

A DNA sequence <SEQ ID 10329> was identified in GBS which encodes amino acid sequence <SEQ ID 10330>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3069

A DNA sequence <SEQ ID 10331> was identified in GBS which encodes amino acid sequence <SEQ ID 10332>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3070

A DNA sequence <SEQ ID 10333> was identified in GBS which encodes amino acid sequence <SEQ ID 10334>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3071

A DNA sequence <SEQ ID 10335> was identified in GBS which encodes amino acid sequence <SEQ ID 10336>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3072

A DNA sequence <SEQ ID 10339> was identified in GBS which encodes amino acid sequence <SEQ ID 10340>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3073

A DNA sequence <SEQ ID 10341> was identified in GBS which encodes amino acid sequence <SEQ ID 10342>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3074

A DNA sequence <SEQ ID 10343> was identified in GBS which encodes amino acid sequence <SEQ ID 10344>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3075

A DNA sequence <SEQ ID 10345> was identified in GBS which encodes amino acid sequence <SEQ ID 10346>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3076

A DNA sequence <SEQ ID 10347> was identified in GBS which encodes amino acid sequence <SEQ ID 10348>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3077

A DNA sequence <SEQ ID 10349> was identified in GBS which encodes amino acid sequence <SEQ ID 10350>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3078

A DNA sequence <SEQ ID 10351> was identified in GBS which encodes amino acid sequence <SEQ ID 10352>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3079

A DNA sequence <SEQ ID 10353> was identified in GBS which encodes amino acid sequence <SEQ ID 10354>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3080

A DNA sequence <SEQ ID 10355> was identified in GBS which encodes amino acid sequence <SEQ ID 10356>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3081

A DNA sequence <SEQ ID 10357> was identified in GBS which encodes amino acid sequence <SEQ ID 10358>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3082

A DNA sequence <SEQ ID 10359> was identified in GBS which encodes amino acid sequence <SEQ ID 10360>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3083

A DNA sequence <SEQ ID 10361> was identified in GBS which encodes amino acid sequence <SEQ ID 10362>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3084

A DNA sequence <SEQ ID 10363> was identified in GBS which encodes amino acid sequence <SEQ ID 10364>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3085

A DNA sequence <SEQ ID 10365> was identified in GBS which encodes amino acid sequence <SEQ ID 10366>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3086

A DNA sequence <SEQ ID 10367> was identified in GBS which encodes amino acid sequence <SEQ ID 10368>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3087

A DNA sequence <SEQ ID 10369> was identified in GBS which encodes amino acid sequence <SEQ ID 10370>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3088

A DNA sequence <SEQ ID 10371> was identified in GBS which encodes amino acid sequence <SEQ ID 10372>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3089

A DNA sequence <SEQ ID 10373> was identified in GBS which encodes amino acid sequence <SEQ ID 10374>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3090

A DNA sequence <SEQ ID 10375> was identified in GBS which encodes amino acid sequence <SEQ ID 10376>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3091

A DNA sequence <SEQ ID 10377> was identified in GBS which encodes amino acid sequence <SEQ ID 10378>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3092

A DNA sequence <SEQ ID 10379> was identified in GBS which encodes amino acid sequence <SEQ ID 10380>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3093

A DNA sequence <SEQ ID 10381> was identified in GBS which encodes amino acid sequence <SEQ ID 10382>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3094

A DNA sequence <SEQ ID 10383> was identified in GBS which encodes amino acid sequence <SEQ ID 10384>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3095

A DNA sequence <SEQ ID 10385> was identified in GBS which encodes amino acid sequence <SEQ ID 10386>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3096

A DNA sequence <SEQ ID 10387> was identified in GBS which encodes amino acid sequence <SEQ ID 10388>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3097

A DNA sequence <SEQ ID 10389> was identified in GBS which encodes amino acid sequence <SEQ ID 10390>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3098

A DNA sequence <SEQ ID 10391> was identified in GBS which encodes amino acid sequence <SEQ ID 10392>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3099

A DNA sequence <SEQ ID 10393> was identified in GBS which encodes amino acid sequence <SEQ ID 10394>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3100

A DNA sequence <SEQ ID 10395> was identified in GBS which encodes amino acid sequence <SEQ ID 10396>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3101

A DNA sequence <SEQ ID 10397> was identified in GBS which encodes amino acid sequence <SEQ ID 10398>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3102

A DNA sequence <SEQ ID 10399> was identified in GBS which encodes amino acid sequence <SEQ ID 10400>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3103

A DNA sequence <SEQ ID 10401> was identified in GBS which encodes amino acid sequence <SEQ ID 10402>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3104

A DNA sequence <SEQ ID 10403> was identified in GBS which encodes amino acid sequence <SEQ ID 10404>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3105

A DNA sequence <SEQ ID 10405> was identified in GBS which encodes amino acid sequence <SEQ ID 10406>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3106

A DNA sequence <SEQ ID 10407> was identified in GBS which encodes amino acid sequence <SEQ ID 10408>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3107

A DNA sequence <SEQ ID 10409> was identified in GBS which encodes amino acid sequence <SEQ ID 10410>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3108

A DNA sequence <SEQ ID 10411> was identified in GBS which encodes amino acid sequence <SEQ ID 10412>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3109

A DNA sequence <SEQ ID 10413> was identified in GBS which encodes amino acid sequence <SEQ ID 10414>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3110

A DNA sequence <SEQ ID 10415> was identified in GBS which encodes amino acid sequence <SEQ ID 10416>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3111

A DNA sequence <SEQ ID 10417> was identified in GBS which encodes amino acid sequence <SEQ ID 10418>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3112

A DNA sequence <SEQ ID 10419> was identified in GBS which encodes amino acid sequence <SEQ ID 10420>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3113

A DNA sequence <SEQ ID 10421> was identified in GBS which encodes amino acid sequence <SEQ ID 10422>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3114

A DNA sequence <SEQ ID 10423> was identified in GBS which encodes amino acid sequence <SEQ ID 10424>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3115

A DNA sequence <SEQ ID 10425> was identified in GBS which encodes amino acid sequence <SEQ ID 10426>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3116

A DNA sequence <SEQ ID 10427> was identified in GBS which encodes amino acid sequence <SEQ ID 10428>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3117

A DNA sequence <SEQ ID 10429> was identified in GBS which encodes amino acid sequence <SEQ ID 10430>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3118

A DNA sequence <SEQ ID 10431> was identified in GBS which encodes amino acid sequence <SEQ ID 10432>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3119

A DNA sequence <SEQ ID 10433> was identified in GBS which encodes amino acid sequence <SEQ ID 10434>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3120

A DNA sequence <SEQ ID 10435> was identified in GBS which encodes amino acid sequence <SEQ ID 10436>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3121

A DNA sequence <SEQ ID 10437> was identified in GBS which encodes amino acid sequence <SEQ ID 10438>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3122

A DNA sequence <SEQ ID 10441> was identified in GBS which encodes amino acid sequence <SEQ ID 10442>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3123

A DNA sequence <SEQ ID 10443> was identified in GBS which encodes amino acid sequence <SEQ ID 10444>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3124

A DNA sequence <SEQ ID 10445> was identified in GBS which encodes amino acid sequence <SEQ ID 10446>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3125

A DNA sequence <SEQ ID 10447> was identified in GBS which encodes amino acid sequence <SEQ ID 10448>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3126

A DNA sequence <SEQ ID 10449> was identified in GBS which encodes amino acid sequence <SEQ ID 10450>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3127

A DNA sequence <SEQ ID 10451> was identified in GBS which encodes amino acid sequence <SEQ ID 10452>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3128

A DNA sequence <SEQ ID 10453> was identified in GBS which encodes amino acid sequence <SEQ ID 10454>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3129

A DNA sequence <SEQ ID 10455> was identified in GBS which encodes amino acid sequence <SEQ ID 10456>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3130

A DNA sequence <SEQ ID 10457> was identified in GBS which encodes amino acid sequence <SEQ ID 10458>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10907> which encodes amino acid sequence <SEQ ID 10908> was also identified.

Example 3131

A DNA sequence <SEQ ID 10459> was identified in GBS which encodes amino acid sequence <SEQ ID 10460>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3132

A DNA sequence <SEQ ID 10461> was identified in GBS which encodes amino acid sequence <SEQ ID 10462>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3133

A DNA sequence <SEQ ID 10463> was identified in GBS which encodes amino acid sequence <SEQ ID 10464>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3134

A DNA sequence <SEQ ID 10465> was identified in GBS which encodes amino acid sequence <SEQ ID 10466>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3135

A DNA sequence <SEQ ID 10467> was identified in GBS which encodes amino acid sequence <SEQ ID 10468>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3136

A DNA sequence <SEQ ID 10469> was identified in GBS which encodes amino acid sequence <SEQ ID 10470>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3137

A DNA sequence <SEQ ID 10471> was identified in GBS which encodes amino acid sequence <SEQ ID 10472>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3138

A DNA sequence <SEQ ID 10473> was identified in GBS which encodes amino acid sequence <SEQ ID 10474>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3139

A DNA sequence <SEQ ID 10475> was identified in GBS which encodes amino acid sequence <SEQ ID 10476>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3140

A DNA sequence <SEQ ID 10477> was identified in GBS which encodes amino acid sequence <SEQ ID 10478>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3141

A DNA sequence <SEQ ID 10479> was identified in GBS which encodes amino acid sequence <SEQ ID 10480>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3142

A DNA sequence <SEQ ID 10481> was identified in GBS which encodes amino acid sequence <SEQ ID 10482>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3143

A DNA sequence <SEQ ID 10483> was identified in GBS which encodes amino acid sequence <SEQ ID 10484>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3144

A DNA sequence <SEQ ID 10485> was identified in GBS which encodes amino acid sequence <SEQ ID 10486>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3145

A DNA sequence <SEQ ID 10487> was identified in GBS which encodes amino acid sequence <SEQ ID 10488>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3146

A DNA sequence <SEQ ID 10489> was identified in GBS which encodes amino acid sequence <SEQ ID 10490>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3147

A DNA sequence <SEQ ID 10491> was identified in GBS which encodes amino acid sequence <SEQ ID 10492>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3148

A DNA sequence <SEQ ID 10493> was identified in GBS which encodes amino acid sequence <SEQ ID 10494>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3149

A DNA sequence <SEQ ID 10495> was identified in GBS which encodes amino acid sequence <SEQ ID 10496>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3150

A DNA sequence <SEQ ID 10497> was identified in GBS which encodes amino acid sequence <SEQ ID 10498>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3151

A DNA sequence <SEQ ID 10499> was identified in GBS which encodes amino acid sequence <SEQ ID 10500>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3152

A DNA sequence <SEQ ID 10501> was identified in GBS which encodes amino acid sequence <SEQ ID 10502>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3153

A DNA sequence <SEQ ID 10503> was identified in GBS which encodes amino acid sequence <SEQ ID 10504>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3154

A DNA sequence <SEQ ID 10505> was identified in GBS which encodes amino acid sequence <SEQ ID 10506>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3155

A DNA sequence <SEQ ID 10509> was identified in GBS which encodes amino acid sequence <SEQ ID 10510>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3156

A DNA sequence <SEQ ID 10511> was identified in GBS which encodes amino acid sequence <SEQ ID 10512>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3157

A DNA sequence <SEQ ID 10513> was identified in GBS which encodes amino acid sequence <SEQ ID 10514>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3158

A DNA sequence <SEQ ID 10515> was identified in GBS which encodes amino acid sequence <SEQ ID 10516>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3159

A DNA sequence <SEQ ID 10517> was identified in GBS which encodes amino acid sequence <SEQ ID 10518>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3160

A DNA sequence <SEQ ID 10519> was identified in GBS which encodes amino acid sequence <SEQ ID 10520>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3161

A DNA sequence <SEQ ID 10521> was identified in GBS which encodes amino acid sequence <SEQ ID 10522>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3162

A DNA sequence <SEQ ID 10523> was identified in GBS which encodes amino acid sequence <SEQ ID 10524>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3163

A DNA sequence <SEQ ID 10525> was identified in GBS which encodes amino acid sequence <SEQ ID 10526>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3164

A DNA sequence <SEQ ID 10527> was identified in GBS which encodes amino acid sequence <SEQ ID 10528>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3165

A DNA sequence <SEQ ID 10529> was identified in GBS which encodes amino acid sequence <SEQ ID 10530>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3166

A DNA sequence <SEQ ID 10531> was identified in GBS which encodes amino acid sequence <SEQ ID 10532>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3167

A DNA sequence <SEQ ID 10533> was identified in GBS which encodes amino acid sequence <SEQ ID 10534>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3168

A DNA sequence <SEQ ID 10535> was identified in GBS which encodes amino acid sequence <SEQ ID 10536>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3169

A DNA sequence <SEQ ID 10537> was identified in GBS which encodes amino acid sequence <SEQ ID 10538>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3170

A DNA sequence <SEQ ID 10539> was identified in GBS which encodes amino acid sequence <SEQ ID 10540>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3171

A DNA sequence <SEQ ID 10541> was identified in GBS which encodes amino acid sequence <SEQ ID 10542>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3172

A DNA sequence <SEQ ID 10543> was identified in GBS which encodes amino acid sequence <SEQ ID 10544>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3173

A DNA sequence <SEQ ID 10545> was identified in GBS which encodes amino acid sequence <SEQ ID 10546>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

SEQ ID 10546 (GBS665) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 137 (lane 8-10; MW 41kDa) and in Figure 187 (lane 5; MW 41kDa). It was also was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 137 (lane 11 & 12; MW 16.1kDa), in Figure 141 (lane 4; MW 16kDa) and in Figure 179 (lane 6; MW 16kDa). Purified GBS665-GST is shown in Figure 243, lane 4.

GBS665-His was purified as shown in Figure 230, lane 7-8.

Example 3174

A DNA sequence <SEQ ID 10547> was identified in GBS which encodes amino acid sequence <SEQ ID 10548>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10909> which encodes amino acid sequence <SEQ ID 10910> was also identified.

Example 3175

A DNA sequence <SEQ ID 10549> was identified in GBS which encodes amino acid sequence <SEQ ID 10550>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3176

A DNA sequence <SEQ ID 10551> was identified in GBS which encodes amino acid sequence <SEQ ID 10552>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3177

A DNA sequence <SEQ ID 10553> was identified in GBS which encodes amino acid sequence <SEQ ID 10554>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3178

A DNA sequence <SEQ ID 10555> was identified in GBS which encodes amino acid sequence <SEQ ID 10556>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3179

A DNA sequence <SEQ ID 10557> was identified in GBS which encodes amino acid sequence <SEQ ID 10558>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3180

A DNA sequence <SEQ ID 10559> was identified in GBS which encodes amino acid sequence <SEQ ID 10560>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3181

A DNA sequence <SEQ ID 10561> was identified in GBS which encodes amino acid sequence <SEQ ID 10562>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3182

A DNA sequence <SEQ ID 10563> was identified in GBS which encodes amino acid sequence <SEQ ID 10564>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3183

A DNA sequence <SEQ ID 10565> was identified in GBS which encodes amino acid sequence <SEQ ID 10566>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3184

A DNA sequence <SEQ ID 10567> was identified in GBS which encodes amino acid sequence <SEQ ID 10568>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3185

A DNA sequence <SEQ ID 10569> was identified in GBS which encodes amino acid sequence <SEQ ID 10570>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3186

A DNA sequence <SEQ ID 10571> was identified in GBS which encodes amino acid sequence <SEQ ID 10572>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3187

A DNA sequence <SEQ ID 10573> was identified in GBS which encodes amino acid sequence <SEQ ID 10574>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3188

A DNA sequence <SEQ ID 10575> was identified in GBS which encodes amino acid sequence <SEQ ID 10576>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3189

A DNA sequence <SEQ ID 10577> was identified in GBS which encodes amino acid sequence <SEQ ID 10578>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3190

A DNA sequence <SEQ ID 10579> was identified in GBS which encodes amino acid sequence <SEQ ID 10580>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3191

A DNA sequence <SEQ ID 10581> was identified in GBS which encodes amino acid sequence <SEQ ID 10582>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3192

A DNA sequence <SEQ ID 10583> was identified in GBS which encodes amino acid sequence <SEQ ID 10584>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3193

A DNA sequence <SEQ ID 10585> was identified in GBS which encodes amino acid sequence <SEQ ID 10586>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3194

A DNA sequence <SEQ ID 10587> was identified in GBS which encodes amino acid sequence <SEQ ID 10588>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3195

A DNA sequence <SEQ ID 10591> was identified in GBS which encodes amino acid sequence <SEQ ID 10592>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3196

A DNA sequence <SEQ ID 10593> was identified in GBS which encodes amino acid sequence <SEQ ID 10594>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3197

A DNA sequence <SEQ ID 10595> was identified in GBS which encodes amino acid sequence <SEQ ID 10596>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3198

A DNA sequence <SEQ ID 10597> was identified in GBS which encodes amino acid sequence <SEQ ID 10598>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10797> which encodes amino acid sequence <SEQ ID 10798> was also identified.

Example 3199

A DNA sequence <SEQ ID 10599> was identified in GBS which encodes amino acid sequence <SEQ ID 10600>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3200

A DNA sequence <SEQ ID 10601> was identified in GBS which encodes amino acid sequence <SEQ ID 10602>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3201

A DNA sequence <SEQ ID 10603> was identified in GBS which encodes amino acid sequence <SEQ ID 10604>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3202

A DNA sequence <SEQ ID 10605> was identified in GBS which encodes amino acid sequence <SEQ ID 10606>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3203

A DNA sequence <SEQ ID 10607> was identified in GBS which encodes amino acid sequence <SEQ ID 10608>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3204

A DNA sequence <SEQ ID 10609> was identified in GBS which encodes amino acid sequence <SEQ ID 10610>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3205

A DNA sequence <SEQ ID 10611> was identified in GBS which encodes amino acid sequence <SEQ ID 10612>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3206

A DNA sequence <SEQ ID 10613> was identified in GBS which encodes amino acid sequence <SEQ ID 10614>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3207

A DNA sequence <SEQ ID 10615> was identified in GBS which encodes amino acid sequence <SEQ ID 10616>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3208

A DNA sequence <SEQ ID 10617> was identified in GBS which encodes amino acid sequence <SEQ ID 10618>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3209

A DNA sequence <SEQ ID 10619> was identified in GBS which encodes amino acid sequence <SEQ ID 10620>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3210

A DNA sequence <SEQ ID 10621> was identified in GBS which encodes amino acid sequence <SEQ ID 10622>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3211

A DNA sequence <SEQ ID 10623> was identified in GBS which encodes amino acid sequence <SEQ ID 10624>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3212

A DNA sequence <SEQ ID 10625> was identified in GBS which encodes amino acid sequence <SEQ ID 10626>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3213

A DNA sequence <SEQ ID 10627> was identified in GBS which encodes amino acid sequence <SEQ ID 10628>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3214

A DNA sequence <SEQ ID 10629> was identified in GBS which encodes amino acid sequence <SEQ ID 10630>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3215

A DNA sequence <SEQ ID 10631> was identified in GBS which encodes amino acid sequence <SEQ ID 10632>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3216

A DNA sequence <SEQ ID 10633> was identified in GBS which encodes amino acid sequence <SEQ ID 10634>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10939> which encodes amino acid sequence <SEQ ID 10940> was also identified.

SEQ ID 10634 (GBS675) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 162 (lane 14 & 15; MW 56kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 163 (lane 2; MW 31kDa) and in Figure 188 (lane 5; MW 31kDa).

Purified GBS675-His is shown in Figure 240, lane 7-8.

Example 3217

A DNA sequence <SEQ ID 10635> was identified in GBS which encodes amino acid sequence <SEQ ID 10636>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3218

A DNA sequence <SEQ ID 10637> was identified in GBS which encodes amino acid sequence <SEQ ID 10638>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3219

A DNA sequence <SEQ ID 10639> was identified in GBS which encodes amino acid sequence <SEQ ID 10640>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3220

A DNA sequence <SEQ ID 10641> was identified in GBS which encodes amino acid sequence <SEQ ID 10642>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3221

A DNA sequence <SEQ ID 10643> was identified in GBS which encodes amino acid sequence <SEQ ID 10644>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3222

A DNA sequence <SEQ ID 10645> was identified in GBS which encodes amino acid sequence <SEQ ID 10646>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3223

A DNA sequence <SEQ ID 10647> was identified in GBS which encodes amino acid sequence <SEQ ID 10648>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3224

A DNA sequence <SEQ ID 10649> was identified in GBS which encodes amino acid sequence <SEQ ID 10650>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3225

A DNA sequence <SEQ ID 10651> was identified in GBS which encodes amino acid sequence <SEQ ID 10652>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3226

A DNA sequence <SEQ ID 10653> was identified in GBS which encodes amino acid sequence <SEQ ID 10654>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3227

A DNA sequence <SEQ ID 10655> was identified in GBS which encodes amino acid sequence <SEQ ID 10656>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3228

A DNA sequence <SEQ ID 10657> was identified in GBS which encodes amino acid sequence <SEQ ID 10658>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3229

A DNA sequence <SEQ ID 10659> was identified in GBS which encodes amino acid sequence <SEQ ID 10660>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3230

A DNA sequence <SEQ ID 10661> was identified in GBS which encodes amino acid sequence <SEQ ID 10662>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3231

A DNA sequence <SEQ ID 10663> was identified in GBS which encodes amino acid sequence <SEQ ID 10664>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3232

A DNA sequence <SEQ ID 10665> was identified in GBS which encodes amino acid sequence <SEQ ID 10666>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10917> which encodes amino acid sequence <SEQ ID 10918> was also identified.

A DNA sequence <SEQ ID 10667> was identified in GBS which encodes amino acid sequence <SEQ ID 10668>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3233

A DNA sequence <SEQ ID 10669> was identified in GBS which encodes amino acid sequence <SEQ ID 10670>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3234

A DNA sequence <SEQ ID 10671> was identified in GBS which encodes amino acid sequence <SEQ ID 10672>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3235

A DNA sequence <SEQ ID 10673> was identified in GBS which encodes amino acid sequence <SEQ ID 10674>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3236

A DNA sequence <SEQ ID 10675> was identified in GBS which encodes amino acid sequence <SEQ ID 10676>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3237

A DNA sequence <SEQ ID 10677> was identified in GBS which encodes amino acid sequence <SEQ ID 10678>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3238

A DNA sequence <SEQ ID 10679> was identified in GBS which encodes amino acid sequence <SEQ ID 10680>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3239

A DNA sequence <SEQ ID 10681> was identified in GBS which encodes amino acid sequence <SEQ ID 10682>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3240

A DNA sequence <SEQ ID 10683> was identified in GBS which encodes amino acid sequence <SEQ ID 10684>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3241

A DNA sequence <SEQ ID 10685> was identified in GBS which encodes amino acid sequence <SEQ ID 10686>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3242

A DNA sequence <SEQ ID 10687> was identified in GBS which encodes amino acid sequence <SEQ ID 10688>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3243

A DNA sequence <SEQ ID 10689> was identified in GBS which encodes amino acid sequence <SEQ ID 10690>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3244

A DNA sequence <SEQ ID 10691> was identified in GBS which encodes amino acid sequence <SEQ ID 10692>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

SEQ ID 10692 (GBS676) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 163 (lane 3-5; MW 66kDa) and in Figure 239 (lane 8; MW 66kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 163 (lane 7 & 8; MW 41kDa) and in Figure 188 (lane 6; MW 41kDa). Purified GBS676-His is shown in Figure 240, lane 4-5. Purified GBS676-GST is shown in Figure 246, lanes 10 & 11.

Example 3245

A DNA sequence <SEQ ID 10693> was identified in GBS which encodes amino acid sequence <SEQ ID 10694>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3246

A DNA sequence <SEQ ID 10695> was identified in GBS which encodes amino acid sequence <SEQ ID 10696>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3247

A DNA sequence <SEQ ID 10697> was identified in GBS which encodes amino acid sequence <SEQ ID 10698>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3248

A DNA sequence <SEQ ID 10699> was identified in GBS which encodes amino acid sequence <SEQ ID 10700>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3249

A DNA sequence <SEQ ID 10703> was identified in GBS which encodes amino acid sequence <SEQ ID 10704>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3250

A DNA sequence <SEQ ID 10705> was identified in GBS which encodes amino acid sequence <SEQ ID 10706>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3251

A DNA sequence <SEQ ID 10707> was identified in GBS which encodes amino acid sequence <SEQ ID 10708>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3252

A DNA sequence <SEQ ID 10709> was identified in GBS which encodes amino acid sequence <SEQ ID 10710>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10803> which encodes amino acid sequence <SEQ ID 10804> was also identified.

Example 3253

A DNA sequence <SEQ ID 10711> was identified in GBS which encodes amino acid sequence <SEQ ID 10712>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10913> which encodes amino acid sequence <SEQ ID 10914> was also identified.

Example 3254

A DNA sequence <SEQ ID 10713> was identified in GBS which encodes amino acid sequence <SEQ ID 10714>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3255

A DNA sequence <SEQ ID 10715> was identified in GBS which encodes amino acid sequence <SEQ ID 10716>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3256

A DNA sequence <SEQ ID 10717> was identified in GBS which encodes amino acid sequence <SEQ ID 10718>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3257

A DNA sequence <SEQ ID 10719> was identified in GBS which encodes amino acid sequence <SEQ ID 10720>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3258

A DNA sequence <SEQ ID 10721> was identified in GBS which encodes amino acid sequence <SEQ ID 10722>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3259

A DNA sequence <SEQ ID 10723> was identified in GBS which encodes amino acid sequence <SEQ ID 10724>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3260

A DNA sequence <SEQ ID 10725> was identified in GBS which encodes amino acid sequence <SEQ ID 10726>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3261

A DNA sequence <SEQ ID 10727> was identified in GBS which encodes amino acid sequence <SEQ ID 10728>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3262

A DNA sequence <SEQ ID 10729> was identified in GBS which encodes amino acid sequence <SEQ ID 10730>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

SEQ ID 10730 (GBS670) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 140 (lane 2-4; MW 45.3kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 140 (lane 5-7; MW 20.4kDa) and in Figure 179 (lane 10; MW 20kDa).

GBS670-His was purified as shown in Figure 230, lane 9-10.

Example 3263

A DNA sequence <SEQ ID 10731> was identified in GBS which encodes amino acid sequence <SEQ ID 10732>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3264

A DNA sequence <SEQ ID 10733> was identified in GBS which encodes amino acid sequence <SEQ ID 10734>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3265

A DNA sequence <SEQ ID 10735> was identified in GBS which encodes amino acid sequence <SEQ ID 10736>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3266

A DNA sequence <SEQ ID 10737> was identified in GBS which encodes amino acid sequence <SEQ ID 10738>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3267

A DNA sequence <SEQ ID 10739> was identified in GBS which encodes amino acid sequence <SEQ ID 10740>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3268

A DNA sequence <SEQ ID 10741> was identified in GBS which encodes amino acid sequence <SEQ ID 10742>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3269

A DNA sequence <SEQ ID 10743> was identified in GBS which encodes amino acid sequence <SEQ ID 10744>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3270

A DNA sequence <SEQ ID 10745> was identified in GBS which encodes amino acid sequence <SEQ ID 10746>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3271

A DNA sequence <SEQ ID 10747> was identified in GBS which encodes amino acid sequence <SEQ ID 10748>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3272

A DNA sequence <SEQ ID 10749> was identified in GBS which encodes amino acid sequence <SEQ ID 10750>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3273

A DNA sequence <SEQ ID 10751> was identified in GBS which encodes amino acid sequence <SEQ ID 10752>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3274

A DNA sequence <SEQ ID 10753> was identified in GBS which encodes amino acid sequence <SEQ ID 10754>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3275

A DNA sequence <SEQ ID 10755> was identified in GBS which encodes amino acid sequence <SEQ ID 10756>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3276

A DNA sequence <SEQ ID 10757> was identified in GBS which encodes amino acid sequence <SEQ ID 10758>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3277

A DNA sequence <SEQ ID 10759> was identified in GBS which encodes amino acid sequence <SEQ ID 10760>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3278

A DNA sequence <SEQ ID 10761> was identified in GBS which encodes amino acid sequence <SEQ ID 10762>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3279

A DNA sequence <SEQ ID 10763> was identified in GBS which encodes amino acid sequence <SEQ ID 10764>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3280

A DNA sequence <SEQ ID 10765> was identified in GBS which encodes amino acid sequence <SEQ ID 10766>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3281

A DNA sequence <SEQ ID 10767> was identified in GBS which encodes amino acid sequence <SEQ ID 10768>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3282

A DNA sequence <SEQ ID 10769> was identified in GBS which encodes amino acid sequence <SEQ ID 10770>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3283

A DNA sequence <SEQ ID 10771> was identified in GBS which encodes amino acid sequence <SEQ ID 10772>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3284

A repeated DNA sequence <SEQ ID 10791> was identified in GBS which encodes amino acid sequence <SEQ ID 10792>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3285

A DNA sequence <SEQ ID 10805> was identified in GBS which encodes amino acid sequence <SEQ ID 10806>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3286

A DNA sequence <SEQ ID 10807> was identified in GBS which encodes amino acid sequence <SEQ ID 10808>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3287

A DNA sequence <SEQ ID 10809> was identified in GBS which encodes amino acid sequence <SEQ ID 10810>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3288

A DNA sequence <SEQ ID 10811> was identified in GBS which encodes amino acid sequence <SEQ ID 10812>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3289

A DNA sequence <SEQ ID 10813> was identified in GBS which encodes amino acid sequence <SEQ ID 10814>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3290

A DNA sequence <SEQ ID 10815> was identified in GBS which encodes amino acid sequence <SEQ ID 10816>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3291

A DNA sequence <SEQ ID 10817> was identified in GBS which encodes amino acid sequence <SEQ ID 10818>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3292

A DNA sequence <SEQ ID 10819> was identified in GBS which encodes amino acid sequence <SEQ ID 10820>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3293

A DNA sequence <SEQ ID 10821> was identified in GBS which encodes amino acid sequence <SEQ ID 10822>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3294

A DNA sequence <SEQ ID 10823> was identified in GBS which encodes amino acid sequence <SEQ ID 10824>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3295

A DNA sequence <SEQ ID 10825> was identified in GBS which encodes amino acid sequence <SEQ ID 10826>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3296

A DNA sequence <SEQ ID 10827> was identified in GBS which encodes amino acid sequence <SEQ ID 10828>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3297

A DNA sequence <SEQ ID 10829> was identified in GBS which encodes amino acid sequence <SEQ ID 10830>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3298

A DNA sequence <SEQ ID 10831> was identified in GBS which encodes amino acid sequence <SEQ ID 10832>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3299

A DNA sequence <SEQ ID 10833> was identified in GBS which encodes amino acid sequence <SEQ ID 10834>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3300

A DNA sequence <SEQ ID 10835> was identified in GBS which encodes amino acid sequence <SEQ ID 10836>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3301

A DNA sequence <SEQ ID 10837> was identified in GBS which encodes amino acid sequence <SEQ ID 10838>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3302

A DNA sequence <SEQ ID 10839> was identified in GBS which encodes amino acid sequence <SEQ ID 10840>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3303

A DNA sequence <SEQ ID 10841> was identified in GBS which encodes amino acid sequence <SEQ ID 10842>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3304

A DNA sequence <SEQ ID 10843> was identified in GBS which encodes amino acid sequence <SEQ ID 10844>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3305

A DNA sequence <SEQ ID 10845> was identified in GBS which encodes amino acid sequence <SEQ ID 10846>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3306

A DNA sequence <SEQ ID 10847> was identified in GBS which encodes amino acid sequence <SEQ ID 10848>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3307

A DNA sequence <SEQ ID 10849> was identified in GBS which encodes amino acid sequence <SEQ ID 10850>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3308

A DNA sequence <SEQ ID 10851> was identified in GBS which encodes amino acid sequence <SEQ ID 10852>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3309

A DNA sequence <SEQ ID 10853> was identified in GBS which encodes amino acid sequence <SEQ ID 10854>. Related sequences are <SEQ ID 10855>, <SEQ ID 10856>, <SEQ ID 10857>, <SEQ ID 10858>, <SEQ ID 10859>, <SEQ ID 10860>, <SEQ ID 10861>, <SEQ ID 10862>, <SEQ ID 10863>, <SEQ ID 10864>, <SEQ ID 10865> and <SEQ ID 10866>. These proteins and their epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3310

A DNA sequence <SEQ ID 10867> was identified in GBS which encodes amino acid sequence <SEQ ID 10868>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3311

A DNA sequence <SEQ ID 10869> was identified in GBS which encodes amino acid sequence <SEQ ID 10870>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3312

A DNA sequence <SEQ ID 10871> was identified in GBS which encodes amino acid sequence <SEQ ID 10872>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3313

A DNA sequence <SEQ ID 10873> was identified in GBS which encodes amino acid sequence <SEQ ID 10874>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3314

A DNA sequence <SEQ ID 10875> was identified in GBS which encodes amino acid sequence <SEQ ID 10876>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3315

A DNA sequence <SEQ ID 10877> was identified in GBS which encodes amino acid sequence <SEQ ID 10878>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3316

A DNA sequence <SEQ ID 10879> was identified in GBS which encodes amino acid sequence <SEQ ID 10880>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3317

A DNA sequence <SEQ ID 10881> was identified in GBS which encodes amino acid sequence <SEQ ID 10882>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3318

A DNA sequence <SEQ ID 10883> was identified in GBS which encodes amino acid sequence <SEQ ID 10884>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3319

A DNA sequence <SEQ ID 10885> was identified in GBS which encodes amino acid sequence <SEQ ID 10886>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3320

A DNA sequence <SEQ ID 10887> was identified in GBS which encodes amino acid sequence <SEQ ID 10888>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3321

A DNA sequence <SEQ ID 10889> was identified in GBS which encodes amino acid sequence <SEQ ID 10890>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3322

A DNA sequence <SEQ ID 10891> was identified in GBS which encodes amino acid sequence <SEQ ID 10892>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3323

A DNA sequence <SEQ ID 10893> was identified in GBS which encodes amino acid sequence <SEQ ID 10894>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3324

A DNA sequence <SEQ ID 10895> was identified in GBS which encodes amino acid sequence <SEQ ID 10896>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3325

A DNA sequence <SEQ ID 10897> was identified in GBS which encodes amino acid sequence <SEQ ID 10898>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3326

A DNA sequence <SEQ ID 10899> was identified in GBS which encodes amino acid sequence <SEQ ID 10900>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3327

A DNA sequence <SEQ ID 10901> was identified in GBS which encodes amino acid sequence <SEQ ID 10902>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3328

A DNA sequence <SEQ ID 10903> was identified in GBS which encodes amino acid sequence <SEQ ID 10904>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3329

Seven rRNA genes were identified in *S.agalactiae*. These are SEQ IDs 12018 to 12024. These rRNA genes are particularly useful for diagnostic purposes and for phlyogenetic studies. An alignment of the rRNA sequences is shown below:

```
12023  ------------------------------------TTTCGAGTCAAAGTCATCAGCGTT
12024  ------------------------------------------------------------
12019  -----TCCAATCATACTTAATTTCACTAATATCTGGATTTTGACATATTCAGTTAATTCT
12021  ...ATCGAATTGAACGGACTCAATTTGGTTGTTATGTAATTTT--ACATAATCTATGATTTCT
12020  ------------------------------------------------------------
12018  ------------------------------------------------------------
12022  ----------------------------CTTCTTTGTTTTCTTTAGAGATATTAACTGTA

12023  TACTGTTACGGCAGCAGTTCCAAGAGTTACTCCACTCACAAGGACTGCTGATAATATTCT
12024  ------------------------------------------------------------
12019  TTTTCATGCTTTTTGAGATAAGCTACTTGTTCTTTTTTTATTACTTTTTTACCTTTCTTT
12021  TGCTCATGCTCTTTGAGATAGGCTAATTGTTCTTTTTTTGTCATTTTTTTATCTTTCTTC
12020  ------------------------------------------------------------
12018  ------------------------------------------------------------
12022  CCCACTTTGGGCGTTAAAATACCTAAAGTAGCCTTTATTAAAGTTGATTTAGCAGCCCCA

12023  TTTTTTCATTTTTATTAAACTACTCCTTTAC--GATAAGACATTAAATATTTTACCAAAA
12024  ------------------------------------------------------------
12019  ACTGCTGACTGTTTGCTATTTTTTACTTCGTTTGACTGACTTTTAGATTCACTATTCATT
12021  ACTTCTGATTGCTTGCTATTTTTTACTTCGTTTGACTGAATTTTATGTTCACTATTCATT
12020  ------------------------------------------------------------
12018  ----------------------------CTTT-GATACAATATTATCAAAATTATATTAA
12022  TTTTCACCTGTTAAGGTAACAAACTCCCCACT-GTCTAAATGGTAATTAACCCCTTCCAG

12023  AATTCACGAAATTATATTACGTCATTGTTACATTTATATTTGAAATCAACTATTTCTAAA
12024  ------------------------------------------------------------
12019  TGACAGCCTGCTAGTAACATCCCAATAATAGATATGGGAATTAACCATTTTACATATTTT
12021  TGACAGCCTCCAAGTATCATCCCAAAAATTGATATGGGAATTAACCATTTTATATATTTT
12020  ------------------------------------------------------------
12018  CGGTAAAGATATTGTTAAAGACCAAACTTGGATTATCAATCGT----TATCAAGAAATTA
12022  CA-CAGGATCGCTATCGTACTGAAAAGTAAGACCACTAACTGTAATATATCGCATGATTA

12023  TGAACCATAATCAAATCTAGAAAACGATAACCTTCTTCTATTCACTCT---ATCAATATA
12024  ------------------------------------------------------------
12019  TTCAACATGCTCTCTTTTCTTAGAAAATAAACTTCCCATGTCAAGTATCTAATAAAAATA
12021  CTCATCATGTTCTCTTTTCTTAGAATATAAATTTTATATATCAAGTATATAATGAAATTA
12020  ------------------------------------------------------------
12018  TTAGTG---ATTTGTCTTTAGGAAGCACTA--------TTGCAGAAGA---AATTACTCG
12022  CCCTTCT--AATTCTCTAGAGAAAAGATCAAGAAAACGTTCTAAAACG---ACCTTTTCG

12023  ATTACTCCATAGTGAAACTAAAAGAGAAATAAAAAAAGAGTATAATTACTCTTAAAATTA
12024  ------------------------------------------------------------
12019  ATTATTATTTACCAGTATGTTAAAACTAATATTAGTATAACAAA-TTTTCACGAGTTTAA
```

```
12021  ACTATTATTCACCAACATTATAAAATTAATTTTAGTATAACAAAATTTTCACGTATTTTT
12020  ----------ATCAAAAAAACATGACCAGTATGAATTAAAGCAACGTATAATCAATGCCT
12018  CTCTAT--AGAGCAGCTAGCTTCACTTCCCATAGAAAATAATCAGTTTTTAT-ATGAT--
12022  TCCTTTGAAAAATGATTTACTAATCTTCCGTAAACCCCTAACGTATTGTCATGATGATGT

12023  TAATATTTACGGAGAATAAGGGATTCGAACCCTTGCGCCAGTTACCCGACCTAACGATTT
12024  ------------------------------------------------------------
12019  TT--TTTTAGTCGTAACATATACACTGAAAAATCTTATTATTTTATACTACCTATCTATC
12021  ATAGTTTTAGTCTTAACATGTAAACAGAAA------A----------------------TC
12020  TAATGCGTAAAGGATACCAGTACGAAGATA----------------------------TC
12018  ---TGTTTTTTAGCAGCCGGTGAAGATA-------------------ACAACGCAAAGTT
12022  GTGTGTTCATCTGCAATGGGTTTAGCAAGT----TCA---------GATAACTCAAAATA

12023  AGCAAACCGTCCTCTTCAGCCTCTTGAG--TAATTCTCCAAATTAATATTAATGGGCACG
12024  ------------------------------------------------------------
12019  ATTCACAAACACTTTTATTACTTCAGAACCTATGACATTTAGGAGTCCTCTTTGAATTTC
12021  ATTTGTATA-----T----------------------TTTAAATGCCCTAATTAAATT--
12020  AAAAGTGC------T----------------------TTAAGAGAATATTTATAAGAT--
12018  AGTTGCA-ACGTTTTTTAATCAAAATGA--CATTCCTGCAAGATATGTTCATCCAAACGA
12022  AGTAATACGAGCATCTTTAGAATCTTTA--TTCGCTTTCAACATATCCTGAGA-AATTAA

12023  AGTGGACTCGAACCACCGACCTCACGCTTATCAGGCGTGCGCTCTAACCACCTGAGCTAC
12024  ------------------------------------------------------------
12019  ATTTAAATGTTGAGTCTCCACTAACTCTTGAAAAATTTCCTTATTATTTCTGCTTGTTTT
12021  ------------------------------AATAATT-----AATATTTATTATTATATA
12020  ------------------------------AATAACTCTCAGACGATGTATT-TTACAGA
12018  AGCAGGAATTATTGTAACTAAAGAACCATG--TAATGCACGAATTATT--CCAG----GA
12022  ACTTTTTACTGCTTTAGTTACAGCTGCCTGACTAATATTTAACTTCTTAGCTAAATCAGA

12023  GCGCCCAAGCAAATGCTTGGTTTTACTTTTATGTAAAGTAAGCGGGTGACGAGAATCGA-
12024  ------------------------------------------------------------
12019  AAACCTTCTATAACGATTGCAATAATGAAAAACAAATATAAGTAATTTTCAGTAACTTTT
12021  AATTCTTCTACAATGA----------AAAAAATAAATATAT--A-TTACAAGTAACATT-
12020  AAAT----TATGATAA-----A----CTATAACAGACGTAT--AAATTGTAGAAAGTTG-
12018  AGTTATGATAAGATTGA------GAACTTATGTCTATACAATGAGGTTCTTGTTATCCCT
12022  ATTTGTCAACTGCTCTT------GTGATAAAAGCATCAGAATGTGTTCTTGCGTATTAGT

12023  -ACTCGCGACAACAGCTTGGAAGGCTGTAGTTTTACCACTAAACTACACCCGCTAAAAAC
12024  ------------------------------------------------------------
12019  TCTCAAAATTACCAGCACAATACAAAAAAGACAAGGCTTCTAAACCTTGTCTTTATAAAT
12021  --TCACAATAAATTATCTAGTAGAAAAAAGACAAGGTTTAGAAACCTTGTCTTTATAAGT
12020  ----GTAGGCTATGAGATTACCTAAAGAAGGCGACTTTATTACAATTCAAAGTTACAAAC
12018  GGATTT---------TTTGG--AGTCACAGAAGATAAC-CAAATTTGTACCTTTTCAAGA
12022  CAATTTAA-CATCACTTTGACAAGTACCAAACAATAATTCATGTTGATTTTCTGCTTTAA

12023  TTATATAATAAATGGCGCGAGACGGAATCGAACCGCCGACACATGGAGCTTCAATCCATT
12024  ------------------------------------------------------------
12019  ATACCGGCGGCCGGGGTCGAACCGGCACGTCCGTGAGGACACTGGATTTTGAGTCCAGCG
12021  ATACCGGCGGCCGGGGTCGAACCGGCACGTCCGTGAGGACACTGGATTTTGAGTCCAGCG
12020  ATGATGGTAGTTTACACCGAACTTG-----GCGTGACACCA-TGGTATTAAAAACAACCG
12018  GGGGGATCTGACATTACTGGATC--------CCTAATTGC--------AGCAGGCATAAA
12022  GCAAGATTTGAC-TCACTAAATGG-------TCTAATTTTTGTTCTAAAACTGTCATATA

12023  GCTCTACCAACTGAGCTACCGAGCCTATTGCGGGAGCAGGATTTGAACCTACGACCTTCG
12024  ------------------------------------------------------------
12019  CGTCTGCCAATTCCGCCACGCCGGCTATCTTAAAACTGGGGTAGCTGGATTCGA--ACCA
12021  CGTCTGCCAATTCCGCCACGCCGGCTATCTTAAAACTGGGGTAGCTGGATTCGA--ACCA
12020  AAAATGCC--CTCATTGGTGTTAATGATCAT---ACTTTAGTAACAGAAAATGATGGTCG
12018  AGCAGACCT-TTATGAGAACTTCACAGATGT----TGATGGTATATTTGCAGCACATCCA
12022  TACCT-CTT-TTTTGTTAACCAGTAAATTATATCACGAAGATATAGAAGAATCAATCATA

12023  GGTTA-TGAGCCCGACGAGCTACCTAGCTGCTCCA-------TCCCGCGATATCTTTAAA
12024  ------------------------------------------------------------
12019  ACGCA-TGAGGGAGTCAAAGTCCCTTGCCTTACCG-------CTTGGCTATACCCCATGA
12021  ACGCA-TGAGGGAGTCAAAGTCCCTTGCCTTACCG-------CTTGGCTATACCCCATGA
12020  ACGC--TGGGTGACACGAGAGCC--TGCAATA----------GTATACTTTCATA
12018  GGT-------GTAGTTAAGAACCCTCACGCTA----------TCCCTGAGCTTACTTATA
12022  GATAGGTGAAGAAGATAAAACCTTTTATCTCAACAACCTAACTTTATAAACTTCTTTGCA
```

```
12023   GGA---------GGATGTGGGATTCGAACCCACGCACGCTTTTACAC--GCCTGACGGTT
12024   ------------------------------------------------------------
12019   AAAGGCG-----AGTGATGGGAATCGAACCCACGAATGTCAGAGCCACAATCTGATGTGT
12021   AAAGGCG-----AGTGATGGGAATCGAACCCACGAATGTCAGAGCCACAATCTGATGTGT
12020   AAA---------AATACTGG---T------TT--AACATTATCGCTA-----TGATACGT
12018   AAGA--------AATGCGTGAATTAGCCTATGCGGGTTTTTCGGTTT-TACATGATGAA-
12022   AAAACCTTTCATACTATTAAAAACACGATCAGCTTTTTTCTCTGTAG-AACACATTGAAA

12023   TTCAAGACCGTTCCCTTCAGCCGGACTTGGGTAATCCTCCATATAACAAAAAATATGGAC
12024   ------------------------------------------------------------
12019   TAACCACTTCACCACACCCGCCATATTAGAAAAAACACGGGCAGTAGGAATCGAACCCAC
12021   TAACCACTTCACCACACCCGCCATATTAGAAAAAACACGGGCAGTAGGAATCGAACCCAC
12020   GAAACTGGTGTCTCCTACTATTGTAATCTAGCAAGT-----CCGTATATCTTGGACCC--
12018   --------GCTTTACTTCCTGCCTATCGTGGCAGAATCCCTCTTGTTATTAAAAATAC--
12022   AAACAGTTGGTCCACTTCCTGTC-ATTAATGCAACATCGGCTCCAGAATTTAACATAC--

12023   CTTGTAGGACTCGAACCTACGACCGCTCGGTTATGAGCCGAGTGCTCTAACCAGTTGAGC
12024   ------------------------------------------------------------
12019   ACTGAAGGTTTTGGAGACCTTAGTTCTACCTTTAAACTATGCCCGTTTACTATGGAGAGA
12021   ACTGAAGGTTTTGGAGACCTTAGTTCTACCTTTAAACTATGCCCGTTTACTATGGAGAGA
12020   --TGAAGCACTCAAGTATATTGACTATGACCTTGATGTCAAAGTATTTGCAGATGGTGAA
12018   ----AAA----TAATCCCCAACAGCCTGGTACAAAAATAGTTTTAAAGCATACTCGTAG-
12022   ----GTTCTTTTATTGTACTTATAACTGGATTTTTAGTAATTGTAATATCCTCGAGTGAA

12023   TAAAGGTCCAAAGTCTCAATAAAATAAATAGCGGCGGAGGGGATCGAACCCCCGACCTCC
12024   ------------------------------------------------------------
12019   GAGGGATTCGAACCCCCGAACCCGAAGGAGCGGATTTACAGTCCGCCGCGTTTAGCCTCT
12021   GAGGGATTCGAACCCCCGAACCCGAAGGAGCGGATTTACAGTCCGCCGCGTTTAGCCTCT
12020   AAAAGACTACTAGATGTGGACGAATATGAACAGCATAAAGYTCAGATGAACT--ATCCTA
12018   --TAACATAGCAGTAACTGG-GATCGCT--TCTGATAGCCGTTTTGCTAGCATAAACGTA
12022   TTTCCCATAGATTTGACCATTAACTGATAATCTGATGACAAAATAGCAGACTTTAATAAA

12023   CGGGTATG-AACCGGACGCTCTAGCCAGCT--GAGCTACACCGCCATAAAAATATATCCA
12024   ------------------------------------------------------------
12019   TCGCTATC-TCTCCTAAGGTATAAATGGCGCGAGACGGAATCGAACCGCCGACACATGGA
12021   TCGCTATC-TCTCCTAAGGTATAAATGGCGCGAGACGGAATCGAACCGCCGACACATGGA
12020   CCGATATT-GATTATATATTAAAGGAAAATGTAAAAATATTGGTAGAATGGATAAATGAG
12018   TCTAAAT--ACTTAATGAATAGA---GAAGTAGGTTTCGGCCGAAAAG----TACTACAA
12022   TCAATATCAACTCTACTTATAGACTTACAATCAATATCTCTAAAAATGGATTTAGTTGAA

12023   TCGGGAAGACAGGATTCGAACCTGCGACACCTTGGTCCCAAACCAAGTACTCTACCAAGC
12024   ------------------------------------------------------------
12019   GCTTCAATCCATTGCTCTACCAACTGAGCTACCGAGCCTATTGCGGGAGCAGGATTTGAA
12021   GCTTCAATCCATTGCTCTACCAACTGAGCTACCGAGCCTATTGCGGGAGCAGGATTTGAA
12020   AATAAAGGCCCCTTTTC-ATCATC--ATATATCAA-TATCTGGTATAAACGGTA------
12018   ATTTTAGAG---GATTTAAATATT---AGTTTTGAACATATGCCAACTGGCATAGATGAT
12022   ATACCAAAATCCGGCTTAACCAGA---ACTATCCAACATGGTCTCAATGTCGGTAAGGGT

12023   TGAGCTACTTCCCGAAAAATATGCAC--CCTAGAGGAGTCGAACCTCTAACCGCCTGATT
12024   ------------------------------------------------------------
12019   CCTACGACCTTCGGGTTATGAGCCCG--ACGAGCTACCTAGCTGCTCCATCCCGCGATAT
12021   CCTACGACCTTCGGGTTATGAGCCCG--ACGAGCTACCTAGCTGCTCCATCCCGCGATAT
12020   --------CCTTGAATTGAAA-------AAGCGCTAACTAAC-ACACTAAATAGTG-TGT
12018   CTATCCATTGT---CTTACGTGAAA---AAGAATTGACACCAATCAAAGAACAAGAAATC
12022   TTAACAATTTCACCTTTACCTAATACTAACGAACATCCCCCCACCAAGACAATAAGGAACA

12023   CGTAGTCAG---GTACTCTATCCAGTTGAGCTAAGGGTGCTAAATATTATA-----TGCC
12024   ------------------------------------------------------------
12019   CTTTAAAGGAGGATGTGGGATTCGAACCCACGCACGCTTTTACACGCCTGACG--GTTTT
12021   CTTTAAAGGAGGATGTGGGATTCGAACCCACGCACGCTTTTACACGCCTGACG--GTTTT
12020   TTTTATTA----ATATCAAATTTAATTACA---ATACTATTGCAAAAATAT----ATACT
12018   TTAAATTACCTAACTCGTAAACTAGAAGTAG--ATTACGTTGACATCCAA----------
12022   TC--ACTACC-AATTTTAAAACCAATAGCAACCATTTCGTCATAGTCCATTTGAAGATTC

12023   GAGGACCGGAATC----GAACCGGTACGATGTTTACCATCGCAGGATTTTAAGTCCTGTG
12024   ------------------------------------------------------------
12019   CAAGACCGTTCCCTTCAGCCGGACTTGGGTAATCCTCCATATAACAAAAAATAGTCCGTA
```

```
12021  CAAGACCGTTCCCTTCAGCCGGACTTGGGTAATCCTCCATATAACAAAAAATAGTCCGTA
12020  TAAAATAAA-------AAAAGTAGAAAGATCACTTTCTACTTTTTTAAGAATAGTCCGTA
12018  CACAATCTATC-------TACAATCGTAATTGTAGGTGAAA-ATATGAAAAGTCAGATTG
12022  CATAATCGATT-------AAGAGCTCTTATTGTAGCAGCAGCATCAGTAGAACCACCCCC

12023  CGTCTGCCAGTTCCGCCACCCCGGCCTCTAACAAGCGAACGACGGGGTTCGAACCCGCGA
12024  ------------------------------------------------------------
12019  CGGGATTCGAACCCGTGTTACCGCCGTGAAAAGGCGGTGTCTTAACCCCTTGACCAACGG
12021  CGGGATTCGAACCCGTGTTACCGCCGTGAAAAGGCGGTGTCTTAACCCCTTGACCAACGG
12020  CGGGATTCGAACCCGTGTTACCGCCGTGAAAAGGCGGTGTCTTAACCCCTTGACCAACGG
12018  GAGTCACTGCAACAGCGACACAAGCCTTATC------AAGAGAAAAA-----ATCAATAT
12022  CAGTC-CTGCACAGACAGGAATGGATTTTTCTAATCTAATATGAACACCTTTATTAATAC

12023  -CCCTCACCTT-----GGCAAGGTGATGTTCTACCACTGAACTACGTTCGCACTAAAGAC
12024  ------------------------------------------------------------
12019  -ACCATATTCTTGATGGGCACGAGTGGACTCGAACCACCGACCTCACGCTTATCAGGCGT
12021  -ACCATATTCTTGATGGGCACGAGTGGACTCGAACCACCGACCTCACGCTTATCAGGCGT
12020  -ACCATATTCTTGATGGGCACGAGTGGACTCGAACCACCGACCTCACGCTTATCAGGCGT
12018  CACCATGAT-----ATCACA-----AGGTTCAAGCGAA--GTCTCCATTATGT-------
12022  CATATTGATTTTTGATTATATCTGCAGCTTTAAACACATCATTATCATTATTTAAAGGCA

12023  ACTATTTATCCTATAAAATTGTAATGCCGGC------------------------------
12024  ------TATCCTATAAAATTGTAATGCCGGC------------------------------
12019  GCGCTCTAACCACCTGAGCTACGCGCCCAAAATAACTTCTAAAATTATAAAGTTAATGCC
12021  GCGCTCTAACCACCTGAGCTACGCGCCCAAG-------CTA--------------------
12020  GCGCTCTAACCACCTGAGCTACGCGCCCAAG-------CTA--------------------
12018  --TCGTTATAAACAGTAAGGATGAAAAAAGAG------------------------CTAT
12022  TTTTGCTACTATCAGAATCGATAACAATACAAT-----CTT---------------CCTT
             **        *

12023  ---TACATGACTTGAACACGCGACCCTCTGATTACAAATCAGATGCTCTACCAACTGAGC
12024  ---TACATGACTTGAACACGCGACCCTCTGATTACAAATCAGATGCTCTACCAACTGAGC
12019  GGCTACATGACTTGAACACGCGACCCTCTGATTACAAATCAGATGCTCTACCAACTGAGC
12021  --TTGCTTGGTTT-----T--TACTTTCTTATA-------A--------------------
12020  --TTGCTTGGTTT-----T--TACTTTCTTATA-------------------A--------
12018  TAAAGCACTATATGAA-ACAT--TCTTCCAAA--AATAGTACCTATTACACTACTTACAC
12022  TAGCTCAGAAATGGTA-ACGTAGTCATTAAGATCAATACTAACCATAATCATAGCTAATT
         *                    *

12023  TAAGCCGGCAATCTACTAATGCGGGTGAAGGGACTTGAACCCCCACGCCGTTAAGCGCCA
12024  TAAGCCGGCAATCTACTAATGCGGGTGAAGGGACTTGAACCCCCACGCCGTTAAGCGCCA
12019  TAAGCCGGCAATCTACTAATGCGGGTGAAGGGACTTGAACCCCCACGCCGTTAAGCGCCA
12021  ----------AG------TAAAGCGGGTGACGAGAATCGAACTC------------------
12020  ----------AG------TAAAGCGGGTGACGAGAATCGAACTC------------------
12018  TATTAGATAGATAA--CAAATCGTCCT-----AAGTAAGCTTA--------CTTAGGACGA
12022  CATGATAACCATCGT-CACATCGTCCTTTAATATCTAATCCTAAATTAAGTTTGGCAGGA
                       *   **          *    * *

12023  GATCCTAAATCTGGTGCGTCTGCCAATTCCGCCACACCCGCATTTCTAAATGACCCGTAC
12024  GATCCTAAATCTGGTGCGTCTGCCAATTCCGCCACACCCGCATTTCTAAATGACCCGTAC
12019  GATCCTAAATCTGGTGCGTCTGCCAATTCCGCCACACCCGCATTTCTAAATGACCCGTAC
12021  -------------------------------GCGACAACAGC------------------
12020  -------------------------------GCGACAACAGC------------------
12018  TTTT----ATTTAGAACATAGGATAGTTTTTCCACTTTTAATCGTAA--------CCACTT
12022  GCTT----TCTCAAAAATTTTCATAAAACCTCCCTAATAAAATATAGAA-T-ATCCATAT
                                       *

12023  TGGGCTCGAACCAGTGACCCATTGATTAAAAGTCAATTGCTCTACCAACTGAGCTAACGA
12024  TGGGCTCGAACCAGTGACCCATTGATTAAAAGTCAATTGCTCTACCAACTGAGCTAACGA
12019  TGGGCTCGAACCAGTGACCCATTGATTAAAAGTCAATTGCTCTACCAACTGAGCTAACGA
12021  --------------T-----------TGGAAGGCTGTAGTTTTACCA-CTAAACTA----
12020  ------------------------TTGGAAGGCTGTAGTTTTACCA-CTAAACTA----
12018  GGTATCA------GTGACA----AATTCGGA--CAATTAAGATGTTAGCCAATCTTAAGG
12022  TATAACATAACAAATGACA----AATTCGGA--CAATTAAGATGCTAGCCAATCTTAAGG
                              *   *  *  *     *   * *  * **

12023  GTCTACGGTCCCGACGGGAATCGAACCCGCGATCTTCGCCGTGACAGGGCGACGTGATAA
12024  GTCTACGGTCCCGACGGGAATCGAACCCGCGATCTTCGCCGTGACAGGGCGACGTGATAA
```

```
12019  GTCTACGGTCCCGACGGGAATCGAACCCGCGATCTTCGCCGTGACAGGGCGACGTGATAA
12021  ----------------------------------------------CACC------------
12020  ----------------------------------------------CACC------------
12018  ATA-ATAATTCCAATAAAAA--------------------AAGGCTAACCAAAGTTAGTC
12022  ATA-ATAATTCCAATAAAAA--------------------AAGGCTAACCAAAGTTAGTC
                                                 *

12023  CCGCTACACTACGGGACCTATGGGAGTTAACGGGATCGAACCGCTGACCCTCTGCTTGTA
12024  CCGCTACACTACGGGACCTATGGGAGTTAACGGGATCGAACCGCTGACCCTCTGCTTGTA
12019  CCGCTACACTACGGGACCTATGGGAGTTAACGGGATCGAACCGCTGACCCTCTGCTTGTA
12021  -CGCT-----------TCTATGGGAGTTAACGGGATCGAACCGCTGACCCTCTGCTTGTA
12020  -CGCT-----------TCTATGGGAGTTAACGGGATCGAACCGCTGACCCTCTGCTTGTA
12018  TCCCTTTA--------TCTACTCCGCCAGTAGGACTCGAACCTACGACATCATGATTAAC
12022  TCCCTTTA--------TCTACTCCGCCAGTAGGACTCGAACCTACGACATCATGATTAAC
        * **             ***            **  *******   ***     ** **

12023  AGGCAGATGCT-CTCCCAGCTGAGCTAAACTCCCTTT--GCTAAGCGACTACCTTATCTC
12024  AGGCAGATGCT-CTCCCAGCTGAGCTAAACTCCCTTT--GCTAAGCGACTACCTTATCTC
12019  AGGCAGATGCT-CTCCCAGCTGAGCTAAACTCCCTTT--GCTAAGCGACTACCTTATCTC
12021  AGGCAGATGCT-CTCCCAGCTGAGCTAAACTCCCTTT--GCTAAGCGACTACCTTATCTC
12020  AGGCAGATGCT-CTCCCAGCTGAGCTAAACTCCCTTT--GCTAAGCGACTACCTTATCTC
12018  AGTCATGCGCTACTACCAACTGAGCTATGGCGGATTATAGCTAAGCGACTACCTTATCTC
12022  AGTCATGCGCTACTACCAACTGAGCTATGGCGGATTATAGCTAAGCGACTACCTTATCTC
       ** **   *** ** *** ********       **   **********************

12023  ACAGGGGGCAACCCCCAACTACTTCCGGCGTTCTAGGGCTTAACTTCTGTGTTCGGCATG
12024  ACAGGGGGCAACCCCCAACTACTTCCGGCGTTCTAGGGCTTAACTTCTGTGTTCGGCATG
12019  ACAGGGGGCAACCCCCAACTACTTCCGGCGTTCTAGGGCTTAACTTCTGTGTTCGGCATG
12021  ACAGGGGGCAACCCCCAACTACTTCCGGCGTTCTAGGGCTTAACTTCTGTGTTCGGCATG
12020  ACAGGGGGCAACCCCCAACTACTTCCGGCGTTCTAGGGCTTAACTTCTGTGTTCGGCATG
12018  ACAGGGGGCAACCCCCAACTACTTCCGGCGTTCTAGGGCTTAACTTCTGTGTTCGGCATG
12022  ACAGGGGGCAACCCCCAACTACTTCCGGCGTTCTAGGGCTTAACTTCTGTGTTCGGCATG
       ************************************************************

12023  AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACTCAAAAT
12024  AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACTCAAAAT
12019  AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACTCAAAAT
12021  AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACTCAAAAT
12020  AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACTCAAAAT
12018  AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACTCAAAAT
12022  AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACTCAAAAT
       ************************************************************

12023  TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTTGGATAAGTC
12024  TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTTGGATAAGTC
12019  TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTTGGATAAGTC
12021  TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTTGGATAAGTC
12020  TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTTGGATAAGTC
12018  TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTTGGATAAGTC
12022  TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTTGGATAAGTC
       ************************************************************

12023  CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCACCACAATTACACTCCTAACCTATCT
12024  CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCACCACAATTACACTCCTAACCTATCT
12019  CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCACCACAATTACACTCCTAACCTATCT
12021  CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCACCACAATTACACTCCTAACCTATCT
12020  CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCACCACAATTACACTCCTAACCTATCT
12018  CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCACCACAATTACACTCCTAACCTATCT
12022  CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCACCACAATTACACTCCTAACCTATCT
       ************************************************************

12023  ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT
12024  ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT
12019  ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT
12021  ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT
12020  ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT
12018  ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT
12022  ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT
```

```
                ************************************************************

12023           GGGCTTCGCACTTAGATGCTTTCAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG
12024           GGGCTTCGCACTTAGATGCTTTCAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG
12019           GGGCTTCGCACTTAGATGCTTTCAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG
12021           GGGCTTCGCACTTAGATGCTTTCAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG
12020           GGGCTTCGCACTTAGATGCTTTCAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG
12018           GGGCTTCGCACTTAGATGCTTTCAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG
12022           GGGCTTCGCACTTAGATGCTTTCAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG
                ************************************************************

12023           CCTTTGGCAAGACAACTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA
12024           CCTTTGGCAAGACAACTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA
12019           CCTTTGGCAAGACAACTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA
12021           CCTTTGGCAAGACAACTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA
12020           CCTTTGGCAAGACAACTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA
12018           CCTTTGGCAAGACAACTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA
12022           CCTTTGGCAAGACAACTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA
                ************************************************************

12023           GCAGATCCTCTCAAATTTCCTACGCCCGCGACGGATAGGGACCGAACTGTCTCACGACGT
12024           GCAGATCCTCTCAAATTTCCTACGCCCGCGACGGATAGGGACCGAACTGTCTCACGACGT
12019           GCAGATCCTCTCAAATTTCCTACGCCCGCGACGGATAGGGACCGAACTGTCTCACGACGT
12021           GCAGATCCTCTCAAATTTCCTACGCCCGCGACGGATAGGGACCGAACTGTCTCACGACGT
12020           GCAGATCCTCTCAAATTTCCTACGCCCGCGACGGATAGGGACCGAACTGTCTCACGACGT
12018           GCAGATCCTCTCAAATTTCCTACGCCCGCGACGGATAGGGACCGAACTGTCTCACGACGT
12022           GCAGATCCTCTCAAATTTCCTACGCCCGCGACGGATAGGGACCGAACTGTCTCACGACGT
                ************************************************************

12023           TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCCTTGGGACCGACTA
12024           TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCCTTGGGACCGACTA
12019           TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCCTTGGGACCGACTA
12021           TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCCTTGGGACCGACTA
12020           TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCCTTGGGACCGACTA
12018           TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCCTTGGGACCGACTA
12022           TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCCTTGGGACCGACTA
                ************************************************************

12023           CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCCGTCGATGTGAACTC
12024           CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCCGTCGATGTGAACTC
12019           CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCCGTCGATGTGAACTC
12021           CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCCGTCGATGTGAACTC
12020           CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCCGTCGATGTGAACTC
12018           CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCCGTCGATGTGAACTC
12022           CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCCGTCGATGTGAACTC
                ************************************************************

12023           TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC
12024           TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC
12019           TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC
12021           TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC
12020           TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC
12018           TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC
12022           TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC
                ************************************************************

12023           CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG
12024           CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG
12019           CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG
12021           CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG
12020           CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG
12018           CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG
12022           CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG
                ************************************************************

12023           CAGTCAAGCTCCCTTATACCTTTACACTCTACGACTGATTTCCAACCAGTCTGAGGGAAC
12024           CAGTCAAGCTCCCTTATACCTTTACACTCTACGACTGATTTCCAACCAGTCTGAGGGAAC
12019           CAGTCAAGCTCCCTTATACCTTTACACTCTACGACTGATTTCCAACCAGTCTGAGGGAAC
```

```
12021  CAGTCAAGCTCCCTTATACCTTTACACTCTACGACTGATTTCCAACCAGTCTGAGGGAAC
12020  CAGTCAAGCTCCCTTATACCTTTACACTCTACGACTGATTTCCAACCAGTCTGAGGGAAC
12018  CAGTCAAGCTCCCTTATACCTTTACACTCTACGACTGATTTCCAACCAGTCTGAGGGAAC
12022  CAGTCAAGCTCCCTTATACCTTTACACTCTACGACTGATTTCCAACCAGTCTGAGGGAAC
       ************************************************************

12023  CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAACTGCCCGTCAGAC
12024  CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAACTGCCCGTCAGAC
12019  CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAACTGCCCGTCAGAC
12021  CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAACTGCCCGTCAGAC
12020  CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAACTGCCCGTCAGAC
12018  CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAACTGCCCGTCAGAC
12022  CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAACTGCCCGTCAGAC
       ************************************************************

12023  ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC
12024  ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC
12019  ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC
12021  ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC
12020  ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC
12018  ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC
12022  ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC
       ************************************************************

12023  CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC
12024  CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC
12019  CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC
12021  CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC
12020  CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC
12018  CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC
12022  CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC
       ************************************************************

12023  ATGTGGTACAGATACTCAATATCAAACTGCAGTAAAGCTCCATGGGGTCTTTCCGTCCTG
12024  ATGTGGTACAGATACTCAATATCAAACTGCAGTAAAGCTCCATGGGGTCTTTCCGTCCTG
12019  ATGTGGTACAGATACTCAATATCAAACTGCAGTAAAGCTCCATGGGGTCTTTCCGTCCTG
12021  ATGTGGTACAGATACTCAATATCAAACTGCAGTAAAGCTCCATGGGGTCTTTCCGTCCTG
12020  ATGTGGTACAGATACTCAATATCAAACTGCAGTAAAGCTCCATGGGGTCTTTCCGTCCTG
12018  ATGTGGTACAGATACTCAATATCAAACTGCAGTAAAGCTCCATGGGGTCTTTCCGTCCTG
12022  ATGTGGTACAGATACTCAATATCAAACTGCAGTAAAGCTCCATGGGGTCTTTCCGTCCTG
       ************************************************************

12023  TCGCGGGTAACCTGCATCTTCACAGGTACTAAAATTTCACCGAGTCTCTCGTTGAGACAG
12024  TCGCGGGTAACCTGCATCTTCACAGGTACTAAAATTTCACCGAGTCTCTCGTTGAGACAG
12019  TCGCGGGTAACCTGCATCTTCACAGGTACTAAAATTTCACCGAGTCTCTCGTTGAGACAG
12021  TCGCGGGTAACCTGCATCTTCACAGGTACTAAAATTTCACCGAGTCTCTCGTTGAGACAG
12020  TCGCGGGTAACCTGCATCTTCACAGGTACTAAAATTTCACCGAGTCTCTCGTTGAGACAG
12018  TCGCGGGTAACCTGCATCTTCACAGGTACTAAAATTTCACCGAGTCTCTCGTTGAGACAG
12022  TCGCGGGTAACCTGCATCTTCACAGGTACTAAAATTTCACCGAGTCTCTCGTTGAGACAG
       ************************************************************

12023  TGCCCAAATCATTACGCCTTTCGTGCGGGTCGGAACTTACCCGACAAGGAATTTCGCTAC
12024  TGCCCAAATCATTACGCCTTTCGTGCGGGTCGGAACTTACCCGACAAGGAATTTCGCTAC
12019  TGCCCAAATCATTACGCCTTTCGTGCGGGTCGGAACTTACCCGACAAGGAATTTCGCTAC
12021  TGCCCAAATCATTACGCCTTTCGTGCGGGTCGGAACTTACCCGACAAGGAATTTCGCTAC
12020  TGCCCAAATCATTACGCCTTTCGTGCGGGTCGGAACTTACCCGACAAGGAATTTCGCTAC
12018  TGCCCAAATCATTACGCCTTTCGTGCGGGTCGGAACTTACCCGACAAGGAATTTCGCTAC
12022  TGCCCAAATCATTACGCCTTTCGTGCGGGTCGGAACTTACCCGACAAGGAATTTCGCTAC
       ************************************************************

12023  CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA
12024  CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA
12019  CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA
12021  CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA
12020  CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA
12018  CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA
12022  CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA
       ************************************************************
```

```
12023  CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCCTATACATCATCT
12024  CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCCTATACATCATCT
12019  CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCCTATACATCATCT
12021  CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCCTATACATCATCT
12020  CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCCTATACATCATCT
12018  CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCCTATACATCATCT
12022  CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCCTATACATCATCT
       ************************************************************

12023  TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCTATTCACTGCGGC
12024  TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCTATTCACTGCGGC
12019  TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCTATTCACTGCGGC
12021  TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCTATTCACTGCGGC
12020  TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCTATTCACTGCGGC
12018  TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCTATTCACTGCGGC
12022  TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCTATTCACTGCGGC
       ************************************************************

12023  TGATCTAAAATCAGCGCCCCTTCTCCCGAAGTTACGGGGCCATTTTGCCGAGTTCCTTAA
12024  TGATCTAAAATCAGCGCCCCTTCTCCCGAAGTTACGGGGCCATTTTGCCGAGTTCCTTAA
12019  TGATCTAAAATCAGCGCCCCTTCTCCCGAAGTTACGGGGCCATTTTGCCGAGTTCCTTAA
12021  TGATCTAAAATCAGCGCCCCTTCTCCCGAAGTTACGGGGCCATTTTGCCGAGTTCCTTAA
12020  TGATCTAAAATCAGCGCCCCTTCTCCCGAAGTTACGGGGCCATTTTGCCGAGTTCCTTAA
12018  TGATCTAAAATCAGCGCCCCTTCTCCCGAAGTTACGGGGCCATTTTGCCGAGTTCCTTAA
12022  TGATCTAAAATCAGCGCCCCTTCTCCCGAAGTTACGGGGCCATTTTGCCGAGTTCCTTAA
       ************************************************************

12023  CGAGAGTTCTCTCGCTCACCTGAGGCTACTCGCCTCGACTACCTGTGTCGGTTTGCGGTA
12024  CGAGAGTTCTCTCGCTCACCTGAGGCTACTCGCCTCGACTACCTGTGTCGGTTTGCGGTA
12019  CGAGAGTTCTCTCGCTCACCTGAGGCTACTCGCCTCGACTACCTGTGTCGGTTTGCGGTA
12021  CGAGAGTTCTCTCGCTCACCTGAGGCTACTCGCCTCGACTACCTGTGTCGGTTTGCGGTA
12020  CGAGAGTTCTCTCGCTCACMTGAGGCTACTCGCCTCGACTACCTGTGTCGGTTTGCGGTA
12018  CGAGAGTTCTCTCGCTCACCTGAGGCTACTCGCCTCGACTACCTGTGTCGGTTTGCGGTA
12022  CGAGAGTTCTCTCGCTCACCTGAGGCTACTCGCCTCGACTACCTGTGTCGGTTTGCGGTA
       ******************* ****************************************

12023  CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTAACTTCGC
12024  CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTAACTTCGC
12019  CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTAACTTCGC
12021  CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTAACTTCGC
12020  CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTAACTTCGC
12018  CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTAACTTCGC
12022  CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTAACTTCGC
       ************************************************************

12023  TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC
12024  TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC
12019  TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC
12021  TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC
12020  TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC
12018  TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC
12022  TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC
       ************************************************************

12023  ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC
12024  ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC
12019  ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC
12021  ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC
12020  ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC
12018  ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC
12022  ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC
       ************************************************************

12023  TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC
12024  TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC
12019  TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC
12021  TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC
```

```
12020  TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC
12018  TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC
12022  TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC
       ************************************************************

12023  GGTCTCTCCTTAGGTCCCGACTAACCCAGGGCGGACGAGCCTTCCCCTGGAAACCTTAGT
12024  GGTCTCTCCTTAGGTCCCGACTAACCCAGGGCGGACGAGCCTTCCCCTGGAAACCTTAGT
12019  GGTCTCTCCTTAGGTCCCGACTAACCCAGGGCGGACGAGCCTTCCCCTGGAAACCTTAGT
12021  GGTCTCTCCTTAGGTCCCGACTAACCCAGGGCGGACGAGCCTTCCCCTGGAAACCTTAGT
12020  GGTCTCTCCTTAGGTCCCGACTAACCCAGGGCGGACGAGCCTTCCCCTGGAAACCTTAGT
12018  GGTCTCTCCTTAGGTCCCGACTAACCCAGGGCGGACGAGCCTTCCCCTGGAAACCTTAGT
12022  GGTCTCTCCTTAGGTCCCGACTAACCCAGGGCGGACGAGCCTTCCCCTGGAAACCTTAGT
       ************************************************************

12023  CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA
12024  CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA
12019  CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA
12021  CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA
12020  CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA
12018  CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA
12022  CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA
       ************************************************************

12023  TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTCACACATAGAACGCTCTCCTACCATGA
12024  TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTCACACATAGAACGCTCTCCTACCATGA
12019  TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTCACACATAGAACGCTCTCCTACCATGA
12021  TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTCACACATAGAACGCTCTCCTACCATGA
12020  TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTCACACATAGAACGCTCTCCTACCATGA
12018  TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTCACACATAGAACGCTCTCCTACCATGA
12022  TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTCACACATAGAACGCTCTCCTACCATGA
       ************************************************************

12023  CACTTTTGTGTCATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA
12024  CACTTTTGTGTCATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA
12019  CACTTTTGTGTCATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA
12021  CACTTTTGTGTCATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA
12020  CACTTTTGTGTCATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA
12018  CACTTTTGTGTCATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA
12022  CACTTTTGTGTCATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA
       ************************************************************

12023  GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA
12024  GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA
12019  GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA
12021  GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA
12020  GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA
12018  GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA
12022  GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA
       ************************************************************

12023  CATCCTAGTTGTCTGTGCAACCCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT
12024  CATCCTAGTTGTCTGTGCAACCCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT
12019  CATCCTAGTTGTCTGTGCAACCCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT
12021  CATCCTAGTTGTCTGTGCAACCCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT
12020  CATCCTAGTTGTCTGTGCAACCCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT
12018  CATCCTAGTTGTCTGTGCAACCCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT
12022  CATCCTAGTTGTCTGTGCAACCCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT
       ************************************************************

12023  AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG
12024  AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG
12019  AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG
12021  AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG
12020  AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG
12018  AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG
12022  AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG
       ************************************************************
```

```
12023  CCGATTATATCTCGTTGGCATTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC
12024  CCGATTATATCTCGTTGGCATTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC
12019  CCGATTATATCTCGTTGGCATTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC
12021  CCGATTATATCTCGTTGGCATTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC
12020  CCGATTATATCTCGTTGGCATTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC
12018  CCGATTATATCTCGTTGGCATTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC
12022  CCGATTATATCTCGTTGGCATTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC
       ************************************************************

12023  TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT
12024  TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT
12019  TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT
12021  TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT
12020  TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT
12018  TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT
12022  TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT
       ************************************************************

12023  TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAATTTCTCCGCTACCCACAAGTCATC
12024  TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAATTTCTCCGCTACCCACAAGTCATC
12019  TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAATTTCTCCGCTACCCACAAGTCATC
12021  TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAATTTCTCCGCTACCCACAAGTCATC
12020  TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAATTTCTCCGCTACCCACAAGTCATC
12018  TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAATTTCTCCGCTACCCACAAGTCATC
12022  TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAATTTCTCCGCTACCCACAAGTCATC
       ************************************************************

12023  CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTTACCTCACCTTCAACC
12024  CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTTACCTCACCTTCAACC
12019  CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTTACCTCACCTTCAACC
12021  CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTTACCTCACCTTCAACC
12020  CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTTACCTCACCTTCAACC
12018  CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTTACCTCACCTTCAACC
12022  CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTTACCTCACCTTCAACC
       ************************************************************

12023  TGCTCATGGGTAGGTCACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA
12024  TGCTCATGGGTAGGTCACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA
12019  TGCTCATGGGTAGGTCACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA
12021  TGCTCATGGGTAGGTCACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA
12020  TGCTCATGGGTAGGTCACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA
12018  TGCTCATGGGTAGGTCACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA
12022  TGCTCATGGGTAGGTCACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA
       ************************************************************

12023  GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC
12024  GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC
12019  GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC
12021  GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC
12020  GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC
12018  GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC
12022  GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC
       ************************************************************

12023  CGGTTCATTCTACAAAAGGCACGCTCTCACCCATTAACGGGCTCGAACTTGTTGTAGGCA
12024  CGGTTCATTCTACAAAAGGCACGCTCTCACCCATTAACGGGCTCGAACTTGTTGTAGGCA
12019  CGGTTCATTCTACAAAAGGCACGCTCTCACCCATTAACGGGCTCGAACTTGTTGTAGGCA
12021  CGGTTCATTCTACAAAAGGCACGCTCTCACCCATTAACGGGCTCGAACTTGTTGTAGGCA
12020  CGGTTCATTCTACAAAAGGCACGCTCTCACCCATTAACGGGCTCGAACTTGTTGTAGGCA
12018  CGGTTCATTCTACAAAAGGCACGCTCTCACCCATTAACGGGCTCGAACTTGTTGTAGGCA
12022  CGGTTCATTCTACAAAAGGCACGCTCTCACCCATTAACGGGCTCGAACTTGTTGTAGGCA
       ************************************************************

12023  CACGGTTTCAGGTTCTATTTCACTCCCCTCCCGGGGTGCTTTTCACCTTTCCCTCACGGT
12024  CACGGTTTCAGGTTCTATTTCACTCCCCTCCCGGGGTGCTTTTCACCTTTCCCTCACGGT
12019  CACGGTTTCAGGTTCTATTTCACTCCCCTCCCGGGGTGCTTTTCACCTTTCCCTCACGGT
12021  CACGGTTTCAGGTTCTATTTCACTCCCCTCCCGGGGTGCTTTTCACCTTTCCCTCACGGT
12020  CACGGTTTCAGGTTCTATTTCACTCCCCTCCCGGGGTGCTTTTCACCTTTCCCTCACGGT
```

```
12018  CACGGTTTCAGGTTCTATTTCACTCCCCTCCCGGGGTGCTTTTCACCTTTCCCTCACGGT
12022  CACGGTTTCAGGTTCTATTTCACTCCCCTCCCGGGGTGCTTTTCACCTTTCCCTCACGGT
       ************************************************************

12023  ACTGGTTCACTATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCCAGATTC
12024  ACTGGTTCACTATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCCAGATTC
12019  ACTGGTTCACTATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCCAGATTC
12021  ACTGGTTCACTATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCCAGATTC
12020  ACTGGTTCACTATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCCAGATTC
12018  ACTGGTTCACTATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCCAGATTC
12022  ACTGGTTCACTATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCCAGATTC
       ************************************************************

12023  CGACGAGATTTCGCGTGTCTCGCCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT
12024  CGACGAGATTTCGCGTGTCTCGCCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT
12019  CGACGAGATTTCGCGTGTCTCGCCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT
12021  CGACGAGATTTCGCGTGTCTCGCCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT
12020  CGACGAGATTTCGCGTGTCTCGCCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT
12018  CGACGAGATTTCGCGTGTCTCGCCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT
12022  CGACGAGATTTCGCGTGTCTCGCCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT
       ************************************************************

12023  AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA
12024  AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA
12019  AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA
12021  AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA
12020  AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA
12018  AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA
12022  AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA
       ************************************************************

12023  ATCCTATATCGCAGTCCTACAACCCCGAAGTGTAAACACTTCGGTTTGCCCTCCTGCCGT
12024  ATCCTATATCGCAGTCCTACAACCCCGAAGTGTAAACACTTCGGTTTGCCCTCCTGCCGT
12019  ATCCTATATCGCAGTCCTACAACCCCGAAGTGTAAACACTTCGGTTTGCCCTCCTGCCGT
12021  ATCCTATATCGCAGTCCTACAACCCCGAAGTGTAAACACTTCGGTTTGCCCTCCTGCCGT
12020  ATCCTATATCGCAGTCCTACAACCCCGAAGTGTAAACACTTCGGTTTGCCCTCCTGCCGT
12018  ATCCTATATCGCAGTCCTACAACCCCGAAGTGTAAACACTTCGGTTTGCCCTCCTGCCGT
12022  ATCCTATATCGCAGTCCTACAACCCCGAAGTGTAAACACTTCGGTTTGCCCTCCTGCCGT
       ************************************************************

12023  TTCGCTCGCCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCCTGCAGCTACTTAGATGT
12024  TTCGCTCGCCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCCTGCAGCTACTTAGATGT
12019  TTCGCTCGCCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCCTGCAGCTACTTAGATGT
12021  TTCGCTCGCCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCCTGCAGCTACTTAGATGT
12020  TTCGCTCGCCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCCTGCAGCTACTTAGATGT
12018  TTCGCTCGCCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCCTGCAGCTACTTAGATGT
12022  TTCGCTCGCCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCCTGCAGCTACTTAGATGT
       ************************************************************

12023  TTCAGTTCACTGCGTCTTCCTTCTCATATCCTTAACAGATATGGATACTAGTCATTAACT
12024  TTCAGTTCACTGCGTCTTCCTTCTCATATCCTTAACAGATATGGATACTAGTCATTAACT
12019  TTCAGTTCACTGCGTCTTCCTTCTCATATCCTTAACAGATATGGATACTAGTCATTAACT
12021  TTCAGTTCACTGCGTCTTCCTTCTCATATCCTTAACAGATATGGATACTAGTCATTAACT
12020  TTCAGTTCACTGCGTCTTCCTTCTCATATCCTTAACAGATATGGATACTAGTCATTAACT
12018  TTCAGTTCACTGCGTCTTCCTTCTCATATCCTTAACAGATATGGATACTAGTCATTAACT
12022  TTCAGTTCACTGCGTCTTCCTTCTCATATCCTTAACAGATATGGATACTAGTCATTAACT
       ************************************************************

12023  AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAAGCAT
12024  AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAAGCAT
12019  AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAAGCAT
12021  AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAAGCAT
12020  AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAAGCAT
12018  AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAAGCAT
12022  AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAAGCAT
       ************************************************************

12023  TTCGTCGTTAGTCACGTCCTTCTTCGGCTTCTAGTGCCAAGGCATCCACCGTGCGCCCTT
```

```
12024  TTCGTCGTTAGTCACGTCCTTCTTCGGCTTCTAGTGCCAAGGCATCCACCGTGCGCCCTT
12019  TTCGTCGTTAGTCACGTCCTTCTTCGGCTTCTAGTGCCAAGGCATCCACCGTGCGCCCTT
12021  TTCGTCGTTAGTCACGTCCTTCTTCGGCTTCTAGTGCCAAGGCATCCACCGTGCGCCCTT
12020  TTCGTCGTTAGTCACGTCCTTCTTCGGCTTCTAGTGCCAAGGCATCCACCGTGCGCCCTT
12018  TTCGTCGTTAGTCACGTCCTTCTTCGGCTTCTAGTGCCAAGGCATCCACCGTGCGCCCTT
12022  TTCGTCGTTAGTCACGTCCTTCTTCGGCTTCTAGTGCCAAGGCATCCACCGTGCGCCCTT
       ************************************************************

12023  ATTAACTTAACCTTATTAACCTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTCACA
12024  ATTAACTTAACCTTATTAACCTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTCACA
12019  ATTAACTTAACCTTATTAACCTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTCACA
12021  ATTAACTTAACCTTATTAACCTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTCACA
12020  ATTAACTTAACCTTATTAACCTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTCACA
12018  ATTAACTTAACCTTATTAACCTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTCACA
12022  ATTAACTTAACCTTATTAACCTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTCACA
       ************************************************************

12023  GCGTTTTCGGTTTATTTTCTTGTTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA
12024  GCGTTTTCGGTTTATTTTCTTGTTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA
12019  GCGTTTTCGGTTTATTTTCTTGTTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA
12021  GCGTTTTCGGTTTATTTTCTTGTTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA
12020  GCGTTTTCGGTTTATTTTCTTGTTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA
12018  GCGTTTTCGGTTTATTTTCTTGTTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA
12022  GCGTTTTCGGTTTATTTTCTTGTTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA
       ************************************************************

12023  TATAGATATTCAATTTTCAATGAACAATTTGAACCTTTCGATTCAATGGAGCCTAGCGGG
12024  TATAGATATTCAATTTTCAATGAACAATTTGAACCTTTCGATTCAATGGAGCCTAGCGGG
12019  TATAGATATTCAATTTTCAATGAACAATTTGAACCTTTCGATTCAATGGAGCCTAGCGGG
12021  TATAGATATTCAATTTTCAATGAACAATTTGAACCTTTCGATTCAATGGAGCCTAGCGGG
12020  TATAGATATTCAATTTTCAATGAACAATTTGAACCTTTCGATTCAATGGAGCCTAGCGGG
12018  TATAGATATTCAATTTTCAATGAACAATTTGAACCTTTCGATTCAATGGAGCCTAGCGGG
12022  TATAGATATTCAATTTTCAATGAACAATTTGAACCTTTCGATTCAATGGAGCCTAGCGGG
       ************************************************************

12023  ATCGAACCGCTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAGGCCCCA
12024  ATCGAACCGCTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAGGCCCCA
12019  ATCGAACCGCTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAGGCCCCA
12021  ATCGAACCGCTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAGGCCCCA
12020  ATCGAACCGCTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAGGCCCCA
12018  ATCGAACCGCTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAGGCCCCA
12022  ATCGAACCGCTGACCTCCTGCGTGCAAAGCAGGCGCTCTCCCAGCTGAGCTAAGGCCCCA
       ************************************************************

12023  CAAGACCTCTCAAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG
12024  CAAGACCTCTCAAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG
12019  CAAGACCTCTCAAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG
12021  CAAGACCTCTCAAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG
12020  CAAGACCTCTCAAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG
12018  CAAGACCTCTCAAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG
12022  CAAGACCTCTCAAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG
       ************************************************************

12023  GTGATCCAGCCGCACCTTCCGATACGGCTACCTTGTTACGACTTCACCCCAATCATCTAT
12024  GTGATCCAGCCGCACCTTCCGATACGGCTACCTTGTTACGACTTCACCCCAATCATCTAT
12019  GTGATCCAGCCGCACCTTCCGATACGGCTACCTTGTTACGACTTCACCCCAATCATCTAT
12021  GTGATCCAGCCGCACCTTCCGATACGGCTACCTTGTTACGACTTCACCCCAATCATCTAT
12020  GTGATCCAGCCGCACCTTCCGATACGGCTACCTTGTTACGACTTCACCCCAATCATCTAT
12018  GTGATCCAGCCGCACCTTCCGATACGGCTACCTTGTTACGACTTCACCCCAATCATCTAT
12022  GTGATCCAGCCGCACCTTCCGATACGGCTACCTTGTTACGACTTCACCCCAATCATCTAT
       ************************************************************

12023  CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC
12024  CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC
12019  CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC
12021  CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC
12020  CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC
12018  CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC
```

```
12022  CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC
       ************************************************************

12023  TCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCGGCGTGCTGAT
12024  TCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCGGCGTGCTGAT
12019  TCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCGGCGTGCTGAT
12021  TCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCGGCGTGCTGAT
12020  TCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCGGCGTGCTGAT
12018  TCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCGGCGTGCTGAT
12022  TCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCGGCGTGCTGAT
       ************************************************************

12023  CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACTGA
12024  CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACTGA
12019  CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACTGA
12021  CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACTGA
12020  CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACTGA
12018  CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACTGA
12022  CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACTGA
       ************************************************************

12023  GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG
12024  GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG
12019  GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG
12021  GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG
12020  GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG
12018  GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG
12022  GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG
       ************************************************************

12023  TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC
12024  TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC
12019  TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC
12021  TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC
12020  TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC
12018  TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC
12022  TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC
       ************************************************************

12023  TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAATA
12024  TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAATA
12019  TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAATA
12021  TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAATA
12020  TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAATA
12018  TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAATA
12022  TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAATA
       ************************************************************

12023  GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
12024  GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
12019  GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
12021  GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
12020  GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
12018  GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
12022  GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
       ************************************************************

12023  TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT
12024  TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT
12019  TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT
12021  TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT
12020  TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT
12018  TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT
12022  TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT
       ************************************************************

12023  GTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTG
12024  GTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTG
```

```
12019  GTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTG
12021  GTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTG
12020  GTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTG
12018  GTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTG
12022  GTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTG
       ************************************************************

12023  TGCGGGCCCCCGTCAATTCCTTTGAGTTTCAACCTTGCGGTCGTACTCCCCAGGCGGAGT
12024  TGCGGGCCCCCGTCAATTCCTTTGAGTTTCAACCTTGCGGTCGTACTCCCCAGGCGGAGT
12019  TGCGGGCCCCCGTCAATTCCTTTGAGTTTCAACCTTGCGGTCGTACTCCCCAGGCGGAGT
12021  TGCGGGCCCCCGTCAATTCCTTTGAGTTTCAACCTTGCGGTCGTACTCCCCAGGCGGAGT
12020  TGCGGGCCCCCGTCAATTCCTTTGAGTTTCAACCTTGCGGTCGTACTCCCCAGGCGGAGT
12018  TGCGGGCCCCCGTCAATTCCTTTGAGTTTCAACCTTGCGGTCGTACTCCCCAGGCGGAGT
12022  TGCGGGCCCCCGTCAATTCCTTTGAGTTTCAACCTTGCGGTCGTACTCCCCAGGCGGAGT
       ************************************************************

12023  GCTTAATGCGTTAGCTGCGGCACTAAGCCCCGGAAAGGGCCTAACACCTAGCACTCATCG
12024  GCTTAATGCGTTAGCTGCGGCACTAAGCCCCGGAAAGGGCCTAACACCTAGCACTCATCG
12019  GCTTAATGCGTTAGCTGCGGCACTAAGCCCCGGAAAGGGCCTAACACCTAGCACTCATCG
12021  GCTTAATGCGTTAGCTGCGGCACTAAGCCCCGGAAAGGGCCTAACACCTAGCACTCATCG
12020  GCTTAATGCGTTAGCTGCGGCACTAAGCCCCGGAAAGGGCCTAACACCTAGCACTCATCG
12018  GCTTAATGCGTTAGCTGCGGCACTAAGCCCCGGAAAGGGCCTAACACCTAGCACTCATCG
12022  GCTTAATGCGTTAGCTGCGGCACTAAGCCCCGGAAAGGGCCTAACACCTAGCACTCATCG
       ************************************************************

12023  TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA
12024  TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA
12019  TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA
12021  TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA
12020  TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA
12018  TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA
12022  TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA
       ************************************************************

12023  GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCCTCCATATATCTACGCA
12024  GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCCTCCATATATCTACGCA
12019  GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCCTCCATATATCTACGCA
12021  GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCCTCCATATATCTACGCA
12020  GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCCTCCATATATCTACGCA
12018  GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCCTCCATATATCTACGCA
12022  GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCCTCCATATATCTACGCA
       ************************************************************

12023  TTTCACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA
12024  TTTCACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA
12019  TTTCACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA
12021  TTTCACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA
12020  TTTCACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA
12018  TTTCACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA
12022  TTTCACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA
       ************************************************************

12023  AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG
12024  AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG
12019  AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG
12021  AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG
12020  AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG
12018  AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG
12022  AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG
       ************************************************************

12023  CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA
12024  CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA
12019  CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA
12021  CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA
12020  CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA
12018  CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA
12022  CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA
```

```
                ************************************************************

12023           CGTAGTTAGCCGTCCCTTTCTGGTTAGTTACCGTCACTTGGTAGATTTTCCACTCCTACC
12024           CGTAGTTAGCCGTCCCTTTCTGGTTAGTTACCGTCACTTGGTAGATTTTCCACTCCTACC
12019           CGTAGTTAGCCGTCCCTTTCTGGTTAGTTACCGTCACTTGGTAGATTTTCCACTCCTACC
12021           CGTAGTTAGCCGTCCCTTTCTGGTTAGTTACCGTCACTTGGTAGATTTTCCACTCCTACC
12020           CGTAGTTAGCCGTCCCTTTCTGGTTAGTTACCGTCACTTGGTAGATTTTCCACTCCTACC
12018           CGTAGTTAGCCGTCCCTTTCTGGTTAGTTACCGTCACTTGGTAGATTTTCCACTCCTACC
12022           CGTAGTTAGCCGTCCCTTTCTGGTTAGTTACCGTCACTTGGTAGATTTTCCACTCCTACC
                ************************************************************

12023           AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT
12024           AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT
12019           AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT
12021           AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT
12020           AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT
12018           AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT
12022           AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT
                ************************************************************

12023           TGCTCGGTCAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC
12024           TGCTCGGTCAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC
12019           TGCTCGGTCAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC
12021           TGCTCGGTCAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC
12020           TGCTCGGTCAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC
12018           TGCTCGGTCAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC
12022           TGCTCGGTCAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC
                ************************************************************

12023           TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCCTCTCAGGTCGGCTATGTATCGTCG
12024           TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCCTCTCAGGTCGGCTATGTATCGTCG
12019           TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCCTCTCAGGTCGGCTATGTATCGTCG
12021           TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCCTCTCAGGTCGGCTATGTATCGTCG
12020           TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCCTCTCAGGTCGGCTATGTATCGTCG
12018           TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCCTCTCAGGTCGGCTATGTATCGTCG
12022           TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCCTCTCAGGTCGGCTATGTATCGTCG
                ************************************************************

12023           CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATACAACGCAGGTCCATCTCACAGTGAA
12024           CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATACAACGCAGGTCCATCTCACAGTGAA
12019           CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATACAACGCAGGTCCATCTCACAGTGAA
12021           CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATACAACGCAGGTCCATCTCACAGTGAA
12020           CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATACAACGCAGGTCCATCTCACAGTGAA
12018           CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATACAACGCAGGTCCATCTCACAGTGAA
12022           CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATACAACGCAGGTCCATCTCACAGTGAA
                ************************************************************

12023           GCAATTGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC
12024           GCAATTGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC
12019           GCAATTGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC
12021           GCAATTGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC
12020           GCAATTGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC
12018           GCAATTGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC
12022           GCAATTGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC
                ************************************************************

12023           GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTCGC
12024           GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTCGC
12019           GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTCGC
12021           GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTCGC
12020           GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTCGC
12018           GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTCGC
12022           GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTCGC
                ************************************************************

12023           AACTCATCAGTCTAGTGTAAACACCAAACCTCAGCGTTCTACTTGCATGTATTAGGCACG
12024           AACTCATCAGTCTAGTGTAAACACCAAACCTCAGCGTTCTACTTGCATGTATTAGGCACG
12019           AACTCATCAGTCTAGTGTAAACACCAAACCTCAGCGTTCTACTTGCATGTATTAGGCACG
```

```
12021  AACTCATCAGTCTAGTGTAAACACCAAACCTCAGCGTTCTACTTGCATGTATTAGGCACG
12020  AACTCATCAGTCTAGTGTAAACACCAAACCTCAGCGTTCTACTTGCATGTATTAGGCACG
12018  AACTCATCAGTCTAGTGTAAACACCAAACCTCAGCGTTCTACTTGCATGTATTAGGCACG
12022  AACTCATCAGTCTAGTGTAAACACCAAACCTCAGCGTTCTACTTGCATGTATTAGGCACG
       ************************************************************

12023  CCGCCAGCGTTCGTCCTGAGCCAGGATCAAACTCTCATTAAAAGTTTGAGCTTTGCTCTT
12024  CCGCCAGCGTTCGTCCTGAGCCAGGATCAAACTCTCATTAAAAGTTTGAGCTTTGCTCTT
12019  CCGCCAGCGTTCGTCCTGAGCCAGGATCAAACTCTCATTAAAAGTTTGAGCTTTGCTCTT
12021  CCGCCAGCGTTCGTCCTGAGCCAGGATCAAACTCTCATTAAAAGTTTGAGCTTTGCTCTT
12020  CCGCCAGCGTTCGTCCTGAGCCAGGATCAAACTCTCATTAAAAGTTTGAGCTTTGCTCTT
12018  CCGCCAGCGTTCGTCCTGAGCCAGGATCAAACTCTCATTAAAAGTTTGAGCTTTGCTCTT
12022  CCGCCAGCGTTCGTCCTGAGCCAGGATCAAACTCTCATTAAAAGTTTGAGCTTTGCTCTT
       ************************************************************

12023  TTCTGTCTCGCTGACAGATTTATTGTTTTTT-GTCATTGACGGATTTACAATGTAAATCC
12024  TTCTGTCTCGCTGACAGATTTATTGTTTTTTTGTCATTGACGGATTTACAATGTAAATCC
12019  TTCTGTCTCGCTGACAGATTTATTGTTTTTT-GTCATTGACGGATTTACAATGTAAATCC
12021  TTCTGTCTCGCTGACAGATTTATTGTTTTTTTGTCATTGACGGATTTACAATGTAAATCC
12020  TTCTGTCTCGCTGACAGATTTATTGTTTTTTTGTCATTGACGGATTTACAATGTAAATCC
12018  TTCTGTCTCGCTGACAGATTTATTGTTTTTT-GTCATTGACGGATTTACAATGTAAATCC
12022  TTCTGTCTCGCTGACAGATTTATTGTTTTTT-GTCATTGACGGATTTACAATGTAAATCC
       ****************************** ****************************

12023  ACCCTGCACATTCGTTCGTCTTGTTCAGTTTTCAAAGGTCTAATGATATATCATAAAAAT
12024  ACCCTGCACATTCGTTCGTCTTGTTCAGTTTTCAAAGGTCTTTGCCTCTCTTGAGACAAC
12019  ACCCTGCACATTCGTTCGTCTTGTTCAGTTTTCAAAGGTCTTTGCCTCTCTTGAGACAAC
12021  ACCCTGCACATTCGTTCGTCTTGTTCAGTTTTCAAAGGTCTTTGCCTCTCTTGAGACAAC
12020  ACCCTGCACATTCGTTCGTCTTGTTCAGTTTTCAAAGGTCTTTGCCTCTCTTGAGACAAC
12018  ACCCTGCACATTCGTTCATCTTGTTCAGTTTTCAAAGGTCTTTGCCTCTCTTGAGACAAC
12022  ACCCTGCACATTCGTTCGTCTTGTTCAGTTTTCAAAGGTCTTTGCCTCTCTTGAGACAAC
       ***************** ********************     * * *    * **

12023  ATATCCATCGGGAAGACAGGATTCGAACCTG-CGACACCTTGGTCCCAAACCAAGTACTC
12024  TTCTATATTCTAGCAAACTTATTCTGCTTTGTCAACTACTTTTTTTTAAGTTGTTAACTA
12019  TTCTATATTCTAGCAAACTTATTCTGCTTTGTCAACTACTTT-TTTTAAGTTGTTAACTA
12021  TTCTATATTCTAGCAAACTTATTCTGCTTTGTCAACTACTTT-TTTTAAGTTGTTAACTA
12020  TTCTATATTCTAGCAAACTTATTCTGCTTTGTCAACTACTTT-TTTTAAGTTGTTAACTA
12018  TTCTATATTCTAGCAAACTTATTCTGCTTTGTCAACTACTTT-TTTTAAGTTGTTTATAA
12022  TTCTATATTCTAGCAAACTTATTCTGCTTTGTCAACTACTTT-TTTTAAGTTGTTAACTA
        * *  **       *    ****     ** * **  ***  *   **       *

12023  TACCAAGCTG--A-GCTACT-TCCCGAAAAA---TATGCACC---CTAGAGGAGTCGAAC
12024  CGCGTTACTAGAA-GCTGCTCTCTCGAGACAACTTATTCATTATACTAAATATTTCTACT
12019  CGCGCTAATAGAA-ACTGCTCTCTCGAGACAACTTATTCATTATACTAAATATTTCTACT
12021  CGCGCTAATAGAA-ACTGCTCTCTCGAGACAACTTATTTAGTTTACTACATCATCTCTTA
12020  CGCGCTAATAGAA-ACTGCTCTCTCGAGACAACTTATTTAGTTTACTACATCATCTCTTA
12018  AATGATAATACAATATTAGGTTCGCTTAAGAACTCATTTAGTATACTATAATTTTTTATT
12022  CGCGCTAATAGAA-ACTGCTCTCTCGAGACAACTTATTTAGTTTACTACATCATCTCTTA
              *   *   *    ** *   * *    **  *     *** *

12023  CTCTAACCGCCTGATTCGTA-GTCAGGTACTCTATCC-------AGTTGA----GCTAAG
12024  TCCTGTCAATACTATTTTTGCATTTTTTCTTTTATTTTTAAA-AAGTTAATATTATTTAT
12019  TCCTGTCAATACTATTTTTGTA---TTTTATAAATTTAGTAT-AGACATAACTATTCCTC
12021  CTTTGTCAACTCTTTTTTCATACT-TTTTCTACATTTTCTGA-AAATGTAGATCAGGCTC
12020  CTTTGTCAACTCTTTTTTCATACT-TTTTCTACATTTTCTGA-AAATGTAGATAGAGCGC
12018  TGTTGTCAATAGGTTTTAAAAA-----------AATCTCAGAGAAAACCCTGAGATTTTT
12022  CTTTGTCAACTCTTTTTTCATACT-TTTTCTACATTTTCTGAAAAAAGTTTCCTGTTGGC
          *  *       **                 *

12023  GGTGCTAAAT--------ATTATATGCCGA-------GGACCGGAATC-------G---A
12024  AGTAACTAAC--------CTTCTATACTTGTTGA-ATGGATAGCATTT-------T---T
12019  TATATTCAATTAAGAGAAATTATATAACCACTATTGAGAAATGTAGTC-------T---A
12021  AA-GCTTAAC---GATTCTTTTTAAAATCATTA-----AATTTTAAAA-------C---A
12020  AAGAAAAAAAGAGGTCTCACCTCTTTTTATTTCTTAGTAACTACTACA-------A---A
12018  TAAATT--ATGTTACAAAGTT--AATTTCCTT-----TAGCTTCAATT---------AAA
12022  TAACACCAATAACATAGAGTTTAAAATTCCATAC--CTAAATTTATTTTATTAGTAAAAA
             *
```

```
12023  ACCGGTACGATGTTTACC-A--TCGCAGGATTTTAAGTCCTGTGCGTCTGC--CAGTTCC
12024  ACCGTTGTCATGTTCAT--A--TTTCATCTTCTTAATTCACAAATTTAAACTTCATCTTC
12019  GCGATTAAATTCTTTGCTCA--TCGAA-AATATCCAATAAATATAATAATGCATAAAACG
12021  AATTTCAGACATGTTGC------CAAA-GTTTTGATATTATTACTATAAT--ATAGTTTG
12020  TCTATTAGGATCGTTACCTT--CAGAATAACTTTCAACACCCTCTATAGT-TGCAATTGT
12018  CCTAGTTCGCCATCTTCACG-CTTGTAAAGGACATTTGTCGTATTATCTTCTGCATCT--
12022  AATAAAAGATGGGCTAGCCATCTTTTATAATATTTGTTTTTTATATTCTTCAGCTTCTTG
                                      *                 *

12023  G-CCA--------CCCCGGCCTCTAACAAGCGAACGACGGGGTTCGAA-CCCGCGACCCT
12024  A-TAAAAAATACCCTTCAAATTTTATCTAAATTTGAAGGGTATTTGAAATTTATAAAGTT
12019  C-CTGCTTACGAAATATAAACAAA-ATTGTTTGCAT--TTCGTAAACAAGCGTTACCTAT
12021  T-AGAGGAGAATAATATGGGCCAA-GAACCTATCAT--CGAATATCAAAATAT----CAA
12020  T-TTATGAACAGTTTTTCGCTCACTGTTACTCATAGGATCCATATGGTAAGGTTCATTAG
12018  --GTATAGATAAAGAAATCATGACCTAAAAGTTCCATTTGCAACAATGCTTCCTCAACAT
12022  GGGTGTAGATAAAACAAA-ATGACCAGGGGTAATCTCGTGCATTTGACGTTCTTGTCCGT

12023  CAC--CTTGGCAAGGTGATGTTCTACCACTGAACTACGT-TCGCACTAAAGACACTATT-
12024  CTT--TAAAAATATATGATGACTTATTTTTTATCTTCTTCTTGCATTTTTTCTTTGATTT
12019  TTA--ACAATATATGATGAGTGTTCCCGCTGAGAATAATTCTCAGCGGTAGACCAGAGCT
12021  TAA--AGTGTATGGGGAAAATGTTGCGGTTGAAGACA----TTAACCTTAAAATTTACCC
12020  TCT--CTAAAACACGCCTAGCTATTTTTTTAGAAAAA----TCAATTAAAGTTTCTGTAC
12018  CCA--TTGGTTTTAGATTAACATTCTTAGTACGTACAAT--T-CTTTGGCTTACTGCTTC
12022  CTTGCTCAATAGCTGGATTATACGGCTGGTGAACACGTT--GACGTTCACTCTCCGGATC
                                         *

12023  ------------------------------------------------------------
12024  CATCGTATGATAACGCTCTTGCTTTATCTTCA---TCATTTTCTGTCTCAGGCATTTTAC
12019  AGACTAAGAATCGATTGATTCCATCATCATAACACTCAACAAAATTGATAAAAATTATAC
12021  TGGT---GATTTTGTTTGTTTCATCGGTA-------CGAGTGGATCAGGTAAAACAACAT
12020  GATGCTCAACGTAGTCATGGACATTAATGGA----TACTGAAAAACTCTTAGAAAAGCGG
12018  TTCATCTGGCTCAGCC-----TCAAATTCTGTTGTGAAAAC---TTGACTTGCTGGAATC
12022  TGGTTCTGGAATAGCTGATAATAGACTCTTCGTATAAGGGTGGATTGGATTGTTATAAAC

12023  ------------------------------------------------------------
12024  CTGTCTCAAAAATCGATTTAATCTGAGCAGCATCAA-----GAGTCTCATATTTTAAGAG
12019  TAATTCA-ATAATTGCCATTGGGGCAGCATCGCCAC--GGCGTGGTTCTGT-TTTAAGAA
12021  TAATGCGTATGGTTAACCATATGTTAAAACCAACAA--ATGGTACTCTATTATTTAAGGG
12020  TCATGAAGATAATT----TTGTGCTAACAACTGCAACGATTTTAATACTTT-TCCATGAT
12018  TTTTCACGATATT--------TTTTCGCAATTTTA--------GTTTTATT-TTTACGA-
12022  ATCATCAGATGTTC---CAACTTCTAACAGTTTCCCCCAATGCATAACACC-GATACGAT

12023  ------------------------------------------------------------
12024  GGCTTCTGCAATTAATTTATGAGTATCACGGTTTTCGTTGATAATATCAGCTGCC--TTA
12019  TACGAGTGTATCCTCCG--TTACGTTCAGCATAACGAGGTGCGATGTCGTC-AAA--AAG
12021  AAAAGATATTTCTACTA--TTAACCCCATTGAATTAAGACGCAGAATTG---GAT--ATG
12020  AACCAATGATGCGCCCAGCTTCTGGCGTTTCTATTTGGAGATTTATTTGTC-GCT--TAC
12018  ---------ATTTGAC------GCTCAATTTT-----ATCAACAACTAAGTCAAT-----
12022  CTGAAATGTATTTTACC--ATAGACAAATCATGTGCGATAAACAAATAAGTCAATCCTTG

12023  ------------------------------------------------------------
12024  TTACGTGCTTCATTAAGAAGGTGACGAACTTCATCATCAATAAGTTGTGCAGTTTGAGCA
12019  TTTTTGAAGAGCTGTTGTTGATGTAT-ATTT-ATCAGAAGCTTCATCATAGTTTTCTGAT
12021  TTATCCAAAACATTGGTTTAATGCCTCATATGACCATTTACGAAAATATAGTTCT-TGTA
12020  TTGTCGTAG--TTTCTATTGTTGCAT-CTAAATCCATCTCATAGATGATATTTTC----A
12018  TGACCCATACATATCTTGTGAAACATCTTCTGCTCGTAAAGTAATAGAATCTATTAAG--
12022  TTCTCTTTGCAATTTTTGCATTAAATTAACAACTTGTGCTTGGATTGAAACATCTAAGGC

12023  ------------------------------------------------------------
12024  GAATATGATTTTTCAGGTGACATTTGACCA--GCCATCATTGCGTGGTTGCCTTCGTATT
12019  GCAATTTCATTACGTACATAAGCAGCAGCT--TGACGACGAGCATGTAAATCACCACGTT
12021  CCAAAATTATTGAAATGGTCAGAAGAAGCT--AAAAGA-GCTAAAGCAAGGGAACTTATT
12020  ACGTATTTAGTCACCTGAGCAGCTGCTACT--TCAATATTAGGAAGTAGGTCAATTTTTT
12018  ----ATTGTTACTTCAACTTTTGCGGTCTT---CTCTCTGTATACTTTGAGGTTGACTCT
12022  AGATATTGGTTCATCAGCAATGATAAATTTAGGCTCTACTGCTAAAGCACGTGCAATCCC
```

```
12023  ------------------------------------------------------------
12024  GAACTGGTCCAAGTTTCTCGCTCATACCATATTCAGTTACCATAGCGCGGGCCATAGCAG
12019  TACCTAGAGTAATCATTTTTTCAACTGTTTTACGGATTTCTTTAGCACGTGCTTCAGTAG
12021  AAATTAGTTGAATTACCCGAAGAA-TATTTGGATCGCTACCCTAGTGAGTTGTCTGGCGG
12020  CAATAGCTTCACTTGTAGTTACAACGTTTTTATCATCAATTTTTGGAAGT--TCTGGTTG
12018  AGTATCTAATTCTTGTGCTTC-----ATTAAAGTATTTTTCAACTTTAGAGAGTTTGGTC
12022  GATACGTTGTCGTTGTCCACCTGAAAATTCATGCGGATAACGTGTTAAATGATCTTTATT

12023  ------------------------------------------------------------
12024  TGGCTTGTTCGAAGTCATTTGAGGCACCTGTTGTCTGAGCGTTGAAAATAATTTCTTCCG
12019  TTACAATTGATTCGTTGATAAGAAGATCGGTTGTCAAATCA--CGAAGCATTGCCTTACG
12021  TCAGCAACAACGTATCGGTGTCATTCGCGCTCTTGCAGCAGACCAAGATATTATTTTAAT
12020  T----GATACGTCTTCTTTTTCAAGCGT-TTCATCAACCTCCTCTATATATTCTTCCACC
12018  TCAACATACTCA--CGAATAGCTTCTG----TTACTTCGATGTTTTCACCACGAAT-ACT
12022  TAACCCTACAAGATCTAATAGGGCCTGAACTTTACTATCACGATCTGATTTTGATTTAGC

12023  ------------------------------------------------------------
12024  CTACACGTCCTCCCATAAGACCTGCTAATTGCTCTTTCATATCATCTTTTGAAAGAAGCA
12019  TTGTGAGCTAGT-------GCGTCCTAGTTTACGGTAAGCCATTATGTCCTCCTATTTTA
12021  GGATGAGCCTTTT-----GGAGCTCTGGATCCTATTACTAGAGAAGGTATTCAAGACTTA
12020  ACATCTACGCTA-------GACGGTACATTCTTAATATTTTTTAACG--CTACCGATTCA
12018  GTATTTAATCAT----ATGAGTACCTCTTTCTTGCGTTGTTAACGCTTTCTATACTCTTA
12022  TAATTTATGTAT----ATCTAAACCT-TCTGCTACGATATCACGAATCTTCATACGGCCG

12023  ------------------------------------------------------------
12024  TTTGATCTTCTTTAGGT----AAAGCAATCATATATCCACCTGCACGACCACGTGGTACG
12019  TTTATCGTTTTTTAATC----CAAGACCTAGATCGGCAAGTTTGATTTTAACTTCTTCAA
12021  GTCAAGTCTCTTCAGG------AAGAAATGGG--GAAAACTATCATCTTAGTTACT-CAT
12020  TTAATATCAGTTACTT----------CGTCGGT-GATACCTTCTATTTCAACTTTTGCTG
12018  TTATAACC-GCTT-----TCATGAAAA----------------------------------
12022  TTTAAGCTAGCCTGAGGATCCTGAAAAATCATCTGAGCGTCTTTACGAAAACTATGTAAT

12023  ------------------------------------------------------------
12024  ATAGTAACTTTATGAACAACTCGCGCATTTGAAAGAATCAAACCGACAATTGTGTGCCCA
12019  GACTCTTACGTCCTAAGTTTCGGACTTTCATCATTTCAGGCTCAGTTTTTTCTG-TTAAA
12021  GA---T-ATGGATGAAGCCCTCAAGTT--AGCAACAAAAATT--ATTGTTATGG-ACAAT
12020  GC---TTTTTACCAAAGCCCAAAAAACCTTTTTTCTCACGTGATACAACTTTTATATGTG
12018  ------------------------------------------------------------
12022  GCTTTACCTTTCAGATGTGAGATCACTTCTCCATTAAAGGTAATTTCTCCATCAGAAATA

12023  ------------------------------------------------------------
12024  GCTTCATGGTAAGCAACCATAGCTCTTTCTCTTTCAGAAATAGTACGATCTTTTTTAGAA
12019  TCAAATACTGTATTAATTCCAGCACGTTTTAAACAGTTATATGAGCGCACTGACAAATCA
12021  GGTAAAATGGTCCAAGAAGGGACACCCAATGATCTCTTACATCATCCTGCTA--------
12020  CCCTCAATCGTGAAATGTTTAACTCTTGTAGTCCTTTTTCAATAGCTTCTTCTACAGTCG
12018  ------------------------------------------------------------
12022  TCATAAAGTTTTAAAATTGAACGTCCAACGGTTGTCTTTCCTGATCCAGATTCCCCAACT

12023  ------------------------------------------------------------
12024  GGACCAGCAATTACACGGTCTTCTGCTTCATCAATATCTGAAGCATCAATAACTTTTTTA
12019  AGTTCCTCAATTGTCCGGTCAAGCACTTTCTCATCGTTCACTTTCTCTGTTTCCTTCATT
12021  ------------------------------------------------------------
12020  CTCCTGTAAATAATACC------------------------------------------
12018  ------------------------------------------------------------
12022  AATCCAAACACTTCACCTTCATAAATGTCAAAACTAACATTATCAATTGCTCTCACTTCA

12023  ------------------------------------------------------------
12024  TTTCGTCGCGCAGCAACTAAAGCAGCTTCATTGAGAACATTCTCCAAATCAGCACCAACA
12019  ACTTCAGTTGCTTTAGCAACCTCTGTTAAATCAGTAAACAAGTTTAAGTGTTCAATTAAG
12021  ------------------------------------------------------------
12020  ------------------------------------------------------------
12018  ------------------------------------------------------------
12022  TTAGCTTTTCCTTTATTGAAGGTCAAAGAAACATTTTTGACTTCAACTAATTTTTTTCGA

12023  ------------------------------------------------------------
12024  AATCCTGGGGTTTGTTGAGCTACTACTTTTAAGTCAACATTATCTGCTAATGGTTTATTT...
12019  ACGCGAGCTGAAAGACCAAGAGCATCCTCAGGAATGA----------------------
12021  ------------------------------------------------------------
```

```
12020       ------------------------------------------------------------
12018       ------------------------------------------------------------
12022       TTTTCAGTCATTAGGCT--------------------------------------------
```

It will be understood that the invention has been described by way of example only and modifications may be made whilst remaining within the scope and spirit of the invention.

TABLE I – THEROETICAL MOLECULAR WEIGHTS FOR GBS PROTEINS

GBS #	expected mol. weight (dalton)		
	GST-fusion	His-fusion	Native
1	78425	53460	49720
2	40035	15070	11330
3	90305	65340	61600
4	43115	18150	14410
5	158835	133870	130130
6	39265	14300	10560
7	44985	20020	16280
8	56315	31350	27610
9	50265	25300	21560
10	96465	71500	67760
11	91515	66550	62810
11d	85905	60940	57200
12	64455	39490	35750
13	40475	15510	11770
14	33325	8360	4620
15	44765	19800	16060
16	73475	48510	44770
17	46745	21780	18040
18	54335	29370	25630
19	46085	21120	17380
20	47625	22660	18920
21	56535	31570	27830
21 long	66435	41470	37730
22	60055	35090	31350
23	60165	35200	31460
24	58405	33440	29700
25	50265	25300	21560
26	118245	93280	89540
28	63795	38830	35090
29	50595	25630	21890
30	44215	19250	15510
31	63795	38830	35090
31d	58735	33770	30030
32	40585	15620	11880
33	71495	46530	42790
34	69295	44330	40590
35	56535	31570	27830
36	59065	34100	30360
37	46965	22000	18260
38	61815	36850	33110
39	65225	40260	36520
41	75235	50270	46530
42	46745	21780	18040
43	58955	33990	30250
44	52355	27390	23650
45	43555	18590	14850
46	59835	34870	31130
47	84255	59290	55550
48	86455	61490	57750
48d	106695	81730	77990
49	59615	34650	30910
50	94155	69190	65450

51	47075	22110	18370
52	55435	30470	26730
53	110215	85250	81510
54	73365	48400	44660
55	36295	11330	7590
56	34865	9900	6160
57	51145	26180	22440
58	128805	103840	100100
59	99215	74250	70510
60	63575	38610	34870
61	68085	43120	39380
62	105485	80520	76780
63	64125	39160	35420
64	112745	87780	84040
65	72485	47520	43780
66	49715	24750	21010
67	120335	95370	91630
68	131225	106260	102520
68d	103065	78100	74360
69	53895	28930	25190
70	74465	49500	45760
70d	59725	34760	31020
71	56755	31790	28050
72	75565	50600	46860
73	72815	47850	44110
74	131225	106260	102520
74d	95475	70510	66770
75	114725	89760	86020
76	198875	173910	170170
77	78535	53570	49830
78	48835	23870	20130
79	58185	33220	29480
79d	50815	25850	22110
80	81835	56870	53130
81	89205	64240	60500
82	40475	15510	11770
83	62585	37620	33880
84	122645	97680	93940
85	70175	45210	41470
86	84035	59070	55330
87	44435	19470	15730
88	73365	48400	44660
89	143325	118360	114620
90	93495	68530	64790
91	88325	63360	59620
92	193595	168630	164890
93	95585	70620	66880
94	77435	52470	48730
95	60605	35640	31900
96	57195	32230	28490
97	138375	113410	109670
98	82055	57090	53350
99	60715	35750	32010
100	53015	28050	24310
101	59395	34430	30690
102	40695	15730	11990
103	56975	32010	28270

104	120005	95040	91300
105	179735	154770	151030
105dNterm	127265	102300	98560
105dCterm	81285	56320	52580
106	85795	60830	57090
107	89535	64570	60830
108	64565	39600	35860
109	75125	50160	46420
109d	70725	45760	42020
110	53895	28930	25190
111 / 190	60165	35200	31460
112	63905	38940	35200
113	59175	34210	30470
114	51915	26950	23210
115	98225	73260	69520
116	73475	48510	44770
117	47515	22550	18810
118	42235	17270	13530
119	109225	84260	80520
120	71385	46420	42680
121	65115	40150	36410
122	46855	21890	18150
123	68305	43340	39600
124	54115	29150	25410
125	57305	32340	28600
126	56865	31900	28160
127	80845	55880	52140
128	39925	14960	11220
129	43775	18810	15070
130	82275	57310	53570
130d	63245	38280	34540
131	89755	64790	61050
132	49055	24090	20350
133	54445	29480	25740
134	42015	17050	13310
135	65225	40260	36520
136	54885	29920	26180
137	63465	38500	34760
138	40145	15180	11440
139	38165	13200	9460
140	43445	18480	14740
141	49935	24970	21230
142	79745	54780	51040
143	33545	8580	4840
144	49165	24200	20460
145	63025	38060	34320
146	107025	82060	78320
147	156965	132000	128260
148	41905	16940	13200
149	62365	37400	33660
150	54665	29700	25960
151	50412	25447	21707
151L	50045	25080	21340
152	45535	20570	16830
153	46965	22000	18260
154	101525	76560	72820
155	62585	37620	33880

156	61265	36300	32560
157	74025	49060	45320
158	52025	27060	23320
159	41025	16060	12320
160	82825	57860	54120
161	95365	70400	66660
162	42015	17050	13310
163	69405	44440	40700
164	42345	17380	13640
165	43555	18590	14850
166	38055	13090	9350
167	50375	25410	21670
168	32555	7590	3850
169	43445	18480	14740
170	64015	39050	35310
170d	59945	34980	31240
171	49825	24860	21120
172	62365	37400	33660
173	96795	71830	68090
174	45095	20130	16390
175	59175	34210	30470
176	55435	30470	26730
177	66215	41250	37510
178	62365	37400	33660
179	58515	33550	29810
180	37615	12650	8910
181	63685	38720	34980
182	90085	65120	61380
182d	87225	62260	58520
183	57855	32890	29150
184	46415	21450	17710
185	40695	15730	11990
186	85685	60720	56980
187	56205	31240	27500
188	61595	36630	32890
189	60165	35200	31460
191	116705	91740	88000
192	69625	44660	40920
193	98005	73040	69300
194	49385	24420	20680
195	81065	56100	52360
195L	147615	122650	118910
195L N-term	91405	66440	62700
196	69515	44550	40810
197	99325	74360	70620
198	73805	48840	45100
199	158285	133320	129580
200	132325	107360	103620
201	74538	49573	45833
202	157295	132330	128590
203	61705	36740	33000
204	39705	14740	11000
205	55985	31020	27280
206	56645	31680	27940
207	44765	19800	16060
208	59725	34760	31020
209	62145	37180	33440

209d	56425	31460	27720
210	60935	35970	32230
210d	53675	28710	24970
211	64895	39930	36190
212	60825	35860	32120
213	45205	20240	16500
214	38935	13970	10230
215	45205	20240	16500
216	91515	66550	62810
217	36075	11110	7370
218	81065	56100	52360
219	56535	31570	27830
220	54555	29590	25850
220	50155	25190	21450
221	41465	16500	12760
222	47405	22440	18700
223	42895	17930	14190
224	45865	20900	17160
225	56645	31680	27940
226	44875	19910	16170
227	46195	21230	17490
228	46525	21560	17820
229	35855	10890	7150
230	51915	26950	23210
231	60935	35970	32230
231d	58735	33770	30030
232	41795	16830	13090
233	35635	10670	6930
234	43115	18150	14410
235	58295	33330	29590
235d	48395	23430	19690
236	46525	21560	17820
237	44215	19250	15510
238	59725	34760	31020
239	63905	38940	35200
240	51475	26510	22770
241	45095	20130	16390
242	43225	18260	14520
243	119455	94490	90750
244	48065	23100	19360
245	48615	23650	19910
246	49605	24640	20900
246d	45975	21010	17270
247	58955	33990	30250
248	92505	67540	63800
248d	70835	45870	42130
249	103835	78870	75130
250	136505	111540	107800
251	52135	27170	23430
252	51695	26730	22990
253	74245	49280	45540
254	59615	34650	30910
255	69075	44110	40370
256	47845	22880	19140
257	60495	35530	31790
258	67975	43010	39270
259	79415	54450	50710

260	48175	23210	19470
261	55765	30800	27060
262	75345	50380	46640
263	63465	38500	34760
264	47185	22220	18480
265	56315	31350	27610
266	51365	26400	22660
267	88655	63690	59950
268	50265	25300	21560
269	60495	35530	31790
270	59285	34320	30580
271	56315	31350	27610
272	118355	93390	89650
272d	98885	73920	70180
273	70945	45980	42240
274	56205	31240	27500
275	47515	22550	18810
276	147945	122980	119240
277	87005	62040	58300
277d	75675	50710	46970
278	52245	27280	23540
279	79415	54450	50710
280	88655	63690	59950
281	74465	49500	45760
281d	71495	46530	42790
282	44765	19800	16060
283		20240	16500
284	67645	42680	38940
285	57525	32560	28820
286	41355	16390	12650
287	61045	36080	32340
287d	57085	32120	28380
288	53675	28710	24970
288d	51035	26070	22330
289	65005	40040	36300
289 long	71825	46860	43120
290	47405	22440	18700
291	63795	38830	35090
292	103505	78540	74800
293	115935	90970	87230
293d N-term	73805	48840	45100
293d C-term	70835	45870	42130
294	75785	50820	47080
295	89425	64460	60720
296	60385	35420	31680
297	100205	75240	71600
298	54335	29370	25630
299	62255	37290	33550
300	130895	105930	102190
301	54885	29920	26180
302	80075	55110	51370
303	53235	28270	24530
304	75125	50160	46420
305	78645	53680	49940
306	67975	43010	39270
307	86675	61710	57970
308	59285	34320	30580

309	62695	37730	33990
310	58845	33880	30140
311	76445	51480	47740
312	64785	39820	36080
313	65995	41030	37290
314	52135	27170	23430
315	51695	26730	22990
316	41795	16830	13090
317	179295	154330	150590
317d N-term	115935	90970	87230
317d C-term	92160	67402	63360
318	70065	45100	41360
319	61925	36960	33220
320	57965	33000	29260
321	83705	58740	55000
322	76628	51663	47923
323	86345	61380	57640
324	86345	61380	57640
325	82605	57640	53900
326	91515	66550	62810
326L	172695	147730	143990
326L N-term	113955	88990	85250
327	279175	254210	250470
327d N-term	139915	114950	111210
327d C-term	167965	143000	139260
328	97602	72637	68897
329	113955	88990	85250
330	83595	58630	54890
331	60825	35860	32120
332	75675	50710	46970
333	63465	38500	34760
333d	57965	33000	29260
334	38275	13310	9570
335	43555	18590	14850
336	67645	42680	38940
337	75235	50270	46530
338	54995	30030	26290
339	76665	51700	47960
339d	72925	47960	44220
340	86565	61600	57860
341	38385	13420	9680
342	61595	36630	32890
343	60385	35420	31680
344	55875	30910	27170
345	40585	15620	11880
346	53895	28930	25190
347	55325	30360	26620
348	58405	33440	29700
349	98335	73370	69630
350	53895	28930	25190
351	82165	57200	53460
352	111315	86350	82610
352d	105485	80520	76780
353	55325	30360	26620
354	42345	17380	13640
355	52135	27170	23430
356	59065	34100	30360

357	40255	15290	11550
358	60495	35530	31790
359	78865	53900	50160
360	73695	48730	44990
361	109005	84040	80300
362	125945	100980	97240
362d N-term	63355	38390	34650
362d C-term	91295	66330	62590
363	53125	28160	24420
364	75015	50050	46310
365	102075	77110	73370
366	68415	43450	39710
367	76885	51920	48180
368	44765	19800	16060
369	142115	117150	113410
370	94595	69630	65890
371	65555	40590	36850
372	55105	30140	26400
373	50265	25300	21560
374	57525	32560	28820
375	66875	41910	38170
376	48065	23100	19360
377	73805	48840	45100
378	58955	33990	30250
379	68855	43890	40150
380	47405	22440	18700
381	66875	41910	38170
382	50815	25850	22110
383	57085	32120	28380
384	77985	53020	49280
385	75675	50710	46970
386	39485	14520	10780
387	54555	29590	25850
388	45645	20680	16940
389	43005	18040	14300
390	62255	37290	33550
391	54775	29810	26070
392	71385	46420	42680
393	55765	30800	27060
394	59725	34760	31020
395	72375	47410	43670
396	34865	9900	6160
397	113625	88660	84920
397d	100865	3740	72160
398	56755	31790	28050
399	55435	30470	26730
400	74135	49170	45430
401	59395	34430	30690
402	78095	53130	49390
403	64455	39490	35750
404	61595	36630	32890
405	45975	21010	17270
406	36955	11990	8250
407	82715	57750	54010
407d	71715	46750	43010
408	45315	20350	16610
409	70395	45430	41690

409d	59600	34842	30800
410	62475	37510	33770
411	41355	16390	12650
412	35965	11000	7260
413	59175	34210	30470
414	50375	25410	21670
415	46195	21230	17490
416	42455	17490	13750
417	77985	53020	49280
418	42125	17160	13420
419	47515	22550	18810
420	67755	42790	39050
421	62915	37950	34210
422	60165	35200	31460
423	74245	49280	45540
424	89975	65010	61270
424	77325	52360	48620
425	116045	91080	87340
426	83815	58850	55110
427	41135	16170	12430
428	55325	30360	26620
429	59175	34210	30470
430	53785	28820	25080
431	54005	29040	25300
432	65665	40700	36960
433	40915	15950	12210
434	44545	19580	15840
642	91845	66880	63140
643	78975	54010	50270
644	49605	24640	20900
645	59725	34760	31020
646	61595	36630	32890
647	55875	30910	27170
648	59835	34870	31130
649	76115	51150	47410
650	51475	26510	22770
651	53345	28380	24640
652	49715	24750	21010
653	44655	19690	15950
654	51255	26290	22550
655	65995	41030	37290
656	57525	32560	28820
657	62805	37840	34100
658	60165	35200	31460
659	60275	35310	31570
660	71495	46530	42790
661	60605	35640	31900
662	62695	37730	33990
663	89535	64570	60830
664	45315	20350	16610
665	41135	16170	12430
666	47075	22110	18370
667	53162	28197	24457
668	43555	18590	14850
669	48505	23540	19800
670	45315	20350	16610
671	36940	12182	8140

672	40130	15372	11330
673	41450	16692	12650
674	45300	20542	16500
675	55970	31212	27170
676	65650	40892	36850
677	54320	29562	25520
678	77750	52992	48950
679	60480	35722	31680
680	64440	39682	35640
681	93040	68282	64240
682	84790	60032	55990
683	15950	44655	19690
684	11880	40585	15620
685	16280	44985	20020
686	21340	50045	25080
687	9350	38055	13090
689	55105	3740	26400

TABLE II – PRIMERS USED TO AMPLIFY GBS*nnn* PROTEINS

Forward primers begin 5'-GGGGACAAGTTTGTACAAAAAAGCAGGC-3' and continue with the sequences indicated in the table below; reverse primers begin 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTT-3' and continue with the sequences indicated in the table. The primers for GBS1 are thus:

Fwd: GGGGACAAGTTTGTACAAAAAAGCAGGC**TCTCAATCTCATATTGTTTCAG**

Rev: GGGGACCACTTTGTACAAGAAAGCTGGGT**TATTTTTAGACATCATAGACA**

The full forward primer sequences are given in the sequence listing as SEQ IDs 10968-11492. The reverse primer sequences are SEQ IDs 11493-12017.

GBS	Forward	Reverse
1	TCTCAATCTCATATTGTTTCAG	ATTTTTAGACATCATAGACA
2	TCTAATTACATTATTACATTTTTG	GGGAATGCCTACAAA
3	TCTGATACTAGTTCAGGAATATC	TTTTTTACTATACTTTTTGT
4	TCTGATACAAGTGATAAGAATACT	TTCCTTTTTAGGCTTACT
5	TCTATTTTTCTTCATAGTCCAC	ATTAGCTTCATTTGTCAG
6	TCTGAATGGGTGTTATTAACTC	AGTTTCTTCTTTAAAATCAT
7	TCTACAAATTCTTATTTTAGCAA	CTCTGAAGCTGTAAAACC
8	TCTGTATCAGTTCAGGCGT	TTTATCAATGTTTGAAACG
9	TCTGCTGCTCTAGGACAAC	TAGTAAATCAAGTTTTTGCA
10	TCTTTTGTTGTTGCCTTATT	ATCCCTTCTATTTTCGA
11	TCTCCACCTATGGAACGT	ATGTAGTGACGTTTCTGTG
11d	TCTCAGAAAGTCTATCGGG	ATGTAGTGACGTTTCTGTG
12	TCTAGTGAGAAGAAAGCAAAT	ATTGGGTGTAAGCATT
13	TCTTCTTGGAATTATTGGAG	CTTAACTCTACCCGTCC
14	TCTGCAATGATTGTAACCAT	TTTTCTCTTATTAAAGAATT
15	TCTGCATCTTATACCGTGAA	ATACCAGCCGTTACTATT
16	TCTGCCGAGAAGGATAAA	TTTAGCTGCTTTTTTAATG
17	TCTGTTTATAAAGTTATTCAAAA	AAATACTACATTTACAGGTG
18	TCTAAGCCTAACAGTCAACA	TTGGTTATTCTCCTTTAAT
19	TCTGATGATAACTTTGAAATGC	ATTATATTTTTGGATATTTC
20	TCTGCAGTGATTGCAAGTC	GGGCTTTTTCTTAAAAA
21	TGTGCTGCATCAAAC	GTTGGCATCCCTTTT
21 Long+A527	TGTGCTGCATCAAAC	CTTTTGATGGGATTGG

22	TGTACTAAACAAAGCCAG	TTGATTTAACGATTTGA
23	TGTCAATTAACCGATAC	TTTATCTCCTCTAAAATAATG
24	TGCTCAAATGATTCAT	CTTTGATAAGTCAGACCA
25	TCTAAAAGTTCACAAGTTACTACT	GTAACCCCAAGCTGAT
26	TCTAGTCATTATTCCATAAAATT	TGATTTTGCAATATCAA
28	TCTAATCATATGCTGATTGAG	TTTTTGTAATTTAAGTACTAA
29	TCAGTTTGGATGTTAAC	TTCTTTTATATTAAGAGCTT
30	TCAACAAATGCAGATG	ATTCGGATAAATGTAGC
31	TGTTTTGTCATTATTGATAG	TCCATTTTTATCCTCAC
31d	TCTCTAACTTGGTTTTTATTAGA	TCCATTTTTATCCTCAC
32	TCTGGTTTAAAAGTGACTGAA	ATGACCTCTACTTTCCA
33	TCTCATCATTTAGGTAAGGAA	CTTGTAATCACTTGGAC
34	TCTGTTAGTAATCGCTACAATC	ATTAATCATGGTATTGGT
35	TCTAATCAAGAAGTTTCAGC	CCATTGTGGAATATCA
36	TCTCGAGTTTTAGCGGATA	TTTGTAAAGCAGTTCTT
37	TCTGTATTATTTTACCAATCACA	ATCATTCATATGATCTCTAGA
38	TTAGGAGTGGTAGTTCAT	ATTTTGATTGATTCTACTC
39	TTTTTATTGTTAGTATTAGC	TTTTGTTTTTTTCAAATA
41	TCTGTTTATCTAGCGGTTAGA	ATCTTCAACGTCCTCC
42	TATAACAGTTTAGTTAGAAGTC	AAAGTCAAAGGAAACTT
43	TTTAAAGGGTTTACATATT	TTCTTTATCTAATTTATAATAG
44	TTTAATACAATTGGTCG	TTGCAATGTTTTTTCT
45	TCTATGGAAAAAATTAGGATT	TAAACTTTGGATAATCTGT
46	TCTAGAGATGAGCAAGAAATA	GTTGAAATTTTGATATGA
47	TCTCAACAGATAGGTCTTTATAA	CTCCTTTACTATATAGCTAACT
48	TTTCTCTATAATTACTTCAAT	TTGTTTGTGAAGTAAAAC
49	TCTAATAAGGCATTATTAGAGG	TGATAATATCTCCATATTTT
50	TCTACACATTTAGTTGACTTAAC	GCATTGGCGCCATA
51	TCTAGTAAACAACACATTTATCTA	TTCTACACGACTTTTATTC
52	TCTCAAGAAACTCATCAGTTG	AAGACCTCCTCGAGAT
53	TCTGCAGAAGACATTGTTACA	TGTTTTTTCTTTCTGTTG
54	TATAATTTTTCGACTAATGA	TGGATTAGTTTGACCTG
55	TCTGACACAGTGTCTTATCCT	TTTATCGTAAGCACTTAGG
56	TCTGTGGAGCAAGTGGCCA	CTCCTTCCAGGCATCG
57	TCTCAAGAACTAAGTAACTTTGA	GTAAAAGTATCTTAAATAGTCA
58	TCTACTGAAACGTTTGAAGG	TGCCATTCCTCCTCT
59	TCTGATGAAGCAACAACTAA	TGTTACCTTTTTATTTTCT
60	TCTAATAAAGATAATCAAAAAACT	TTTTTCATGCGATTGA
61	TGTTTCTTTTTTATTCCA	GAGACGTTTCTTATACCTT
62	TATTACTTTGATGGTAGTTT	TGTACCATATGTTCTCTCT
63	TCTGTTCAATCATTAGCAAA	AAAAGTTGGACTACTTTC
64	TTTAAAGGTAATAAGAAGTTG	TCGTTTTCCACCC
64d	TCTAGTCAAGTTGACTCTGTTA	TCGTTTTCCACCC
65	TCTCAAAACCAGGTGACTG	ATTTGGGTAAATATAGTAAA
66	TTAAGATTTTATAACAACGA	TTTACGACTAACCTCAAC
67	TCTAATGTTTTAGGGGAAA	AATTCCTTTTGGTGG
68	TCCCAAAAGACTTTTG	GGCAGAATACACCTTC
68d	TCCCAAAAGACTTTTG	GGCTGACGTCGACGCA
69	TCTAAAGTTTTAGCCTTTGA	AACTCTCTTAATATATTCTTCT
70	TCTGAAATGGCTTTAG	GTCTTTTTCAATATTCTGT
70d	TCTACTAACTTATTGAGTAGAATCA	GTCTTTTTCAATATTCTGT
71	TGTAGCTCAAAATCTCAT	CTTCTCCTTAGGAGTAACG
72	TCTAGTTTATCTATTAAAGATGCC	ATTATTATCAATTAATAACTCTT
73	TCTATCAAAGAGGCGGTAA	GTCAAACATACTTCCAAA
74	TCTAAAGAGGATAAAAAGCTAG	TTTCGTCGTATAAGCA
74d	TCTAGTGTTTCAGGTAGTAGTG	TTTCGTCGTATAAGCA
75	TCTAAAAAATTAAAACACTCAA	TGTCCTCATTTTTTCAG

76	TCTGATGAAGTTACAACTTCAG	AATACTTGCTGGAACAG
77	TTATTCCAAAGTAAAATAAA	GTCTTTCTTCAATTTTGG
78	TCTCATAACCATCACTCAGAACACATGT	GTCGTGATTTTTATGAGT
79	TCTCCCAAGAATAGGATAAA	CCCAAACTGGCATAAC
79d	TCTAGTCAGTATGAGTCACAGA	CCCAAACTGGCATAAC
80	TCTGCAGAAGTGTCACAAGA	TGAAGGACGTTTGTTG
81	TCTTTTGATGGATTTTT	TTTTTTTAGTTTAAGGCTA
82	TCTACAAATGAAAAACGAAC	GTCCACCTTCCGAT
83	TCTGAAATTAAACTCAAAAATATT	AACATTGTTTTTCCTTTC
84	TCTCATACTCAAGAACACAAAA	ATGGTGATGATGACCT
85	TCTCCTAAGAAGAAATCAGATAC	ATTAACATTTTGAGGGT
86	TCTGCAGAACTAACTCTTTTAA	TTTTGCAAAATCAACA
87	TCTGCGGATACATATAATAACTA	GAATAAATAACTGTATTTTTT
88	TCTTACCAAAAAATGACG	ATTTTCATTAATTTCCTCT
89	TCTGAAGAGCTTACCAAAAC	GATAGCTAATTGGTCTGT
90	TCTAGATATACAAATGGAAATTT	TAAAAGATGAGCTTCTCG
91	TCTAAAAAAGGACAAGTAAATG	AATTTCAATATAGCGACG
92	TCTGATTCTGTCATAAATAAGC	CTTGTTTGTCTTTACCTT
93	TCTGAATTTTCACGAGAAA	ATTATCCTTCAAAGCTG
94	TACCAATTAGGTAGCTATAA	TGTGTCATATAATGTAACCA
95	TCTGTTAATACAAAAACACTTCT	TGATCTTAATTTTCGAG
96	TCTGGTCAGTCTAAAAATGAAG	CCAAACAGGTTGATCT
97	TCTAGCCAGGAGGTATATG	ATTTACATCAGACTGTGAC
98	TCTGAAACTATTAATCCAGAAA	TTTATGGCCAATAACA
99	TCTACAAGTATGAACCATCAA	TTTTTTAGTAGTTGTCAATT
100	TCTAAGGGGCCAAAAGTAG	GTAAGCTGAATTTTCGA
101	TCTATTACTTTAGAAAAATTTATAGA	ACGAGAGTGGTTATTGG
102	TCTGCCTTTTACTTTGGCA	TTTCTTCACTCTTTCTAGAG
103	TCTATTTTTTCCTTGATCAT	CGGCCAGTTTTTTTCTT
104	TCTGGTGAAACCCAAGATA	AACACCTGGTGGGCGT
105	TTAACAATTCATGGACC	ACTATTTCTAATTGCTCTG
105d	TTAACAATTCATGGACC	TGGTCCCGGTGCGCCA
105d	TCTCAAGGACCTCCCGGTG	ACTATTTCTAATTGCTCTG
106	TCTCAAAATCAAAATTCACA	CTTAGCAGATTCATCCC
107	TCTCTGGAGCCTTTTATTT	TTTACTATTTGAAAATTGG
108	TCTGGTAATCGTTCAGATAAG	TTTCATAGGAACTTGTATT
109	TCTATCCAGCAGATCAACT	GTCCACACCTGCGACT
109d	TCTAAACGGGTTCGCTATG	GTCCACACCTGCGACT
110	TCTGTAAAATTAGTATTCGCAC	TTTACCTAAGTAATATTCTGA
111.19	TCTGTTAGCGTTGATAAGGC	TCCCCGTCTTTTTTGT
112	TCTACAATTAAAAATCTCACTG	GTCGTAATCATAAAAGCC
113	TCTAGTAAAATCAAAATTGTAACG	TTCATAACGAACCATAAC
114	TCTAATCTTTTAATTATGGGTT	TTTGAGTTCTAGCAACG
115	TTTCAATACTATTTAAAAGG	TTTTTTATCTTCTTCTTGC
116	TCTACCGAGGAGCCATTAA	TTTTAAAACCTGGTAAAC
117	TCTGAACAATCACAAAAAACA	TCAGCTCGTACTGTTT
118	TCTATGGTGACGGTGCTGG	GTCCTCCTCAATTGGT
119	TCTAGTCAGCCGGTAGGGG	CTCTTTTATACGCGATG
120	TCTGGTGGAGCATTTGCTA	GTTATTTGCTCGTTGTT
121	TCTAATAAAGATAATCAAAAAACT	TTTCTCAAATGTTTTCAT
122	TCTGCTGCCACCAAGAAAG	TTTCAAATGATCTACAGC
123	TCTACAACAAATGTAATGGC	GGCTAGTGTCTGTCCG
124	TCAATGAATTTTTCATTT	ACCATCTATTTTTACCCC
125	TCTACAAAATATCAGCGAATG	AGAACCCGCACTCTCA
126	TCTACTAAGCAAGCAATGTC	GAACGCAACGGCTGCT
127	TCTACAAAAGAATATCAAAATTAT	TTTCATATCAAAAACTATCG
128	TCGACTAATTCGTTAAA	TTCTTTATCTCTTAATGCTT

129	TTTGAAATAGTATTGGAAA	CACAACAGTTATTTTTTCA
130	TCTATATTTTCTATTTTTTATTATGT	AGGCCCTTCTGAGTAG
130d	TCTAAAAAACAACTTCACAAC	AGGCCCTTCTGAGTAG
131	TCTAAAACAGATATTGAAATAGC	AAATAATCCAATGGCTG
132	TCTATTAAATATTATCATTTGCA	CTTTTCAAGCTTTTTCC
133	TCTGCTTTACGGAACCTTG	AAAATGATCAGTTTGAGG
134	TCTACTATTTCTCAACAACAATAC	TTTTTGGCTTAAGAAAG
135	TCTGAAAAAAAGAGTAGTTCAAC	CTTACGATACATTTTAAATTG
136	TCTAATCAATTATCAGAAATCA	TTCTTTTTTTTACTTTAGCG
137	TCTCAAGAGTATAAAACAAAAGAG	CCATTGCAATCCAGCA
138	TCTGCTGTATTTACACTCGTC	ATGTTTATGGCTTGCT
139	TCTGGCGGCAAGATAAAAT	TTTTTGATAAATCCCC
140	TCTGATGGGTTAAAGAATAATG	ATATGTGTATTCATCCTTT
141	TCTGATGTTGTAATTAGTGGAG	TACTTCTATTTTTCCATCTG
142	TTCGAATTAAGAGAAAGA	GTAATGCAATAAATCAAAA
143	TCTAGCTTTTTAGTGATTTCA	GGATTTTAGTTTCGCA
144	TATACGCATAGTGGAAC	CCCATTGATTTCGTCG
145	TCTGTTATTATCAGGGGCG	TACCTCTTTCAATACCAC
146	TCTGTTAGTCGTTCTCCGA	ATTACCGTTAGGTACTGTA
147	TCTGAGGAGCAAGAATTAAA	GGTATGGTTAACAGAATC
148	TCTATTCTAACAAAAGCAAGT	ATATACCCTAGACTTTTTGA
149	TCTAGTGGGCGTTCATGGA	AGGAGTTTTATTGATGATAT
150	TCTGATACCCCTAATCAACTA	AAATGATTGTGGAAAAA
151	TGCAGGAGCTGTCCGC	ATCAAAGAAGTTGACATTG
151 Long	TCTGTCCGCATTGGTAAAG	ATCAAAGAAGTTGACATTG
152	TCTAACTGCTTAGAAAATGAA	GTTAGATAAATTAACCAGTG
153	TCTAACAACTCCAGCA	CCCTTTGCTTCGTTGT
154	TCTGGAAAGGTCAGTGCAG	TTCCACAAGTCCGATT
155	TCTATTTTATTTTCAGATGAAC	TTGTTTGATTCGTCCT
156	TCTGCATCAGATGTTCAGA	ACTACCAAACTGCTGG
157	TCTAGTGACGTTGACAAATA	TTGTGTATTTTTAGTTAGGT
158	TCTATGACCATTTACTTCAATA	GTGGATAAAATTCGAAA
159	TCTCAAACTATTTTGACGC	CAGACTGACTAGGAGCT
160	TCTGATGAATATCTACGTGTCG	GACTTGTAATTGATTCGC
161	TCTGATGAGGTGGACTATAACA	GAAGGCACCACCACCT
162	TCTATTTTCTTGCTCTTAGTTG	GTTGTATAGATGAGTTAATCTG
163	TCTGAAACTGTCATTCAACTTG	ACGGTTTTTAAAGAATG
164	TATTTTTTAACAACAAAAAA	TTTTTCTTTATCTTCTGTG
165	TCTCCAATTTTTATTGGTTT	CGATTTTGTAAGAGCTT
166	TCTGCATCTTATACCGTGAA	CGACGAAGCTATTTCT
167	TCTACAATTTATATTGCTTGG	TAAGGCTTGCATTTTG
168	TCTGTTGGATTGATGTTGG	TTTTCCTAAAAATTTTCC
169	TGGAAACAAATCACAG	GGCATCTCCTAGCTTT
170	TCTGCAATAGTTTTTACTTTTTT	TGATAAAGGTAGTTCTACAC
170d	TCTGGTTCTTATCATTTAACAA	TGATAAAGGTAGTTCTACAC
171	TCTGCTAGACCCAAACAGT	TTTTAGATGTTTTTGTGG
172	TACACTCATATTGTTGAAAA	ATGATTGATAATTTTAAGC
173	TCTAATAGTACTGAGACAAGTGC	TGCTTTTTGATATGCC
174	TCTGCTTATGTCGTCAATTT	TAAAATAAAGTTCAGAAAAG
175	TCTGAATTACCTTCGTTTATC	TTTCTCCCTTGACTTTC
176	TCTAAACATCCGATACTTAATG	CTTTTTCTCAGATGCTT
177	TCTAATTATCCTTTTGCGA	GACATTGAAACGGAAT
178	TCTGGACTACGCGGAGTAT	TTTTATCAATGATGTTGA
179	TCTGCTATTGGAGCAGCTG	CATATGACGCAAACGC
180	TCTGATAAAGAAGGGATAGAGG	AGCCTCTTTTCTTGTT
181	TCTAAAGAAAAATCACAAACTG	ACGATTATCAACAAAGTT
182	TCTCAAAATAATAAAAAAGTAAAA	CATTCTTTTAAATACAAATC

182d	TCTCAAAATAATAAAAAAGTAAAA	GGGTTTGAAAGTTTTC
183	TCAAATGGTCAATCTAGC	TTTAACTTTAATTACTGGAAT
184	TCTAAGGATTCAAAAATCCC	TTTTTTAATAAGCTTCGA
185	TCTGGGCAACCATCTACAT	TTTTTTGTAAACTTCCTG
186	TCTCATTCACAGGATAGCA	CTTAGATACATTGTTTTTTTC
187	TCTGGACGAGGAGAAGTATC	CTTTCTTTTCTTACTTGC
188	TCACAATCTTCTCAAAA	TTTATTATTTTTAATACTTGAA
189	TCTGATAAGTCAGCAAACCC	CTTCAACTGTTGATAGAGC
191	TCTATCACGACATTACAGACT	TCCTTTAGCAGGAGCT
192	TCTAGATATTTAACTGCTGGT	GTTATACATGTTGTCTGAAG
193	TCTATAAAATATCAAGATGATTTT	CCAAATAATAACACGTTT
194	TTAGAAGTCAGAGAGCAG	GCTATCCCTTTCCAAT
195	TCTATTATGGAGACGGGTA	TGTATTTTTAATTTGTTTTC
195L	TCTTTGAATAATAAAGGTGTCG	TGTATTTTTAATTTGTTTTC
195LN	TCTTTGAATAATAAAGGTGTCG	CAAACTTTTAACATTTAATG
196	TCTATTTCCTCAAATTTTTACG	ATAGTGTAAGCTACCAGC
197	TCTAATTTTTATAAGCTCTTG	GTCATCATATTCCTGAAA
198	TCTGCGCTTAAAGAATTAA	TGTTCGGCGTAAGATT
199	TTTTTAAAAGAAATTGAAA	ATTGGTCATTTCTTGAG
200	TTTCGTAAATATAATTTTGA	AACAGATTTATTGGTTGG
201	TCTAGCGATACCTTTAATTTT	AGACTCATCAACTTTTTCT
202	TCTATGCTGATTAAGTCGC	GAACCCTGAAGGGTAG
203	TGTGGTAAAACTGGACT	CCAATTGTATTTTTCAAC
204	TCTAAGACAGGAGCACCCGT	ATTTATACTACCTGTTGAATC
205	TGCGAGTCAATTGAGC	TTTAAATTTGTAGTCTTTAATA
206	TCTACAAATACTTTGAAAAAAGA	CTCTTTTACTTTTCCAAAA
207	TCTAATTTATTTAAACGTTCCT	CCCTCCCTTAAGAGAA
208	TCTAAAAAGCGGCTAGTCA	TTGACGATGTTGCATC
209	TCTGGACAAAAATCAAAAATA	TTTCGAATTATTGTGACT
209d	TCTGGACAAAAATCAAAAATA	GTATTGTTGTTGCCTG
210	TCTGGAGGAAAATTTCAGAA	TTTTTGATTTCCCTTTC
210d	TCTACCTCATATCCTTTTATTT	TTTATAGTGTGTTTGCAA
211	TGTGGACATCGTGGTG	TTTGCTAGGAACTTTGA
212	TCTAAGACTAAAAAAATCATCA	TGATTCAATTCCTTTTC
213	TCTAAACACACCAGTAAAGAA	TTTTTCCTCTACTTTCTTA
214	TCTAAAAATAAAAAAATCTTATTT	TTTGCTCACCTCCACA
215	TTAATAAAAGGATTATTGTCA	CAATAACTTCTGTAAAATAAA
216	TCTGCTCGTTTAATACCACA	TTCACCCTTAAAATAATT
217	TCTAACACTAACATCCCTAGC	TGCATTTTTCCCTTCT
218	TCTAGAGGGAAGGTTATTTAC	CTCCAGTAAAGTATTAGTATTT
219	TCTATCAATAAAGTAACAGCTCA	GTGAGGTTTTGGTAATT
220	TCTAGAACACTATTTAGAATGATAT	TGCATATAAGTTTTTTAGC
220d	TACTATGCGAATCACAG	TGCATATAAGTTTTTTAGC
221	TCTAGTTTAGCATTGCAAAT	CTCATCTAAAGTGCTATCC
222	TCTACATTTTATAAAAAGACGG	CTCGTATTTAGGCAACT
223	TCTAAGAAAATACGAAGCTATAC	ATTGGATATGCCATAAA
224	TCTGGAGGAAATGAAATATTA	GACTTTTTGATGTTTACTTT
225	TCTGGTATGTCTAATAAGGAAAT	TTCTTTACTATAAACATCTTCA
226	TCTAACAAACTTATTACAGAAAA	AGCATTTAAAGTTGAATGT
227	TCTGTTTCATATGAAAAAGTCC	GTTAGTCTCTTCAAGATCA
228	TCTAGTAGAGGTATTTTTTTACAA	AAGACCTACCGCCCAA
229	TCTGAACGTCGGGTAAGTC	TACTTCTTTCTCTTTCAATT
230	TTTTTAATCGATTTTATTT	CTTAGTGTTCCGATATGA
231	TCATTAATTATTCTTACGGT	TCTTGTTTTAAGAGCAGA
231d	TCTTTATACGTTGTTAAACA	TCTTGTTTTAAGAGCAGA
232	TGGCTAAGTAAGCATGAG	ATCATGTTTTCCCTCAA
233	TTCCCAGCTAGCTGTC	ATCTGATATATCCGTTTTAT

234	TCTATAGAAATTGCTGTATTAATT	TTTTTTGTCTCCTTTTTTA
235	TCTATTCGATTTCTTATTCTTG	AAAGACACGATAAACATAAG
235d	TCTGACTCAACCACAGTCTC	AAAGACACGATAAACATAAG
236	TCTGCAGACCTTACAAGTCA	ATTTGCAACTTCTTGTATA
237	TCTATTGTATTTGCTATTGCA	TTTAAAAGTATCCTTAAATAAG
238	TCTGATATTTTTTCAGCTATTGA	CTTCCTCCTCAATAGTTG
239	TCTGTTAGTGCTGCTATTGAA	TTCTCCTCCCCCATTA
240	TCTAAGAAGCTTACTTTTATTTG	ATCCAAACGAGTGAAAT
241	TCAAAAGGATATTCAAGA	AGGTGTTGTTGTATTTTC
242	TCTCATAATATATTAAGATTTTTAGG	CTTTCTAAGTTTATTAAACATA
243	TCTATTCTTGGTCAAGATGT	GGCATCTGTTACCTTG
244	TCTCATGAAAATGTTAAAAAAG	AAACAACTCCATTATTTTT
245	TCTAAGTCAACGGTAACAAA	TAAACGTTGAAGAGCAT
246	AGGAAACGTTTTTCCT	CTTATCATATCTTGTTAAATCA
246d	TCTAACCATAAGGGAAAAGTA	CTTATCATATCTTGTTAAATCA
247	TCTGCTAAACAATTAATTGGT	TTGCCATGGGTTATAG
248	TCTTTGATGGTGTTGTTATTC	AGAATTAAAATTTTCATGC
248d	TCTAAAACTTATTTGTCAAATG	AGAATTAAAATTTTCATGC
249	TGGGCTTACCATACTG	TTTTTTAGATGTTTTATGTG
250	TCTGGCCTTAATCTTAAGC	CTCTTTTACTTTAGCTTCA
251	TCTCAATATTTTTTGAAACAAG	TTTCAAACTCCAGCCA
252	TTTATTTCAGGTTATATCAA	GGAGTGCCTTTCTACT
253	TCTGAAAATTGGAAGTTTGC	TTCATATCGTAAAGCATC
254	TCTATTGAAAAGGGAGTTG	ATCGTCAACCTTAACG
255	TCTATTGTTGGTAGAGAAATCA	TTTTACTTGACGTCTCAC
256	TATCATGTAAAAATTGATCA	GTCTTCCATTAATATTCCC
257	TCTGATTTTTTATACAAAGGAGG	CCAATTATTTTGAAAGTTC
258	TCTGAACGTTATACAGATAAAATG	ATTTTTTTGAATAATATAATCC
259	TCTCTTTCTCGTAAAAAAGAG	TTTATTATCAGAAAAGGC
260	TCTACTCTTGTCTTAGTTGTTTAT	ATTCAAAAAATTTTTCAA
261	TCTATAAAGAAAGCTGAAAATC	CGAAACGTCAGGTAAA
262	TCTATAAAAAATGCTATAGCATA	ACTTATTTTTGATAATATTTCTT
263	TCTCAGCCTTCTAAACTACTTC	ATCAGCATTTCTACGAA
264	TCTGATTTGTTTAGCATGTTG	ATGTAGACTCCTAATGATTT
265	TCTCTTGCTTCCCTGATTT	TTTACTGTTCCTTTCGC
266	TCTCATCAATCAAATCATTATC	GAGATTAATTTGATTATATTTT
267	TCTATCTTTATTATCGGACAA	AACATCATTTCCTCCC
268	TCTAAAGAATTTATTAAAGAATGG	GTTGATAGTTCCAAAACG
269	TCTGCAGATGATGGTGGTT	TAAATGTGTTCCTACTAAATT
270	TTAAATGATGCAATAACAA	CATCAATAGCCGAGCTG
271	TTGCTGGATTATCCTC	TTTATTTTCCAAATGACA
272	TCTGTATTTATGGCAAATAAGA	TTCACTCGGAGTTGGAG
272d	TCTATGAGTTCTCTGGAAGTT	TTCACTCGGAGTTGGAG
273	TCTGGTGTCCTCAACTCTG	AATGTAAATGACAAAGGTA
274	TCTGTTCATGATTTTGGTGA	GTTTTTTAATGGTTTGC
275	TCTGGGGTTTGGTTTTATA	TTTATCATAAGCATCTAGAC
276	TCTCAATCAGACATTAAAGCA	CTGATCTCTTGTTGATGC
277	TCTATTTGGAGGGGGGAAA	AAGCAGGGGAGCAATA
277d	TCTACCAAATTTGACTGGG	AAGCAGGGGAGCAATA
278	TCTGTTACGTTTTTCTTAT	CTGAGCAACACCTGTC
279	TCTAAAAAGAAAAGTTTAATTAGC	GGCAATTTTGTGGCAA
280	TTTGATTTTTTTAAGAAAA	TTGCTTAGTTAATGGCT
281	TCTAAGAAATTAATTATAGGTATTT	AGGCGTTGAATATAATTC
281d	TCTGGTTTTTCGTTTTTGA	AGGCGTTGAATATAATTC
282	TCTCTATTCTCAGATGAAACAA	CTTTTCAACTCCAAACA
283	TCTGTTAAATTAAAATCGTTACTG	GAGTTGTCTTTTTTTGTC
284	TCTATGCAACGATTAGGAC	GCAATCACAATTGACAT

285	TTAGGTGAAAGCAAATC	CTTTGTCTGCTTCACTT
286	TCTGGAGGATTTTATATGAAAG	TTGTATCTTCTCCTGACC
287	TCTGCACACACACCTACTAGT	TTGGTTAATCGTCTTG
287d	TCTAACAATCGTTCAAAGC	TTGGTTAATCGTCTTG
288	TCTAAAAAGTTTTTAAAAGTTTT	TTTAGTTACTTTCATAAATGG
288d	TGGAATAATCATCAGTCA	TTTAGTTACTTTCATAAATGG
289	TCTCAATCTAAAGGGCAAA	ATATAATTCCTCTAAAACTAGC
289L	TCTCAATCTAAAGGGCAAA	CCACTTCAAATTAACTAAC
290	TATTACTTATCAAAAGAAAAGG	ATTCCTTGAACACGAA
291	TCTCAAGTATTAAATGACAATGG	GTGCCATTCATTCTCT
292	TTGAATCGTAAAAAAAGG	TTGTCCTGTGAACTGTG
293	TCTATGGGTCTAGCAACAA	AGGGTTTATTTGTTGAAG
293d N-term	TCTATGGGTCTAGCAACAA	TCCTGATTTATCCACTG
293d C-term	TCTGTTACAGCTAAACACGG	AGGGTTTATTTGTTGAAG
294	TCTGGTCATTTTAGTGAAAAA	CAAAATACCTAAGCTAGC
295	TCTAGCGACATAAAAATCAT	ACGAACTTCCATAACC
296	TCTAAAGGTATTATTTTAGCG	GGCTTCTCCAATCAAA
297	TCTATTCAGATTGGCAAATT	TTGAGTTAATGGATTGTT
298	TCTACTAAATTTATTGTTGATTCA	TAGCGTTATTTCACTGTG
299	TTTGAAATACTTAAACCTG	TTTCTCCGCCCAGTCA
300	TCTGCTTCTACAAATAATGTTTC	CCGTTTATTCTTTCTACTG
301	TCTGTAATTAATATTGAGCAAGC	CATATCTGTTGCATCAAT
302	TCTGAAATCAACACTGAAATAG	AACTGGCTTTTTAGTCAG
303	TCTACAAGGCATATAAAAATTTC	TTTATTATTTAATTCTTCAATA
304	TCTAACGAAATCAAATGCCC	GTCTTTTAGAGCATCGA
305	TCTGGACGAGTAATGAAAACA	CTCTCCTCTAAGACTTTCG
306	TCTGGGAAAAAAATTGTTTT	TCCTTTTGTTACTTTTGC
307	TCTAAATTTACAGAACTTAACTTAT	TTTATCGCCTTTGTTG
308	ATGACACAGATGAATTTTA	ATGTTCAGGTTCTCCG
309	TTGCAACTTGGAATTG	TTCCATTATCTTCAAGTTA
310	TCTGCTAAAGAGAGGGTAGAT	CTCTTCTTCATTTTTCTTA
311	TCAATTATTACTGATGTTTAC	TTTTTTTAAGTTGTAGAATG
312	TCTACTGCAACTAAACAACAT	GTTTTTTGATGCTTCTTG
313	TCTAAACGTATTGCTGTTTTA	TTTACTACTTTGGTTGGC
314	TCTAAATTTTATCTTGTTAGACAC	GTGTGTCATTTTGACCT
315	TCTATAGGGGATTATTCAGTAA	TCCTTCAAGATCATTTAA
316	TCTACTGAACGAACATTCGA	ACCTCCTTTTCTTTCATT
317	TCTAATAAGCCATATTCAATAG	ATCTTCTCCTAACTTACCC
317d N-term	TCTAATAAGCCATATTCAATAG	ACTAGCTAGATTCTTAACGC
317d C-term	TCTGACTTGAATGGCAATAT	ATCTTCTCCTAACTTACCC
318	TCTATTGATTTTATTATTTCTATTG	GCCTCTTTCTCCAAAT
319	TTAAAACATTTTGGTAGTAA	ATGTCCTGTTATATCTTCTT
320	TCTACTATTTATGACCAAATTG	GCGTTGAATAATGGTT
321	TCTAAAAATAAAAAAGATCAGTT	TATTTCTTTAGTTTCTTCAA
322	TCTCAAGAAACAGATACGACG	TAATAAAAATTATATAAGAACCT
323	TCTGGTAATGAGTCAAAGAAC	TTCTGTCTTATAAGCATAAG
324	TCTGGAAGTAAATCAGCTTC	TTTTTTATAAGCATGTGTA
325	TCTGCTTGGCAACTTGTTC	ATGAGACATAAGGTCTTG
326	TCTGGCATCTCAGACTTACC	GTTGGAGCTCCTACTG
326L	TCTAAATTCAAATCTGGGG	GTTGGAGCTCCTACTG
326L N-term	TCTAAATTCAAATCTGGGG	CATTTCTTTGGTTAAAGC
327	TCTGGAGGGAAAATGAATC	TATCTCGAGTGCTATTTG
327d N-term	TCTGGAGGGAAAATGAATC	CTCTTCATCGACATAGTAA
327d C-term	TCTGGCAACTTCAAAGCAT	TATCTCGAGTGCTATTTG
328	TCTGACCAAGTCGGTGTCC	ATTTTACAGTAGTGGAGTTT
329	TCTAAATCAAAGACCTCTTCTA	TGTCCTCATTTTTTCA
330	TCTAATAAACGCGTAAAAATC	TTTAACAGTACGAACACG

331	TCTACCAGAACAGTAGCAAT	CCCCCTGTTTTTAAAAT
332	TCTACAAAAAACCTGTTATTAA	ACCCTCATATGATTCC
333	TCTATTGATATACAAAAAATAAAA	TTTAAAATAATGATACATCTC
333d	TCTGGATCATTGAGGGCAA	TTTAAAATAATGATACATCTC
334	TCTAATTTAGTAAAAGTGAATAGTG	TAACCCCGTCTCAACA
335	TCTGAAGAAGAAAAATATTTTGA	TATTTTCGTTTTCTCAAA
336	TCTCAGGTTGAAGTTGACTTA	TTTCTCCAAATAATCTCTC
337	TCTGAAACAGATTCGTTTGTA	CCTACTTTTAGTTTTAGAAGA
338	TCTGCTATAATAGACAAAAAG	GAAATCATAGCTTCCC
339	TCGAAACCGATTAAGAT	ACCTTTTACTTTTGGTAGT
339d	TCTCAAGTCATGCGCTATG	ACCTTTTACTTTTGGTAGT
340	TCTGGATTTCTCTATAATTACTTC	TTGTTTGTGAAGTAAAACG
341	TCTGGAAAACCATTGTTAAC	TAATTTAAAAATTGCATAAA
342	TCTCAGAAAATTGAAGGTATT	TTTCGTTACCATATCTAGA
343	TCTGAAATGCAAGTTCAAA	TAAATCATGGAAACTAGC
344	TCTGCACAACGCAGAATGT	AAAGCCCAACCTTCCG
345	TCTAAAAACCTGAATTGGG	GTTTCCACGTCCTTTC
346	TCTAATAAAATAGCTAATACAGAAG	AAGTTTATTCAAATCTGG
347	TCTATTGATATTCATTCTCATATC	AATGTAATGGTTTTTTAATA
348	TCTACTGGATCTAAAAAATTAGC	AGCTAAAATACCTAACCAG
349	TCTAAAGATCGCTTATATAATAAA	ATTTTTTAAACGACTCAT
350	TCTGCAAAAGATATAATTAAGGTT	AGCGGAACGGTGAATA
351	TCAGAAGATCAAAAACA	ATAATCTAAACTATCAGCTCT
352	TCTACTTTTTTTAAAAAGCTAAA	ATCTCCTATTGTAATTTTGA
352d	TCTGGTACAGATAGTAAATTTGG	ATCTCCTATTGTAATTTTGA
353	TCTACAATGTTAAAAATTGAAA	CACCTCTTTTGTCAGA
354	TCTATTAAAGAACTAAAAGAATTT	TTTGTTAGCGAGTAAGTC
355	TCTCGCTCACTACCTT	TTTATCATCCTCCTTAATAA
356	TCTAAATTCTATATTATTGATGATG	AAACGTTTTACTCTGTAAAA
357	TTGGAACATTTTTATATTAT	AAATAAGAATGTTAAAAGAGC
358	TTTTATACAATTGAAGAGC	TTCCCCAAAAATTTCT
359	TCAAGAAATAATTACGGT	ACGCAGTCCCATTTTC
360	TCTATAATGAAGGCGGTCT	CTGGCATGAGGTCTCA
361	TCTAGCGTATATGTTAGTGGA	CCTTTTTTCAATAATAGC
362	TCTACTAAACCACAGGGGG	ATCTTTAATCTTACCATCC
362d N-term	TCTACTAAACCACAGGGGG	TGCTGCTACTGCAATG
362 C-term	TCTGGTAATGAAGGAAATATCAC	ATCTTTAATCTTACCATCC
363	TCTCTCGAATTAAAAAATATTG	TAAATTCCTTTGTTGTAATA
364	TCTAACTATATGGGTATGGGC	ACCATCAGTTGTCACC
365	TCTGGAACTGCTACATATAGTAGG	TATTGACCAGTGCACG
366	TGGCTTGACATTATTTT	TTTTTTTGAATTTGTAAAAG
367	TCTAAGAAATTAAAAATATTCCC	AGAGATTATTTTTATTTTAAAT
368	TCTAAAATCATTATTCAACGT	TTTATTTTTAGTATCTAAAACG
369	TCTAGTAGAATGATTCCAGG	TTTAGAAACTCCAAGTATCTC
370	TCTACCGAATTTAATGACG	GTTAATTTGACTATTGATATATT
371	TCTAAAGATAGATATATTTTAGCAG	TAAACTCTCAAAAGCTAAAC
372	TCAGAAAAATATTCCACT	ACGTTCTTCTCTGGCT
373	TCTGAAATTGGTCAGCAAA	ACTTAAATGGAACAACC
374	TCTAAGTTCGAAAATATAATATATG	TTTGCCTAAAAAATTAGG
375	TCTGAAAAAGAAACTATTTTAAGT	GGCTTTCCTCCCTTCA
376	TCTAAAGAAAAGAAAAATTTGG	TTCATCTTTTTCAATATCA
377	TCTGGTAATAAACTGATGTATCA	GTGAGAGTGTCTTTGTTT
378	TCTGAAGATCAACTCACTATATTT	CAGATTTTTAGCTACTTGTC
379	TCTCAAATTACCCGAGAAG	TCTAGAGCGCTTTATAAG
380	TCTCTTAAAAGATTACTTACTGAAG	TTTTCTAATAGTTAGAAGCC
381	TCTCTTGGGATAGCTCACA	TTTTAAATGTGCAGAGA
382	TCTATAAAGTTTAAATTATTTTTTAA	ATTTATAATTTCCTTGGG

383	TCTATTTTACAGACGAATATACTAT	TCTATAATATCTCTCTAAAGTGA
384	TCTAGAATAATTGTTGTCGG	CCTCGCTAACATATCAC
385	TCTAATGTAAAAAAACGC	AGCTCTTACAGTCTTGC
386	TCTCTAGTATCAAAGGAGAAAGC	TTGTCTGAGTGACCAA
387	TCTGGTATGTTGTTAGCA	ATAATATGAAATATGTTGTTCA
388	TCTCTTATGATAATAAATTCATTCG	TCCGCAGAGTAAAAAA
389	TCTATGAATAGTGAACATAAAATT	TTCATAAATGTGCCAA
390	TCTAGGGAAACTTACTGGA	TTCATCTCTGCTCACC
391	TCTAAAAAAGTCATCGATTTAA	TTCTCCTTCAGCTTTTA
392	TCTATTACATATGATTTCACAAG	GTCATTTTTTCTAAAGTTTG
393	TCTAATAAATCTTGGTTGAGAA	TTTTTGTAGTTGTTTCAAT
394	TCTCCTATGTTGTCTGTTGG	TTTCATTAGATAACTATTCAGC
395	TCTACTTATCAAAAAACAGTTG	TATAGACTGAAGATAATTAATTAA
396	TTTGTCAAAGGGATTT	AAATCGATTAATCAAGTC
397	TCTAAATTATTTGATAAGTTTATAGA	TCTAAAGTAGTCCTTTAGACTA
397d	TCTAAAACTGCTACAGTTAG	TCTAAAGTAGTCCTTTAGACTA
398	TATTTAGAACAATTAAAAGAGG	TTTGTCCATAATCATTTC
399	TCTAAAGTTTTAGTAGTTGATGAT	GGTAGATATGCCTAACATT
400	TCTAAAATAGTTGAAGGCG	GTTTCCTTCCAAAAAA
401	TCTGGAATTGAATTTAAAAATG	TCCATGCTTAATAGCC
402	TCTGGAAAATATTTTGGTACAG	ATCTAAACCAATTTCTGTAC
403	TCTGAGGTTAGAATGGTAACTC	GTCCACAAAAACGTCT
404	TCTAAAATAGATGACCTAAGAAA	TAGATGTTCTACGGAGAA
405	TTGAAAATTCAGTATTATCA	AAAGATGGCAAGCCAT
406	TCTGATAAAAATAATTTAGAAGACT	TCTCTCTCCACACCATA
407	TCTAAAATTGACATGAGGAA	CTTACCTCCTGTGGCT
407d	TCTAAAATTGACATGAGGAA	CTTTTGTTGGTTACCTC
408	TCTAACCACTTACTTAACCTCA	TATTGTTAAATATGATGAAATG
409	TCTAAGGTAGTAGTAGCTATTGAT	ATGATTATACAAATTGATTAAT
409d	TCTACTGAAGAGAGAAATCCT	ATGATTATACAAATTGATTAAT
410	TCTGCTTTATTATCAGTTATTGTC	TCCCTCTTCCTTGACA
411	TCTAAAGACTATATTAACAGAATATT	AACGTTTTTGAGCTTT
412	TCTGGATTTTTTGCACAGC	TTTTGTCTTAAACGTTCT
413	TCTATTGTTGGTGAACAAGA	TTTAGATAGTCTAGCCATTT
414	TTAAATCAATATTTTCTGC	ACGGCTTGGGGCAGAG
415	TCTGAGCGAATTCCTGTTC	TACCATTATCCGTGCT
416	TCTGAAGTCATTCGTGAACA	ACTATTAAACTCCAATGTTA
417	TCAAAACAATATGATTATATC	GCGCATTGTAACAAAT
418	TCTAGCAAGCCTAATGTTG	TTTTGGTAAAAGGTCTG
419	TCTGATTTAAATAATTACATCGC	TCCTGGAAAGTTCATC
420	TCTAAACGTGAATTACTACTCG	TAGTTTATCTAAAGCGTTC
421	TCTATACGCCAGTTTTTAAG	TTTATGTATAGAAACAGCAG
422	TTTTCGAGCGATTTTG	AATGTACATAACAATAGAGAGC
423	TCTGTAACCAAAGTTGAAGAG	CAACGATCCCAAGAAC
424	TCTATGAAAGATTTTATTGAATG	GCCATTCTTACCTCCT
424d	TCTATGAAAGATTTTATTGAATG	ACGTTTTTTCTGACCG
425	TCTATAGCCTTTAATAGTTTATTT	TATAAAATAAATTTGAAGATCT
426	TCTD440ACAGTTTATAATATAAACCATG	ATCATCTTGTACCAACTC
427	TATTCTTTTGAAGAACTTTT	GCCAATAAATTCACGG
428	TCTATAAAAATTTTGATCCC	AGTCTGTTTTTTAACAAAAG
429	TCTAATCATTCCATTGAATC	TGGTTTTAGAACAACTTTA
430	TTACAAAAAAAATATCGG	AATTAAGCTGAAAATGAC
431	TCTGCGGCTCAATTAGCTG	ATTATATTCTTTTAATTTGTCA
432	TCTCGTACCTTCAAACCAG	CTTACGACGTCCTGGA
433	TCTATTAAAGCAACTTTTACTC	GTGTGTCATGACTACTGTAC
434	TCAATTTTTCAGACAACA	TGAGTAGAGCACAAGC
642	TCTAGAAAACGTAATGATACATT	GAAACGAATACGTTCTT

643	TCTGATTGTCAAATTACACCA	ACTACCTACCGTTTTCAC
644	TCTATTTTTCGTGGTGATAA	TTTGATGGTAACAGTCG
645	TTTTTTAATATTGAATATCAC	AGAAAGGCGCTCTTCT
646	TCTAAGGGAGTCCAATATATG	TATCTTTAATAAAGCCCTA
647	TCTCGTCGCATGAATACCA	CATCCCATAAATTTGTT
648	TCTATAGAATTTTCAGGGC	CAAGACATTTCTTAAAGC
649	TCTGCTACTCACTCTAACTCAG	TTTTGTTTTAGCGATG
650	TGCTCTTCTTCAAATACT	TTTTAAACCATGCTGT
651	TCTCTAACACCATTTACAAAAG	TTTGTAAAGACCTTCTTT
652	TCTCAACAAGGTATTATGGATA	TTCCTCGTTTATTAATTT
653	TCTAAAATTTTAGGTACACCA	AAAGAAAAGATGTGCC
654	TCTGGAAAAATGGTTAAGAA	CTGTGCAGGCTCAAAT
655	TCTAAATTCGTCCGAACCGT	AATTGTCCAGTCTAAGTTA
656	TCTGGTCTTCCAACGCAGC	ATTTAGTGTTATTTCTCCTG
657	TGCTCAGGTAAAACAT	TTTTTTAAGTGATGATGAA
658	TCTGAAAGCAAATCTTTGC	CTTTGTCTGCTTCACTT
659	TGTGCTAATTGGATTG	TTTTGGGGTTACTTTAC
660	TGTGGAAATGTCGGAG	TTTTGCTGAAATAATGTT
661	TGTCAGTCAAACCACA	ATCATACGAATGCAAC
662	TCTGCTAGTTTTTATTTTTTCC	TTTTTCATATTTTTTCAAA
663	TGTGGAAGTAAATCAGC	ATTATTTTTATAAGCATGTG
664	TCTGTTAAATTAAAATCGTTACTG	GAGTTGTCTTTTTTTTGTC
665	TCTATTGCTGGTCCTAGTG	GATAAGCACTTTCCTTAA
666	TTATTTTTTGGAAATTGG	GCCTAAAAACCAATCA
667	TCTGCTGTATTTACACTCGTC	ATGTTTATGGCTTGCT
668	TTTTATATGAAAGAACAACA	TTGTATCTTCTCCTGACC
669	TCAATTATTATTGGGTTAA	ATATACCCTAGACTTTTTGA
670	TCTCCTAAATTAACCCTAGTCT	GGCTTTAAAGTTCGATA
671	TCTAGTCTTGCGAAGGCAG	TTTATCGTAAGCACTTAGG
672	TCTGTATTTACACTCGTCTTACA	ATGTTTATGGCTTGCTT
673	TCTGGAGGATTTTATATGAAAG	TTGTATCTTCTCCTGACC
674	TCTGTTAAATTAAAATCGTTACTG	GAGTTGTCTTTTTTTTGTCT
675	TCTGGTTCATCAGACAAACA	TTCAACTTGATTGCCA
676	TCTGTAGTTAAAGTTGGTATTAACG	TTTTGCAATTTTTGC
677	TCTGTATTAGAAGTACATGCTGA	TTTTAATGCTGTTTGAA
678	TCTGAGACACCAGTAATGGC	TTTTTTAGCTAAGGCTG
679	TCTGCTAACAAGCAGGATC	TTTTGCTAAACCTTCTG
680	TCTAATAAGTCCAGTAACTCTAAG	ATTCATATTAACACGATGC
681	TCTGCTTTTGATGTAATTATGC	TTTGCGTTTTGGAGGG
682	TCTATTAACTATGAGGTTAAAGC	TGCACCTTGATGGCGA
683	TCTGTAATTGTTGAACTTAGTTTG	CCATAATATTTGATGCTG
684	TCTCTTAGGAAGTATAAGCAAA	TTCTAATCCTACAGCATG
685	TCTAAAATTTGTCTGGTTGG	AAAAATTCCTCCTAAATTAA
686	TCTGACTTTTATGATATCAATCTT	AAAGTTTTGACTATTACTGATAG
687	TATGCTATTATGCAAAAAG	TGGGGGAGATAGTTATG
688	TCTGCAATCGTTTCAGCAG	TTGACAGAAAGCTAATTG

TABLE III – RESULTS FOR *in vivo* GBS CHALLENGE

GBS #	% survival	
	Pre-Immune	Post-Immune
1	18.7	22.2
4gst	19.4	37.2
4his	25.0	75.0
8	14.3	42.1
10	29.1	36.0
15	30.0	60.9

GBS #	% survival	
	Pre-immune	Post-Immune
110	11.1	30.0
113	17.6	29.4
114	40.0	52.2
117	27.8	36.8
119	36.4	52.2
139	23.1	26.7

16	33.3	53.8
18	29.4	50.0
21	5.9	10.0
22	36.8	63.1
24	38.5	41.4
25	28.6	85.7
32	20.0	25.0
35	0.0	17.6
45	26.7	37.5
48	20.0	25.0
52	14.2	17.3
53	23.8	29.2
54	22.7	44.0
55	50.0	52.9
57	33.3	55.6
58	6.7	11.8
62	15.8	36.4
63	21.4	42.9
65	3.7	23.3
67	23.5	27.8
71	13.3	26.7
73	28.6	39.1
80	38.8	56.5
84	33.3	37.5
85	30.8	62.5
90	14.3	22.7
94	25.0	30.0
95	16.7	23.1
98	5.9	11.1
100	26.9	42.9
103	16.7	52.9
106	10.0	18.2

150	21.6	44.4
153	25.0	30.0
155	22.6	36.8
157	14.3	31.8
158	22.6	40.0
163	29.6	37.9
164	25.0	43.8
173	17.9	38.7
176	20.0	38.9
177	21.7	33.3
181	5.0	21.7
186	41.2	52.6
188	11.8	23.5
189	21.4	31.6
195	32.1	64.7
206	33.3	50.0
211	30.8	33.3
232	50.0	57.1
233	34.8	55.2
236	57.1	70.6
243	46.7	52.9
263	15.4	35.7
273	61.5	75.0
276	23.8	44.4
296	25.0	28.6
297	13.3	23.5
298	20.0	22.2
302	30.0	52.2
304	33.3	40.9
305	42.1	70.0
316	38.5	42.9
318	7.1	15.8

TABLE IV – COMPARISON OF GBS*nnn* NUMBERING AND SEQ ID NUMBER

GBS numbering	Sequence listing
GBS1	SEQ ID 3532 & 8736
GBS2	SEQ ID 4530 & 8818
GBS3	SEQ ID 6266 & 8958
GBS4	SEQ ID 2 & 8786
GBS5	SEQ ID 2598 & 8674
GBS6	SEQ ID 398 & 8496
GBS7	SEQ ID 8790 & 9798
GBS8	SEQ ID 8694
GBS9	SEQ ID 4540 & 8822
GBS10	SEQ ID 8718
GBS11	SEQ ID 5884 & 8930
GBS12	SEQ ID 8764 & 9692
GBS13	SEQ ID 8484
GBS14	SEQ ID 5406 & 8892
GBS15	SEQ ID 4 & 8710
GBS16	SEQ ID 944 & 8538
GBS17	SEQ ID 1770 & 8602
GBS18	SEQ ID 6860 & 9002
GBS19	SEQ ID 4422 & 8812
GBS20	SEQ ID 308 & 8488
GBS21	SEQ ID 8762

GBS numbering	Sequence listing
GBS345	SEQ ID 2442
GBS346	SEQ ID 2768
GBS347	SEQ ID 2766
GBS348	SEQ ID 8658
GBS349	SEQ ID 2360
GBS350	SEQ ID 8698
GBS351	SEQ ID 2970
GBS352	SEQ ID 8692
GBS353	SEQ ID 3454
GBS354	SEQ ID 8754
GBS355	SEQ ID 8752
GBS356	SEQ ID 8724
GBS357	SEQ ID 8720
GBS358	SEQ ID 3184
GBS359	SEQ ID 3948
GBS360	SEQ ID 3926
GBS361	SEQ ID 8770
GBS362	SEQ ID 8768
GBS363	SEQ ID 3816
GBS364	SEQ ID 1452
GBS365	SEQ ID 1398

GBS22	SEQ ID 8584
GBS23	SEQ ID 8512
GBS24	SEQ ID 1694 & 8598
GBS25	SEQ ID 3180 & 8714
GBS26	SEQ ID 8820
GBS27	SEQ ID 8774
GBS28	SEQ ID 8738
GBS29	SEQ ID 8744
GBS30	SEQ ID 8860
GBS31	SEQ ID 8702
GBS32	SEQ ID 8910 & 10142
GBS33	SEQ ID 5734 & 8912
GBS34	SEQ ID 5750 & 8916
GBS35	SEQ ID 8908
GBS36	SEQ ID 8542
GBS37	SEQ ID 8564
GBS38	SEQ ID 2122 & 8642
GBS39	SEQ ID 8480
GBS40	SEQ ID 8654
GBS41	SEQ ID 1176 & 8562
GBS42	SEQ ID 4856 & 8850
GBS43	SEQ ID 672 & 8520
GBS44	SEQ ID 9000
GBS45	SEQ ID 9018
GBS46	SEQ ID 1834 & 8608
GBS47	SEQ ID 8588
GBS48	SEQ ID 8594 & 8596
GBS49	SEQ ID 8494 & 9490
GBS50	SEQ ID 1236 & 8566
GBS51	SEQ ID 5410
GBS52	SEQ ID 3920
GBS53	SEQ ID 8586
GBS54	SEQ ID 3442
GBS55	SEQ ID 9020 & 10338
GBS56	SEQ ID 2510 & 8668
GBS57	SEQ ID 8854
GBS58	SEQ ID 8664
GBS59	SEQ ID 3744
GBS60	SEQ ID 8760
GBS61	SEQ ID 8776
GBS62	SEQ ID 2244
GBS63	SEQ ID 390
GBS64	SEQ ID 374
GBS65	SEQ ID 8544
GBS66	SEQ ID 3028
GBS67	SEQ ID 3746
GBS68	SEQ ID 4012
GBS69	SEQ ID 4916
GBS70	SEQ ID 3718
GBS71	SEQ ID 8906
GBS72	SEQ ID 1348
GBS73	SEQ ID 220
GBS74	SEQ ID 5872
GBS75	SEQ ID 8926
GBS76	SEQ ID 5862
GBS77	SEQ ID 3256
GBS78	SEQ ID 3262
GBS79	SEQ ID 3264
GBS80	SEQ ID 8780

GBS366	SEQ ID 8574
GBS367	SEQ ID 1340
GBS368	SEQ ID 1598
GBS369	SEQ ID 4822
GBS370	SEQ ID 8844
GBS371	SEQ ID 4926
GBS372	SEQ ID 4956
GBS373	SEQ ID 5062
GBS374	SEQ ID 8878
GBS375	SEQ ID 326
GBS376	SEQ ID 5380
GBS377	SEQ ID 5468
GBS378	SEQ ID 5570
GBS379	SEQ ID 8918
GBS380	SEQ ID 156
GBS381	SEQ ID 8934
GBS382	SEQ ID 8610
GBS383	SEQ ID 4738
GBS384	SEQ ID 8836
GBS385	SEQ ID 1094
GBS386	SEQ ID 9038
GBS387	SEQ ID 8558
GBS388	SEQ ID 9040
GBS389	SEQ ID 8516
GBS390	SEQ ID 8952
GBS391	SEQ ID 8522
GBS392	SEQ ID 6220
GBS393	SEQ ID 8966
GBS394	SEQ ID 8960
GBS395	SEQ ID 6276
GBS396	SEQ ID 8468
GBS397	SEQ ID 6262
GBS398	SEQ ID 8806
GBS399	SEQ ID 1960
GBS400	SEQ ID 3154
GBS401	SEQ ID 3170
GBS402	SEQ ID 4236
GBS403	SEQ ID 8798
GBS404	SEQ ID 8800
GBS405	SEQ ID 8508
GBS406	SEQ ID 8506
GBS407	SEQ ID 6484
GBS408	SEQ ID 9042
GBS409	SEQ ID 6678
GBS410	SEQ ID 4064
GBS411	SEQ ID 9044
GBS412	SEQ ID 9046
GBS413	SEQ ID 272
GBS414	SEQ ID 8946
GBS415	SEQ ID 8944
GBS416	SEQ ID 6044
GBS417	SEQ ID 1874
GBS418	SEQ ID 5146
GBS419	SEQ ID 2638
GBS420	SEQ ID 2104
GBS421	SEQ ID 2108
GBS422	SEQ ID 714
GBS423	SEQ ID 6884
GBS424	SEQ ID 4874

GBS81	SEQ ID 2706
GBS82	SEQ ID 2898
GBS83	SEQ ID 8772
GBS84	SEQ ID 4182
GBS85	SEQ ID 216
GBS86	SEQ ID 2978
GBS87	SEQ ID 3452
GBS88	SEQ ID 5694
GBS89	SEQ ID 2682
GBS90	SEQ ID 8476
GBS91	SEQ ID 8938
GBS92	SEQ ID 8964 & 10238
GBS93	SEQ ID 2848
GBS94	SEQ ID 1592
GBS95	SEQ ID 2224
GBS96	SEQ ID 2130
GBS97	SEQ ID 800
GBS98	SEQ ID 8746
GBS99	SEQ ID 4240
GBS100	SEQ ID 8782
GBS101	SEQ ID 6902
GBS102	SEQ ID 6894
GBS103	SEQ ID 6
GBS104	SEQ ID 8778
GBS105	SEQ ID 1400
GBS106	SEQ ID 8502
GBS107	SEQ ID 6026
GBS108	SEQ ID 8532
GBS109	SEQ ID 4116
GBS110	SEQ ID 6832
GBS111	SEQ ID 8842
GBS112	SEQ ID 8904
GBS113	SEQ ID 300
GBS114	SEQ ID 8968
GBS115	SEQ ID 5164
GBS116	SEQ ID 5152
GBS117	SEQ ID 8962
GBS118	SEQ ID 2508
GBS119	SEQ ID 8814
GBS120	SEQ ID 8874
GBS121	SEQ ID 3826
GBS122	SEQ ID 9006
GBS123	SEQ ID 6310
GBS124	SEQ ID 260
GBS125	SEQ ID 3872
GBS126	SEQ ID 6736
GBS127	SEQ ID 8816
GBS128	SEQ ID 752
GBS129	SEQ ID 8990
GBS130	SEQ ID 9004
GBS131	SEQ ID 6198
GBS132	SEQ ID 8730
GBS133	SEQ ID 474
GBS134	SEQ ID 9008
GBS135	SEQ ID 8882
GBS136	SEQ ID 1188
GBS137	SEQ ID 3960
GBS138	SEQ ID 9052
GBS139	SEQ ID 884

GBS425	SEQ ID 3978
GBS426	SEQ ID 3976
GBS427	SEQ ID 6958
GBS428	SEQ ID 3398
GBS429	SEQ ID 3402
GBS430	SEQ ID 8840
GBS431	SEQ ID 8902
GBS432	SEQ ID 8534
GBS433	SEQ ID 2558
GBS434	SEQ ID 8590
GBS435	SEQ ID 484
GBS436	SEQ ID 8472
GBS437	SEQ ID 466
GBS438	SEQ ID 362
GBS439	SEQ ID 900
GBS440	SEQ ID 8536
GBS441	SEQ ID 936
GBS442	SEQ ID 940
GBS443	SEQ ID 998
GBS444	SEQ ID 1776
GBS445	SEQ ID 8634
GBS446	SEQ ID 2048
GBS447	SEQ ID 1654
GBS448	SEQ ID 8592
GBS449	SEQ ID 1634
GBS450	SEQ ID 1630
GBS451	SEQ ID 2098
GBS452	SEQ ID 2062
GBS453	SEQ ID 8636
GBS454	SEQ ID 1734
GBS455	SEQ ID 1690
GBS456	SEQ ID 1684
GBS457	SEQ ID 8656
GBS458	SEQ ID 8650
GBS459	SEQ ID 2152
GBS460	SEQ ID 2148
GBS461	SEQ ID 2394
GBS462	SEQ ID 2778
GBS463	SEQ ID 8688
GBS464	SEQ ID 8684
GBS465	SEQ ID 8682
GBS466	SEQ ID 2694
GBS467	SEQ ID 2350
GBS468	SEQ ID 8660
GBS469	SEQ ID 2998
GBS470	SEQ ID 2988
GBS471	SEQ ID 2924
GBS472	SEQ ID 2910
GBS473	SEQ ID 2882
GBS474	SEQ ID 2878
GBS475	SEQ ID 2856
GBS476	SEQ ID 8690
GBS477	SEQ ID 3112
GBS478	SEQ ID 3432
GBS479	SEQ ID 3460
GBS480	SEQ ID 3504
GBS481	SEQ ID 8734
GBS482	SEQ ID 8740
GBS483	SEQ ID 3606

GBS140	SEQ ID 8632
GBS141	SEQ ID 1768
GBS142	SEQ ID 8600
GBS143	SEQ ID 9054
GBS144	SEQ ID 2238
GBS145	SEQ ID 8700
GBS146	SEQ ID 8696
GBS147	SEQ ID 8526
GBS148	SEQ ID 9010
GBS149	SEQ ID 8732
GBS150	SEQ ID 3736
GBS151	SEQ ID 3188
GBS152	SEQ ID 3952
GBS153	SEQ ID 3904
GBS154	SEQ ID 4024
GBS155	SEQ ID 8796
GBS156	SEQ ID 4646
GBS157	SEQ ID 4812
GBS158	SEQ ID 5504
GBS159	SEQ ID 8628
GBS160	SEQ ID 8924
GBS161	SEQ ID 8922
GBS162	SEQ ID 168
GBS163	SEQ ID 224
GBS164	SEQ ID 1102
GBS165	SEQ ID 3672
GBS166	SEQ ID 8712
GBS167	SEQ ID 4214
GBS168	SEQ ID 9016
GBS169	SEQ ID 4346
GBS170	SEQ ID 8982
GBS171	SEQ ID 6720
GBS172	SEQ ID 6704
GBS173	SEQ ID 8788
GBS174	SEQ ID 6150
GBS175	SEQ ID 62
GBS176	SEQ ID 8478
GBS177	SEQ ID 8876
GBS178	SEQ ID 6078
GBS179	SEQ ID 8848
GBS180	SEQ ID 3062
GBS181	SEQ ID 1924
GBS182	SEQ ID 3774
GBS183	SEQ ID 4796
GBS184	SEQ ID 1978
GBS185	SEQ ID 1046
GBS186	SEQ ID 8470
GBS187	SEQ ID 844
GBS188	SEQ ID 3410
GBS189	SEQ ID 6986
GBS190	SEQ ID 8842
GBS191	SEQ ID 1814
GBS192	SEQ ID 8618
GBS193	SEQ ID 2382
GBS194	SEQ ID 3912
GBS195	SEQ ID 8
GBS196	SEQ ID 4944
GBS197	SEQ ID 5486
GBS198	SEQ ID 8896

GBS484	SEQ ID 3562
GBS485	SEQ ID 3552
GBS486	SEQ ID 3762
GBS487	SEQ ID 3756
GBS488	SEQ ID 3732
GBS489	SEQ ID 3730
GBS490	SEQ ID 3704
GBS491	SEQ ID 3698
GBS492	SEQ ID 3252
GBS493	SEQ ID 3244
GBS494	SEQ ID 3238
GBS495	SEQ ID 8722
GBS496	SEQ ID 8716
GBS497	SEQ ID 3876
GBS498	SEQ ID 3858
GBS499	SEQ ID 8758
GBS500	SEQ ID 4022
GBS501	SEQ ID 4106
GBS502	SEQ ID 1406
GBS503	SEQ ID 8580
GBS504	SEQ ID 4578
GBS505	SEQ ID 4566
GBS506	SEQ ID 8832
GBS507	SEQ ID 8830
GBS508	SEQ ID 4644
GBS509	SEQ ID 8828
GBS510	SEQ ID 8826
GBS511	SEQ ID 4892
GBS512	SEQ ID 4970
GBS513	SEQ ID 4974
GBS514	SEQ ID 8862
GBS515	SEQ ID 8864
GBS516	SEQ ID 8866
GBS517	SEQ ID 8868
GBS518	SEQ ID 9012
GBS519	SEQ ID 5068
GBS520	SEQ ID 8870
GBS521	SEQ ID 5228
GBS522	SEQ ID 322
GBS523	SEQ ID 8492
GBS524	SEQ ID 8894
GBS525	SEQ ID 5430
GBS526	SEQ ID 5414
GBS527	SEQ ID 5524
GBS528	SEQ ID 8898
GBS529	SEQ ID 5670
GBS530	SEQ ID 5630
GBS531	SEQ ID 5588
GBS532	SEQ ID 1324
GBS533	SEQ ID 8914
GBS534	SEQ ID 8550
GBS535	SEQ ID 8568
GBS536	SEQ ID 1288
GBS537	SEQ ID 5798
GBS538	SEQ ID 8920
GBS539	SEQ ID 158
GBS540	SEQ ID 8482
GBS541	SEQ ID 184
GBS542	SEQ ID 9048

GBS199	SEQ ID 1162
GBS200	SEQ ID 8936
GBS201	SEQ ID 4550
GBS202	SEQ ID 8666
GBS203	SEQ ID 6478
GBS204	SEQ ID 1996
GBS205	SEQ ID 18
GBS206	SEQ ID 8552
GBS207	SEQ ID 3822
GBS208	SEQ ID 3916
GBS209	SEQ ID 3918
GBS210	SEQ ID 3738
GBS211	SEQ ID 4680
GBS212	SEQ ID 8750
GBS213	SEQ ID 8500
GBS214	SEQ ID 8498
GBS215	SEQ ID 9022
GBS216	SEQ ID 8606
GBS217	SEQ ID 9024
GBS218	SEQ ID 8652
GBS219	SEQ ID 8646
GBS220	SEQ ID 2730
GBS221	SEQ ID 9028
GBS222	SEQ ID 3842
GBS223	SEQ ID 8794
GBS224	SEQ ID 9026
GBS225	SEQ ID 8834
GBS226	SEQ ID 4966
GBS227	SEQ ID 5030
GBS228	SEQ ID 5050
GBS229	SEQ ID 9056
GBS230	SEQ ID 1296
GBS231	SEQ ID 5810
GBS232	SEQ ID 5830
GBS233	SEQ ID 4722
GBS234	SEQ ID 1106
GBS235	SEQ ID 8560
GBS236	SEQ ID 6162
GBS237	SEQ ID 8706
GBS238	SEQ ID 4246
GBS239	SEQ ID 8980
GBS240	SEQ ID 8986
GBS241	SEQ ID 9030
GBS242	SEQ ID 9032
GBS243	SEQ ID 8678
GBS244	SEQ ID 6554
GBS245	SEQ ID 8994
GBS246	SEQ ID 6864
GBS247	SEQ ID 8856
GBS248	SEQ ID 454
GBS249	SEQ ID 8620
GBS250	SEQ ID 8634
GBS251	SEQ ID 2258
GBS252	SEQ ID 8648
GBS253	SEQ ID 2526
GBS254	SEQ ID 2710
GBS255	SEQ ID 2966
GBS256	SEQ ID 3424
GBS257	SEQ ID 3550

GBS543	SEQ ID 8932
GBS544	SEQ ID 5880
GBS545	SEQ ID 44
GBS546	SEQ ID 9014
GBS547	SEQ ID 12
GBS548	SEQ ID 8614
GBS549	SEQ ID 8612
GBS550	SEQ ID 4720
GBS551	SEQ ID 4710
GBS552	SEQ ID 1086
GBS553	SEQ ID 1088
GBS554	SEQ ID 1138
GBS555	SEQ ID 8748
GBS556	SEQ ID 5968
GBS557	SEQ ID 774
GBS558	SEQ ID 1192
GBS559	SEQ ID 1196
GBS560	SEQ ID 1268
GBS561	SEQ ID 8518
GBS562	SEQ ID 8676
GBS563	SEQ ID 2296
GBS564	SEQ ID 2300
GBS565	SEQ ID 8950
GBS566	SEQ ID 694
GBS567	SEQ ID 680
GBS568	SEQ ID 6300
GBS569	SEQ ID 8956
GBS570	SEQ ID 8972
GBS571	SEQ ID 8970
GBS572	SEQ ID 3300
GBS573	SEQ ID 3304
GBS574	SEQ ID 8726
GBS575	SEQ ID 8810
GBS576	SEQ ID 4418
GBS577	SEQ ID 8808
GBS578	SEQ ID 4382
GBS579	SEQ ID 4378
GBS580	SEQ ID 1932
GBS581	SEQ ID 8622
GBS582	SEQ ID 8624
GBS583	SEQ ID 1962
GBS584	SEQ ID 8708
GBS585	SEQ ID 8672
GBS586	SEQ ID 6444
GBS587	SEQ ID 8976
GBS588	SEQ ID 8804
GBS589	SEQ ID 8514
GBS590	SEQ ID 8510
GBS591	SEQ ID 630
GBS592	SEQ ID 8504
GBS593	SEQ ID 514
GBS594	SEQ ID 8978
GBS595	SEQ ID 6738
GBS596	SEQ ID 6712
GBS597	SEQ ID 6686
GBS598	SEQ ID 6674
GBS599	SEQ ID 6662
GBS600	SEQ ID 8988
GBS601	SEQ ID 8578

GBS258	SEQ ID 3752
GBS259	SEQ ID 8756
GBS260	SEQ ID 4162
GBS261	SEQ ID 1530
GBS262	SEQ ID 8572
GBS263	SEQ ID 1616
GBS264	SEQ ID 8824
GBS265	SEQ ID 4554
GBS266	SEQ ID 4652
GBS267	SEQ ID 4980
GBS268	SEQ ID 5038
GBS269	SEQ ID 5534
GBS270	SEQ ID 1998
GBS271	SEQ ID 8570
GBS272	SEQ ID 22
GBS273	SEQ ID 5994
GBS274	SEQ ID 774
GBS275	SEQ ID 2308
GBS276	SEQ ID 8942
GBS277	SEQ ID 8954
GBS278	SEQ ID 8524
GBS279	SEQ ID 6292
GBS280	SEQ ID 6254
GBS281	SEQ ID 4458
GBS282	SEQ ID 4444
GBS283	SEQ ID 9034
GBS284	SEQ ID 6456 & 8974
GBS285	SEQ ID 8802
GBS286	SEQ ID 9036
GBS287	SEQ ID 5354
GBS288	SEQ ID 5374
GBS289	SEQ ID 8616
GBS290	SEQ ID 8680
GBS291	SEQ ID 8530
GBS292	SEQ ID 8998
GBS293	SEQ ID 8582
GBS294	SEQ ID 8604
GBS295	SEQ ID 2722
GBS296	SEQ ID 2658
GBS297	SEQ ID 3024
GBS298	SEQ ID 8704
GBS299	SEQ ID 3268
GBS300	SEQ ID 4170
GBS301	SEQ ID 8576
GBS302	SEQ ID 8670
GBS303	SEQ ID 8554
GBS304	SEQ ID 5846
GBS305	SEQ ID 208
GBS306	SEQ ID 212
GBS307	SEQ ID 8992
GBS308	SEQ ID 8880
GBS309	SEQ ID 3386
GBS310	SEQ ID 286
GBS311	SEQ ID 3964
GBS312	SEQ ID 4660
GBS313	SEQ ID 4090
GBS314	SEQ ID 8556
GBS315	SEQ ID 1766
GBS316	SEQ ID 2000

GBS602	SEQ ID 8948
GBS603	SEQ ID 6132
GBS604	SEQ ID 5282
GBS605	SEQ ID 5302
GBS606	SEQ ID 8884
GBS607	SEQ ID 5314
GBS608	SEQ ID 8886
GBS609	SEQ ID 8888
GBS610	SEQ ID 8890
GBS611	SEQ ID 6028
GBS612	SEQ ID 8474
GBS613	SEQ ID 5092
GBS614	SEQ ID 8872
GBS615	SEQ ID 6052
GBS616	SEQ ID 8940
GBS617	SEQ ID 1824
GBS618	SEQ ID 6600
GBS619	SEQ ID 6608
GBS620	SEQ ID 6620
GBS621	SEQ ID 864
GBS622	SEQ ID 8640
GBS623	SEQ ID 8996
GBS624	SEQ ID 9050
GBS625	SEQ ID 2812
GBS626	SEQ ID 8858
GBS627	SEQ ID 8852
GBS628	SEQ ID 8784
GBS629	SEQ ID 6950
GBS630	SEQ ID 4502
GBS631	SEQ ID 4492
GBS632	SEQ ID 4488
GBS633	SEQ ID 8728
GBS634	SEQ ID 3066
GBS635	SEQ ID 8838
GBS636	SEQ ID 4772
GBS637	SEQ ID 8626
GBS638	SEQ ID 8984
GBS639	SEQ ID 8546
GBS640	SEQ ID 6780
GBS641	SEQ ID 900
GBS642	1312
GBS643	1772
GBS644	1956
GBS645	2726
GBS646	3348
GBS647	3770
GBS648	4934
GBS649	5076
GBS650	5446
GBS651	5602
GBS652	5610
GBS653	5760
GBS654	6096
GBS655	6656
GBS656	9324
GBS657	10782
GBS658	8802
GBS659	9344
GBS660	9410

GBS317	SEQ ID 4210
GBS318	SEQ ID 8548
GBS319	SEQ ID 892
GBS320	SEQ ID 916
GBS321	SEQ ID 8846
GBS322	SEQ ID 8540
GBS323	SEQ ID 2102
GBS324	SEQ ID 8490
GBS325	SEQ ID 8900
GBS326	SEQ ID 8630
GBS327	SEQ ID 5856
GBS328	SEQ ID 6016
GBS329	SEQ ID 8928
GBS330	SEQ ID 8792
GBS331	SEQ ID 922
GBS332	SEQ ID 1004
GBS333	SEQ ID 1786
GBS334	SEQ ID 1784
GBS335	SEQ ID 1782
GBS336	SEQ ID 1886
GBS337	SEQ ID 2010
GBS338	SEQ ID 8638
GBS339	SEQ ID 2080
GBS340	SEQ ID 8594 & 8596
GBS341	SEQ ID 2280
GBS342	SEQ ID 2266
GBS343	SEQ ID 8644
GBS344	SEQ ID 8662

GBS661	9428
GBS662	9286
GBS663	9294
GBS664	9034
GBS665	10546
GBS666	10610
GBS667	9052
GBS668	9036
GBS669	9010
GBS670	10730
GBS671	9020
GBS672	9052
GBS673	9036
GBS674	9034
GBS675	10634
GBS676	10692
GBS677	10746
GBS678	9330
GBS679	9404
GBS680	6668
GBS681	4264
GBS682	6762
GBS683	9290
GBS684	9614
GBS685	10454
GBS686	2774
GBS687	4620
GBS688	10224

TABLE V – NUCLEOTIDES DELETED IN EXPRESSION OF GBS*nnn* PROTEINS

GBS	Deleted nucleotides
11d	1-153
31d	1-129
64d	1-165
68d	2029-2796
70d	1-402
74d	1-975
79d	1-201
105dN	2689-4119
105dC	1-2688
105d	1-2688
109d	1-120
130d	1-518
170d	1-111
182d	1596-1674
195C	1-1710
195N	1711-3243
209d	757-912
210d	1-99 & 777-879
220d	1-120
231d	1-54
235d	1-270
246d	1-75
248d	1-591

GBS	Deleted nucleotides
272d	1-531
277d	1-318
281d	1-54
287d	1-108
288d	1-72
293C	1-1229
293N	1230-2379
317N	1729-4107
317C	1-2379
326N	1707-2652
326dN	2326-3927
327N	3034-6831
327C	1-3033
333d	1-150
339d	1-111
352d	1-158
362N	1707-2652
362C	1-1706
397d	1-348
399d	1-111
407d	1174-1473
409d	1-297
424d	1327-1671

TABLE VI – PREDICTED FUNCTIONS FOR CERTAIN SEQ IDs

SEQ ID	Function
6	manganese ABC transporter, ATP-binding protein (psaB)
12	iron (chelated) ABC transporter, permease protein (psaC)
18	peptidyl-prolyl cis-trans isomerase, cyclophilin-type
26	chorismate binding enzyme (pabB)
30	probable transposase (insertion sequence IS861)
42	peptidase, M20/M25/M40 family
44	drug transporter
50	ribosomal protein L11 (rplK)
54	ribosomal protein L1 (rplA)
62	peptide ABC transporter, permease protein
66	peptide ABC transporter, permease protein
78	uridylate kinase (pyrH)
84	ribosome recycling factor (frr)
104	PhoH family protein (phoH)
110	MutT/nudix family protein superfamily
116	tetracenomycin polyketide synthesis O-methyltransferase TcmP
134	phosphopantetheine adenylyltransferase (coaD)
140	PDZ domain protein
144	5-nucleotidase family protein
156	VanZF-related protein
158	ABC transporter, ATP-binding/permease protein
162	ABC transporter, ATP-binding/permease protein
168	BioY family protein
180	acetyl-CoA acetyltransferase
188	endonuclease III (nth)
196	glucokinase (gkl)
200	rhodanese family protein
204	elongation factor Tu family protein (typA)
212	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-
216	cell division protein DivIB
220	cell division protein FtsA (ftsA)
224	cell division protein FtsZ (ftsZ)
236	ylmH protein (ylmH)
240	cell division protein DivIVA (divIVA)
244	isoleucyl-tRNA synthetase (ileS)
252	MutT/nudix family protein
256	ATP-dependent Clp protease, ATP-binding subunit ClpE (clpE)
268	methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cycloh
274	exodeoxyribonuclease VII, large subunit (xseA)
278	exodeoxyribonuclease VII, small subunit (xseB)
282	geranyltranstransferase (ispA)
286	hemolysin A
290	transcriptional repressor
296	DNA repair protein RecN (recN)
300	degV family protein (degV)
322	peptide ABC transporter, permease protein (oppC)
326	peptide ABC transporter, ATP-binding protein (oppD)
328	peptide ABC transporter, ATP-binding protein (oppF)
348	4-diphosphocytidyl-2C-methyl-D-erythritol kinase (ispE)
352	adc operon repressor AdcR (adcR)
356	zinc ABC transporter, ATP-binding protein (adcC)
370	tyrosyl-tRNA synthetase (tyrS)
374	penicillin-binding protein 1B (pbp1B)
378	DNA-directed RNA polymerase, beta subunit (rpoB)
382	dna-directed rna polymerase beta' chain

390	competence protein CglA (cglA)
406	acetate kinase (ackA)
410	transcriptional regulator
418	pyrroline-5-carboxylate reductase (proC)
422	glutamyl-aminopeptidase (pepA)
432	thioredoxin family protein
436	tRNA binding domain protein (pheT)
440	methyltransferase
442	single-strand DNA-binding protein, authentic point mutation (ssbB)
454	GAF domain protein (lytS)
466	lrgB protein (lrgB)
474	oligopeptide ABC transporter, permease protein
476	peptide ABC transporter, ATP-binding protein
480	peptide ABC transporter, ATP-binding protein (oppF)
484	PTS system, IIABC components (treB)
488	alpha amylase family protein (treC)
494	transcriptional regulator, BglG family
506	transcriptional regulator, BglG family
508	PTS system, IIB component
514	PTS system, IIC component
518	transketolase, N-terminal subunit (tktA)
528	ribosomal protein S15 (rpsO)
546	cysteinyl-tRNA synthetase (cysS)
554	RNA methyltransferase, TrmH family, group 3
562	DegV family protein (degV)
572	ribosomal protein S9 (rpsI)
576	integrase, phage family
580	transcriptional regulator
596	recombination protein
626	transcriptional regulator MutR
630	transporter
640	amino acid ABC transporter, permease protein (opuBB)
642	glycine betaine/L-proline transport ATP binding subunit (proV)
654	lectin, alpha subunit precursor
662	transcriptional regulator
664	acetyltransferase, GNAT family
666	acetyltransferase, GNAT family (rlmJ)
670	acetyltransferase, GNAT family
676	transcriptional regulator, tetR family domain protein
680	ABC transporter efflux protein, DrrB family
690	IS1381, transposase OrfA/OrfB, truncation
714	magnesium transporter, CorA family
718	oxidoreductase, Gfo/Idh/MocA family
722	valyl-tRNA synthetase (valS)
730	acetyltransferase, GNAT family
746	methyltransferase
750	bacteriophage L54a, integrase
754	DNA-damage-inducible protein J
774	cation efflux system protein
778	oxidoreductase, aldo/keto reductase family
784	alcohol dehydrogenase, zinc-containing
790	3-oxoadipate enol-lactone hydrolase/4-carboxymuconolactone decarboxylas
804	ribonucleoside-diphosphate reductase, alpha subunit (nrdE)
808	nrdI protein (nrdI)
812	Ribonucleotide reductases
824	elaA protein (elaA)
828	RNA methyltransferase, TrmA family
832	RecX family protein
840	-identity (jag)

844	membrane protein, 60 kDa (yidC)
856	UTP-glucose-1-phosphate uridylyltransferase (galU)
864	rhomboid family protein
884	MORN motif family
892	transcriptional regulator
896	adenylosuccinate lyase (purB)
908	phosphoribosylaminoimidazole carboxylase, catalytic subunit (purE)
912	phosphoribosylamine--glycine ligase (purD)
916	phosphosugar-binding transcriptional regulator
920	acetyl xylan esterase
922	ROK family protein (gki)
926	N-acetylneuraminate lyase (nanA)
936	sugar ABC transporter, permease protein
940	sugar ABC transporter, permease protein (msmF)
952	LysM domain protein, authentic frameshift
956	zoocin A endopeptidase
958	phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydr
962	acetyltransferase, GNAT family family
964	phosphoribosylglycinamide formyltransferase (purN)
968	phosphoribosylformylglycinamidine cyclo-ligase (purM)
972	amidophosphoribosyltransferase (purF)
980	phosphoribosylformylglycinamidine synthase
984	phosphoribosylaminoimidazole-succinocarboxamide synthase (purC)
1042	oligoendopeptidase F (pepF)
1060	ebsC protein
1068	hydrolase, haloacid dehalogenase-like family
1076	riboflavin synthase, beta subunit (ribH)
1082	riboflavin biosynthesis protein RibD (ribD)
1086	Mn2+/Fe2+ transporter, NRAMP family
1094	peptidase, U32 family
1116	HPr(Ser) kinase/phosphatase (hprK)
1130	oxidoreductase
1148	signal recognition particle-docking protein FtsY (ftsY)
1152	Cof family protein
1156	Cof family protein
1172	vicX protein (vicX)
1176	sensory box sensor histidine kinase (vicK)
1180	DNA-binding response regulator (vicR)
1184	amino acid ABC transporter, ATP-binding protein
1188	amino acid ABC transporter, amino acid-binding protein (fliY)
1192	amino acid ABC transporter, permease protein
1196	amino acid ABC transporter, permease protein
1208	DNA-binding response regulator (vicR)
1210	threonyl-tRNA synthetase (thrS)
1214	glycosyl transferase, group 1
1218	glycosyl transferase, group 1 (cpoA)
1222	alpha-amylase (amy)
1230	proline dipeptidase (pepQ)
1238	haloacid dehalogenase-like hydrolase superfamily
1244	mannonate dehydratase (uxuA)
1248	glucuronate isomerase
1254	transcriptional regulator, GntR family
1268	sodiumgalactoside symporter family protein
1270	D-isomer specific 2-hydroxyacid dehydrogenase family protein
1282	transcriptional regulator, LysR family
1290	ABC transporter, ATP-binding protein (potA)
1296	DedA family protein
1308	MutT/nudix family protein family
1310	phosphoserine phosphatase SerB (serB)

1312	septation ring formation regulator EzrA
1320	hydrolase, haloacid dehalogenase-like family (gph)
1340	sensor histidine kinase (vncS)
1348	transmembrane protein Vexp3 (vex3)
1352	ABC transporter, ATP-binding protein (vex2)
1358	transmembrane protein Vexp1 (vex1)
1366	transposase
1374	integrase, phage family
1390	holin 2
1398	minor structural protein
1400	host specificity protein
1404	minor structural protein
1406	PblA
1486	homeobox protein drg11
1488	reverse transcriptase
1496	p22 erf-like protein
1498	gp157
1500	tropomyosin 2
1512	gp49 homologous
1526	transcriptional regulator-related protein
1566	chorismate mutase
1572	PTS system component
1576	PTS system, IIB component
1580	PTS system IIA component
1584	lactose phosphotransferase system repressor (lacR)
1594	adhesion lipoprotein (lmb)
1602	GTP pyrophosphokinase (relA)
1606	2`,3`-cyclic-nucleotide 2`-phosphodiesterase (cpdB)
1616	iron ABC transporter, iron-binding protein
1620	DNA-binding response regulator
1630	PTS system component
1634	PTS system component (manM)
1638	PTS system component (manL)
1642	PTS system component
1658	response regulator BlpR (blpR)
1676	phosphate transport system regulatory protein PhoU
1680	phosphate ABC transporter, ATP-binding protein (pstB)
1684	phosphate ABC transporter, permease protein (pstA)
1690	phosphate ABC transporter, permease protein (pstC)
1694	probable hemolysin precursor
1704	ribosomal protein L11 methyltransferase (prmA)
1710	transcriptional regulator, MerR family (skgA)
1714	acetyltransferase, GNAT family
1716	MutT/nudix family protein
1722	spermidine N1-acetyltransferase
1726	ATPase, AAA family
1736	ABC transporter domain protein
1738	Helix-turn-helix domain protein
1748	integrase, phage family
1756	Helix-turn-helix domain protein
1762	bacteriophage L54a, integrase
1768	LPXTG-motif cell wall anchor domain protein
1776	membrane protein
1778	conjugal transfer protein
1780	IS1381, transposase OrfA/OrfB, truncation
1802	transcriptional regulator (rstR-1)
1806	transcriptional regulator
1808	FtsK/SpoIIIE family protein
1814	aggregation substance

1818	mercuric reductase
1822	transcriptional regulator, MerR family
1824	Mn2+/Fe2+ transporter, NRAMP family
1830	ABC transporter, ATP-binding protein (eplF)
1848	Helix-turn-helix domain protein
1850	type 2 phosphatidic acid phosphatase(PAP2), family
1858	Abortive infection protein family
1868	aminotransferase, class-V
1874	glutathione reductase (gor)
1882	chorismate synthase (aroC)
1886	3-dehydroquinate synthase (aroB)
1900	sulfatase family protein
1914	ABC transporter, ATP-binding protein
1920	smf protein (Smffamily)
1924	transferrin receptor
1928	iron compound ABC transporter, ATP-binding protein
1932	iron compound ABC transporter, permease protein
1942	acetyltransferase, CysE/LacA/LpxA/NodL family
1952	GTP-binding protein
1958	carbon starvation protein A
1960	response regulator (lytR)
1962	GAF domain protein (lytS)
2000	extracellular protein
2004	diarrheal toxin (yukA)
2024	carbamoyl-phosphate synthase, large subunit (carB)
2028	carbamoyl-phosphate synthase, small subunit (carA)
2032	aspartate carbamoyltransferase (pyrB)
2036	dihydroorotase, multifunctional complex type (pyrC)
2040	orotate phosphoribosyltransferase (pyrE)
2048	membrane protein
2062	phosphate ABC transporter, permease protein (pstA-2)
2064	phosphate ABC transporter, ATP-binding protein (pstB)
2070	phosphate transport system regulatory protein PhoU
2072	aminopeptidase N (pepN)
2076	DNA-binding response regulator (arlR)
2080	sensor histidine kinase (arlS)
2088	signal recognition particle protein (ffh)
2102	peptide ABC transporter, peptide-binding protein
2104	integrase/recombinase, phage integrase family
2108	sensor histidine kinase
2112	DNA-binding response regulator (vicR)
2118	ABC transporter, ATP-binding protein
2122	nisin-resistance protein
2130	lipoprotein
2136	gid protein (gid)
2140	transcriptional regulator, GntR family
2142	GMP synthase (guaA)
2152	branched-chain amino acid ABC transporter, permease protein (livM)
2154	branched-chain amino acid ABC transporter, ATP-binding protein (livG)
2156	branched-chain amino acid ABC transporter, ATP-binding protein (livF)
2160	acetoin utilization protein AcuB
2174	DNA polymerase III, delta prime subunit (holB)
2186	copper homeostasis protein (cutC)
2190	phosphoserine aminotransferase (serC)
2202	methylated-DNA--protein-cysteine S-methyltransferase (ogt)
2208	exodeoxyribonuclease III (xth)
2214	PTS system, IIC component
2224	tellurite resistance protein TehB (tehB)
2246	icaA protein

2250	acetyltransferase, GNAT family
2258	oxidoreductase, short chain dehydrogenase/reductase family (fabG)
2266	oxidoreductase, Gfo/Idh/MocA family family
2268	glyoxalase family protein
2272	UDP-N-acetylglucosamine pyrophosphorylase (glmU)
2276	MutT/nudix family protein
2284	5-methylthioadenosine/S-adenosylhomocysteine nucleosidase (mtf)
2296	phosphatidate cytidylyltransferase (cdsA)
2300	membrane-associated zinc metalloprotease
2308	autolysin (flgJ)
2312	DNA polymerase III, alpha subunit, Gram-positive type
2320	nitroreductase family protein superfamily
2326	4-hydroxy-2-oxoglutarate aldolase/2-deydro-3-deoxyphosphogluconate aldo
2328	carbohydrate kinase, PfkB family
2336	oxidoreductase, short chain dehydrogenase/reductase family (fabG)
2338	PTS system, IIA component (manL)
2342	glucuronyl hydrolase
2346	PTS system, IIB component (manL)
2350	PTS system, IIC component (manM)
2364	sugar binding transcriptional regulator RegR (regR)
2368	polypeptide deformylase (def)
2380	oxidoreductase, Gfo/Idh/MocA family
2382	endopeptidase O (pepO)
2394	Na+/H+ antiporter
2404	transcriptional regulator
2410	replication initiation protein RepRC
2412	bacteriophage L54a, antirepressor
2416	e11
2422	replicative DNA helicase (dnaB)
2432	GTP-binding protein
2440	arpR protein
2444	gene 17 protein
2458	integrase/recombinase, phage integrase family
2468	bacteriophage L54a, phage D3 terminase
2472	protease
2500	PblB
2504	sensor histidine kinase
2514	N-acetylmuramoyl-L-alanine amidase
2518	KH domain protein
2522	ribosomal protein S16 (rpsP)
2526	permease
2528	ABC transporter, ATP-binding protein
2538	carbamoyl-phosphate synthase, large subunit
2540	carbamoyl-phosphate synthase, small subunit (carA)
2550	transcriptional regulator, LysR family
2554	ribosomal protein L27 (rpmA)
2562	ribosomal protein L21 (rplU)
2572	glycerophosphoryl diester phosphodiesterase
2582	nitroreductase family protein
2586	dipeptidase (pepV)
2614	GTP-binding protein HflX (hflX)
2618	galactose-1-phosphate uridylyltransferase (galT)
2626	oxidoreductase, short chain dehydrogenase/reductase family
2630	single-stranded-DNA-specific exonuclease RecJ (recJ)
2638	adenine phosphoribosyltransferase (apt)
2646	Bcl-2 family protein
2654	oxidoreductase, DadA family protein
2658	glucose-1-phosphate thymidylyltransferase (rfbA)
2664	dTDP-4-dehydrorhamnose 3,5-epimerase (rfbC)

2682	hyaluronidase
2686	mutator MutT protein (mutX)
2690	MutT/nudix family protein
2694	membrane protein
2702	acetolactate synthase (ilvK)
2706	adherence and virulence protein A (pavA)
2714	ABC transporter, permease protein (rbsC)
2722	metallo-beta-lactamase superfamily protein
2734	ribose 5-phosphate isomerase (rpiA)
2738	phosphopentomutase (deoB)
2742	purine nucleoside phosphorylase, family 2 (deoD)
2750	purine nucleoside phosphorylase (deoD)
2762	capsular polysaccharide biosynthesis protein Cps4A (cps4A)
2768	cpsb protein
2770	cpsc protein
2772	CpsE
2774	CpsF
2776	CpsVG
2778	CpsVH
2780	CpsVM
2782	CpsVN
2784	glycosyl transferase domain protein
2786	glycosyl transferase, family 2/glycosyl transferase family 8
2790	CpsVK
2794	CpsL
2796	neuB protein
2798	UDP-N-acetylglucosamine 2-epimerase
2800	hexapeptide transferase family protein
2802	NeuA
2808	uracil-DNA glycosylase (ung)
2818	DNA topoisomerase IV, B subunit (parE)
2822	DNA topoisomerase IV, A subunit (parC)
2826	branched-chain amino acid aminotransferase (ilvE)
2842	glycerol kinase (glpK)
2848	aerobic glycerol-3-phosphate dehydrogenase (glpD)
2874	ABC transporter, ATP-binding protein
2882	PTS system component (bglP)
2886	glutamate 5-kinase (proB)
2890	gamma-glutamyl phosphate reductase (proA)
2898	cell division protein FtsL (ftsL)
2904	penicillin-binding protein 2X (pbpX)
2910	phospho-N-acetylmuramoyl-pentapeptide-transferase (mraY)
2914	ATP-dependent RNA helicase, DEAD/DEAH box family (deaD)
2918	ABC transporter, substrate-binding protein
2924	amino acid ABC transporter, permease protein
2928	amino acid ABC transporter, ATP-binding protein
2932	thioredoxin reductase (trxB)
2940	NAD+ synthetase (nadE)
2944	aminopeptidase C (pepC)
2952	recombination protein U (recU)
2966	Uncharacterized protein family UPF0020 family
2974	autoinducer-2 production protein LuxS (luxS)
2978	KH domain protein
2986	ABC transporter, ATP-binding protein
2994	DNA-binding response regulator (vraR)
3000	guanylate kinase (gmk)
3004	DNA-directed RNA polymerase, omega subunit
3008	primosomal protein N (priA)
3012	methionyl-tRNA formyltransferase (fmt)

3016	Sun protein (sun)
3020	protein phosphatase 2C
3032	sensor histidine kinase
3034	DNA-binding response regulator (vraR)
3036	cof family protein/peptidyl-prolyl cis-trans isomerase, cyclophilin typ
3040	S1 RNA binding domain protein (rpsA)
3044	pyruvate formate-lyase-activating enzyme
3062	PTS system, IIB component (celA)
3066	PTS system, cellobiose-specific IIC component (celB)
3068	formate acetyltransferase (pfl)
3072	transaldolase
3080	cysteine synthase A (cysK)
3088	comF operon protein 1 (comFA)
3092	competence protein ComF
3096	ribosomal subunit interface protein (yfiA)
3104	tryptophanyl-tRNA synthetase (trpS)
3108	carbamate kinase (arcC)
3116	ornithine carbamoyltransferase (argF)
3124	arginine deiminase (arcA)
3134	transcriptional regulator, Crp/Fnr family
3138	inosine-5'-monophosphate dehydrogenase (guaB)
3140	MutR
3142	transporter
3146	recF protein (recF)
3158	peptidase, M16 family
3166	ABC transporter, ATP-binding protein
3170	ABC transporter, ATP-binding protein
3178	LysM domain protein (lytN)
3180	immunodominant antigen A (lsaA)
3184	L-serine dehydratase, iron-sulfur-dependent, alpha subunit (sdhA)
3188	L-serine dehydratase, iron-sulfur-dependent, beta subunit (sdhB)
3202	DHH subfamily 1 protein
3206	ribosomal protein L9 (rplI)
3210	replicative DNA helicase (dnaB)
3216	ribosomal protein S4 (rpsD)
3224	transcriptional regulator, TetR family
3236	membrane protein
3238	choline transporter (proWX)
3240	glycine betaine/L-proline transport ATP binding subunit (proV)
3242	DNA-binding response regulator
3244	Histidine kinase-, DNA gyrase B-, phytochrome-like ATPase family
3246	ornithine carbamoyltransferase (argF)
3248	carbamate kinase (arcC)
3252	membrane protein
3256	sensory box histidine kinase VicK
3258	DNA-binding response regulator
3268	Helix-turn-helix domain protein
3278	integrase
3284	ribosomal protein L33 (rpmG)
3288	ribosomal protein L32 (rpmF)
3300	YitT family protein
3304	YitT family protein
3320	DNA mismatch repair protein MutS (mutS)
3324	cold-shock domain family protein-related protein
3336	drug transporter
3340	Holliday junction DNA helicase RuvA (ruvA)
3352	recA protein (recA)
3386	oxidoreductase, Gfo/Idh/MocA family
3390	acetyltransferase, GNAT family

3394	anaerobic ribonucleoside-triphosphate reductase activating protein (nrd
3412	ABC transporter, permease protein (rbsC)
3414	ABC transporter, ATP-binding protein (nrtC)
3416	PTS system, mannose-specific IIAB components (manL)
3420	Cof family protein
3432	xanthine/uracil permease family protein
3440	acetyltransferase, GNAT family
3442	transcriptional regulator (cps4A)
3448	HIT family protein (hit)
3460	ABC transporter, permease protein
3472	Uncharacterized BCR, YhbC family COG0779 superfamily
3484	ribosomal protein L7A family
3496	esterase
3500	transcriptional repressor, CopY (copY)
3504	cation-transporting ATPase, E1-E2 family
3508	cation-binding protein-related protein
3520	DNA polymerase I (polA)
3534	DNA-binding response regulator (saeR)
3536	sensor histidine kinase (saeS)
3562	drug resistance transporter, EmrB/QacA subfamily
3566	peptidase M24 family protein
3570	peptidase M24 family protein (pepQ)
3572	cytidine/deoxycytidylate deaminase family protein
3584	translation elongation factor P (efp)
3592	N utilization substance protein B (nusB)
3596	sugar-binding transcriptional regulator, LacI family (scrR)
3600	sucrose-6-phosphate dehydrogenase (scrB)
3606	PTS system IIABC components (scrA)
3610	fructokinase (scrK)
3614	mannose-6-phosphate isomerase, class I (manA)
3622	phospho-2-dehydro-3-deoxyheptonate aldolase (aroH)
3626	holo-(acyl-carrier-protein) synthase (acpS)
3630	alanine racemase (alr)
3634	autolysin (usp45)
3636	ATP-dependent DNA helicase RecG (recG)
3642	shikimate 5-dehydrogenase (aroE)
3652	Cof family protein
3668	ferredoxin-related protein
3676	peptidase t (pepT)
3684	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase (mur
3692	iron compound ABC transporter, substrate-binding protein
3698	FecCD transport family protein (sirB)
3704	iron compound ABC transporter, permease protein (sirB)
3710	inorganic pyrophosphatase, manganese-dependent (ppaC)
3714	pyruvate formate-lyase-activating enzyme (pflA)
3718	CBS domain protein
3730	acid phosphatase
3736	LPXTG-motif cell wall anchor domain protein
3738	LPXTG-site transpeptidase family protein
3742	LPXTG-site transpeptidase family protein
3744	cell wall surface anchor family protein
3746	cell wall surface anchor family protein
3752	glycosyl transferase, group 1 family protein domain protein
3754	EpsQ protein
3756	polysaccharide extrusion protein
3768	dTDP-glucose 4-6-dehydratase
3782	glycosyl transferase domain protein
3788	dTDP-4-dehydrorhamnose reductase (rfbD)
3796	RNA polymerase sigma-70 factor (rpoD)

3802	DNA primase (dnaG)
3816	ABC transporter, ATP-binding protein Vexp2 (vex2)
3818	permease
3820	transmembrane protein Vexp3
3822	transmembrane protein Vexp3
3832	endopeptidase O (pepO)
3834	endopeptidase O (pepO)
3840	serine protease, subtilase family
3842	exotoxin 2
3844	CylK
3854	glycine cleavage system T protein
3856	CylE
3858	ABC transporter homolog CylB
3862	acyl carrier protein homolog AcpC (acpP)
3864	3-oxoacyl-(acyl-carrier-protein) reductase (fabG)
3868	CylD
3876	membrane protein
3912	LPXTG-site transpeptidase family protein
3916	LPXTG-site transpeptidase family protein
3918	LPXTG-site transpeptidase family protein
3920	LPXTG-motif cell wall anchor domain protein
3928	chaperonin, 33 kDa (hslO)
3932	Tn5252, Orf 10 protein
3934	transposase OrfAB, subunit B
3948	psr protein
3952	shikimate kinase (aroK)
3964	enolase (eno)
3972	MutT/nudix family protein
3976	glycosyl transferase, group 1
3978	preprotein translocase, SecA subunit (secA)
3986	preprotein translocase SecY family protein
3990	glycosyl transferase, family 8
3992	glycosyl transferase, family 2
3998	glycosyl transferase, family 8
4000	glycosyl transferase, family 2/glycosyl transferase family 8
4002	glycosyl transferase, family 8
4012	LPXTG-motif cell wall anchor domain protein (clfB)
4016	transcriptional regulator
4018	excinuclease ABC, B subunit (uvrB)
4022	Abortive infection protein family
4024	amino acid ABC transporter, amino acid-binding protein/permease protein
4026	amino acid ABC transporter, ATP-binding protein
4034	GTP-binding protein, GTP1/Obg family (obg)
4042	aminopeptidase PepS (pepS)
4050	ribosomal small subunit pseudouridine synthase A (rsuA)
4060	lactoylglutathione lyase (gloA)
4064	glycosyl transferase family protein
4072	alkylphosphonate utilization operon protein PhnA (phnA)
4078	glucosamine--fructose-6-phosphate aminotransferase (isomerizing) (glmS)
4090	Phosphofructokinase
4094	DNA polymerase III, alpha subunit (dnaE)
4098	transcriptional regulator, GntR family
4102	ABC transporter, ATP-binding protein
4106	ABC transporter, ATP-binding protein
4116	FtsK/SpoIIIE family protein
4122	Helix-turn-helix domain protein
4152	Helix-turn-helix domain protein
4158	excisionase
4160	transposase

4166	chloramphenicol acetyltransferase (cat)
4174	PilB-related protein
4178	acetyltransferase
4182	Leucine Rich Repeat domain protein
4190	nucleoside diphosphate kinase (ndk)
4206	Protein of unknown function superfamily
4218	hydrolase, haloacid dehalogenase-like family (pho2)
4226	oxygen-independent coproporphyrinogen III oxidase
4236	phosphoglucomutase/phosphomannomutase family protein (femD)
4240	Gram-positive signal peptide, YSIRK family domain protein
4256	cobyric acid synthase (cobQ)
4260	lipoate-protein ligase A (lplA)
4264	branched-chain alpha-keto acid dehydrogenase E3 component, lipoamide de
4266	pyruvate dehydrogenase complex, E2 component, dihydrolipoamide acetyltr
4270	pyruvate dehydrogenase complex, E1 component, pyruvate dehydrogenase be
4286	magnesium transporter, CorA family
4294	exonuclease RexB (rexB)
4302	phenylalanyl-tRNA synthetase, beta subunit (pheT)
4324	ATP synthase F1, epsilon subunit (atpC)
4328	ATP synthase F1, beta subunit (atpD)
4332	ATP synthase F1, gamma subunit (atpG)
4338	ATP synthase F1, alpha subunit (atpA)
4342	ATP synthase F1, delta subunit (atpH)
4346	ATP synthase F0, B subunit (atpF)
4350	ATP synthase, F0 subunit A (atpB)
4354	proton-translocating ATPase, c subunit-related protein
4360	glycogen synthase (glgA)
4362	glycogen biosynthesis protein GlgD (glgD)
4366	1,4-alpha-glucan branching enzyme (glgB)
4368	pullulanase
4382	ribonuclease BN
4396	acetyltransferase, GNAT family
4398	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (murA)
4402	thiamine-phosphate pyrophosphorylase (thiE)
4406	phosphomethylpyrimidine kinase (thiD)
4410	transcriptional regulator, Deg family (tenA)
4414	ABC transporter, ATP-binding protein
4426	S-adenosylmethionine synthetase (metK)
4440	DNA polymerase III, gamma and tau subunits (dnaX)
4444	GAF domain protein
4448	uridine kinase (udk)
4452	ATP-dependent RNA helicase, DEAD/DEAH box family
4458	peptidoglycan GlcNAc deacetylase (pgdA)
4462	glyceraldehyde-3-phosphate dehydrogenase, NADP-dependent (gapN)
4466	phosphoenolpyruvate-protein phosphotransferase (ptsI)
4470	phosphocarrier protein hpr
4474	NrdH-redoxin-related protein
4478	ribonucleoside-diphosphate reductase 2, alpha subunit (nrdE)
4498	glycosyl transferase, family 8
4504	alanyl-tRNA synthetase (alaS)
4512	alkyl hydroperoxide reductase, subunit F (ahpF)
4516	alkyl hydroperoxide reductase, subunit C (ahpC)
4520	ribosomal protein S2 (rpsB)
4524	translation elongation factor Ts (tsf)
4532	transcriptional regulator CtsR (ctsR)
4536	ATP-dependent Clp protease, ATP-binding subunit (clpC)
4540	deoxynucleoside kinase
4544	NifR3/Smm1 family protein
4548	chaperonin, 33 kDa (hslO)

4558	glutamate--cysteine ligase (gshA)
4562	Helix-turn-helix domain, fis-type protein
4566	perfringolysin O regulator protein (pfoR)
4570	adenylosuccinate synthetase (purA)
4578	SgaT protein (sgaT)
4582	PTS system, IIB component (sgaT)
4586	PTS system, IIA component (mtlA)
4590	hexulose-6-phosphate synthase
4594	hexulose-6-phosphate isomerase
4598	L-ribulose-5-phosphate 4-epimerase (araD)
4606	sugar binding transcriptional regulator RegR
4610	D-isomer specific 2-hydroxyacid dehydrogenase family protein (serA)
4622	transcriptional regulator, BglG family
4632	glycine betaine/L-proline transport ATP binding subunit (proV)
4636	amino acid ABC transporter, permease protein
4644	Na+/H+ exchanger family protein (kefB)
4648	glyoxylase family protein
4652	LPXTG-site transpeptidase family protein
4656	DNA gyrase, A subunit (gyrA)
4660	L-lactate dehydrogenase (ldh)
4664	NADH oxidase (nox)
4680	lipoprotein (bmpD)
4690	pantothenate kinase (coaA)
4694	ribosomal protein S20 (rpsT)
4698	amino acid ABC transporter, amino acid-binding protein (aatB)
4702	amino acid ABC transporter, ATP-binding protein
4726	ribosomal large subunit pseudouridine synthase B (rluB)
4734	Uncharacterized ACR, COG1354
4738	integrase/recombinase, phage integrase family (xerD)
4742	CBS domain protein
4746	phosphoesterase
4750	HAM1 protein
4768	transcriptional regulator, biotin repressor family
4792	amino acid ABC transproter, permease protein
4796	amino acid ABC transporter, substrate-binding protein
4798	6-aminohexanoate-cyclic-dimer hydrolase
4800	transcription elongation factor GreA (greA)
4804	Uncharacterized BCR, YceG family COG1559
4812	UDP-N-acetylmuramate--alanine ligase (murC)
4822	Snf2 family protein
4828	GTP-binding protein (b2511)
4832	primosomal protein DnaI (dnaI)
4844	sensor histidine kinase (arlS)
4846	DNA-binding response regulator (arlR)
4852	heat shock protein HtpX (htpX)
4870	potassium uptake protein, Trk family
4874	ABC transporter, ATP-binding protein
4888	phosphoglycerate kinase (pgk)
4896	transcriptional regulator, MerR family
4900	glutamine synthetase, type I (glnA)
4904	secreted 45 kd protein (usp45)
4908	metallo-beta-lactamase superfamily protein
4916	glycoprotease family protein
4926	glycoprotease family protein (gcp)
4938	ribosomal protein S14p/S29e (rpsN)
4952	exonuclease (dnaQ)
4956	transcriptional regulator, merR family
4958	cyclopropane-fatty-acyl-phospholipid synthase (cfa)
4970	1,4-dihydroxy-2-naphthoate octaprenyltransferase (menA)

4972	pyridine nucleotide-disulphide oxidoreductase (ndh)
4974	cytochrome d oxidase, subunit I (cydA)
4976	cytochrome d ubiquinol oxidase, subunit II (cydB)
4980	transport ATP-binding protein CydD
4988	polyprenyl synthetase (ispB)
4990	X-pro dipeptidyl-peptidase (pepX)
4998	drug transporter
5002	universal stress protein family
5004	glycerol uptake facilitator protein (glpF)
5012	cppA protein (cppA)
5034	exodeoxyribonuclease V, alpha subunit (recD)
5038	Signal peptidase I
5042	ribonuclease HIII (rnhC)
5062	transcriptional regulator
5068	maltose ABC transporter, permease protein (malD)
5072	maltose ABC transporter, permease protein (malC)
5088	ABC transporter, ATP-binding protein
5092	ABC transporter, permease protein
5106	spspoJ protein (spo0J)
5114	DNA polymerase III, beta subunit (dnaN)
5118	Diacylglycerol kinase catalytic domain (presumed) protein
5138	transcription-repair coupling factor (mfd)
5142	S4 domain protein
5156	MesJ/Ycf62 family protein
5160	hypoxanthine phosphoribosyltransferase (hpt)
5164	cell division protein FtsH (ftsH)
5172	hydrolase, haloacid dehalogenase-like family (b2690)
5178	transcriptional regulator, MarR family
5182	3-oxoacyl-(acyl-carrier-protein) synthase III (fabH)
5190	enoyl-(acyl-carrier-protein) reductase (fabK)
5194	malonyl CoA-acyl carrier protein transacylase (fabD)
5198	3-oxoacyl-[acyl-carrier protein] reductase (fabG)
5200	3-oxoacyl-(acyl-carrier-protein) synthase II (fabF)
5202	acetyl-CoA carboxylase, biotin carboxyl carrier protein (accB)
5206	(3R)-hydroxymyristoyl-(acyl-carrier-protein) dehydratase (fabZ)
5210	acetyl-CoA carboxylase, biotin carboxylase (accC)
5214	acetyl-CoA carboxylase, carboxyl transferase, beta subunit (accD)
5218	acetyl-CoA carboxylase, carboxyl transferase, alpha subunit (accA)
5224	seryl-tRNA synthetase (serS)
5234	PTS system, mannose-specific IID component
5246	ribosomal large subunit pseudouridine synthase, RluD subfamily (rluD)
5254	GTP pyrophosphokinase (relA)
5266	ribose-phosphate pyrophosphokinase (prsA)
5270	aminotransferase, class-V
5274	DNA-binding protein
5282	Domain of unknown function
5290	platelet activating factor
5296	transcriptional regulator, AraC family
5302	voltage-gated chloride channel family protein
5318	spermidine/putrescine ABC transporter, ATP-binding protein (potA)
5320	UDP-N-acetylenolpyruvoylglucosamine reductase (murB)
5324	bifunctional folate synthesis protein (folK)
5328	dihydroneopterin aldolase (folB)
5332	dihydropteroate synthase (folP)
5336	GTP cyclohydrolase I (folE)
5344	rarD protein (rarD)
5348	homoserine kinase (thrB)
5354	Polysaccharide deacetylase family (icaB)
5362	osmoprotectant transporter, BCCT family (opuD)

5384	thiol peroxidase (psaD)
5388	hydrolase
5390	transcriptional regulator, GntR family
5402	gls24 protein
5424	uncharacterized domain 1
5440	cation efflux family protein
5454	dihydroorotate dehydrogenase A (pyrDa)
5458	beta-lactam resistance factor (fibB)
5462	beta-lactam resistance factor (fibA)
5474	HD domain protein
5482	cation-transporting ATPase, E1-E2 family
5486	fructose-1,6-bisphosphatase (fbp)
5488	iron-sulfur cluster-binding protein
5492	peptide chain release factor 2 (prfB)
5496	cell division ABC transporter, ATP-binding protein FtsE (ftsE)
5504	carboxymethylenebutenolidase-related protein
5506	metallo-beta-lactamase superfamily protein
5514	DNA polymerase III, epsilon subunit/ATP-dependent helicase DinG
5520	asparaginyl-tRNA synthetase (asnS)
5526	inosine-uridine preferring nucleoside hydrolase (iunH)
5528	general stress protein 170
5534	Uncharacterised protein family superfamily
5538	Uncharacterized BCR, COG1481
5546	zinc ABC transporter, zinc-binding adhesion liprotein (adcA)
5560	isochorismatase family protein (entB)
5566	3-hydroxybutyryl-CoA dehydrogenase
5572	pyruvate phosphate dikinase (ppdK)
5574	glutamyl-tRNA(Gln) amidotransferase, C subunit (gatC)
5580	glutamyl-tRNA(Gln) amidotransferase, A subunit (gatA)
5594	GTP-binding protein
5612	iojap-related protein
5626	transcriptional regulator SkgA (skgA)
5630	glycerol uptake facilitator protein (glpF)
5634	dihydroxyacetone kinase family protein
5638	dihydroxyacetone kinase family protein
5640	transcriptional regulator, tetR family
5646	dihydroxyacetone kinase family protein
5654	glutamine amidotransferase, class I
5666	peptidase, M20/M25/M40 family
5668	ABC transporter, ATP-binding protein
5686	pur operon repressor (purR)
5690	cmp-binding-factor 1 (cbf1)
5694	competence-induced protein Ccs50 (ccs50)
5702	ribulose-phosphate 3-epimerase (rpe)
5710	rRNA (guanine-N1-)-methyltransferase (rrmA)
5712	dimethyladenosine transferase (ksgA)
5718	primase-related protein
5726	endosome-associated protein
5728	CG17785 gene product
5734	dltD protein (dltD)
5738	D-alanyl carrier protein-related protein
5742	dltB protein (dltB)
5754	DNA-binding response regulator (arlR)
5756	ribosomal protein L34 (rpmH)
5766	penicillin-binding protein 4 (pbp4)
5770	intein-containing protein
5774	NifU family protein
5778	aminotransferase, class-V
5782	Uncharacterized protein family (UPF0051) family

5786	ABC transporter, ATP-binding protein
5790	glycosyl transferase domain protein (llm)
5794	transcriptional regulator MecA (mecA)
5798	undecaprenol kinase
5806	amino acid ABC transporter, amino acid-binding protein/permease protein
5808	amino acid ABC transporter, ATP-binding protein
5834	riboflavin biosynthesis protein RibF (ribF)
5850	type I restriction-modification system, S subunit
5860	lipoprotein
5862	aggregation substance
5866	ID479
5896	type II DNA modification methyltransferase Spn5252IP (spn5252IMP)
5916	ribosomal protein L10 (rplJ)
5922	ATP-dependent Clp protease, ATP-binding subunit ClpC (clpC)
5926	homocysteine S-methyltransferase (mmuM)
5932	transcriptional regulator, TetR family
5938	GTP-binding protein (cgpA)
5952	thymidylate synthase (thyA)
5956	condensing enzyme, FabH-related
5960	hydroxymethylglutaryl-CoA reductase, degradative
5974	gene_idK21C13.21~pir||T04769~strong similarity to unknown protein, put
5976	FMN-dependent dehydrogenase family protein
5980	phosphomevalonate kinase
5986	diphosphomevalonate decarboxylase (mvaD)
5990	mevalonate kinase (mvk)
5994	Histidine kinase-, DNA gyrase B-, phytochrome-like ATPase family (PhoR1
6002	GTP pyrophosphokinase (relA)
6006	transposase for insertion sequence element is904
6016	5'-nucleotidase family
6018	polypeptide deformylase (def)
6022	NADP-specific glutamate dehydrogenase (gdhA)
6026	ABC transporter, ATP-binding/permease protein
6028	ABC transporter, ATP-binding/permease protein
6030	acetyltransferase, GNAT family family
6032	ABC transporter, ATP-binding protein
6040	degV family protein (degV)
6056	carbohydrate kinase, PfkB family (fruB)
6064	beta-lactam resistance factor (flbB)
6070	2-dehydropantoate 2-reductase
6076	PTS system component
6078	pyridine nucleotide-disulphide oxidoreductase family protein (trxB)
6082	tRNA (guanine-N1)-methyltransferase (trmD)
6092	c5a peptidase precursor
6100	ParA
6102	transposase family protein (orfA)
6116	Tn5252, relaxase
6120	Tn5252, Orf 10 protein
6124	mercuric reductase
6126	transcriptional regulator, MerR family
6132	cation transport ATPase, E1-E2 family
6138	cation-transporting ATPase, E1-E2 family
6140	cation-transporting ATPase, E1-E2 family
6144	cation-transporting ATPase, E1-E2 family
6146	transcriptional repressor, CopY (copY)
6150	cadmium resistance transporter
6158	membrane protein
6162	flavoprotein (dfp)
6170	lipoate-protein ligase A
6174	FMN oxidoreductase (nemA)

6178	Bacterial luciferase superfamily
6182	glycine cleavage system H protein (gcvH)
6186	Domain of unknown function
6194	lipoate-protein ligase A (lplA)
6198	formate--tetrahydrofolate ligase (fhs)
6202	cardiolipin synthetase (cls)
6220	aminotransferase, class II (aspB)
6222	RNA methyltransferase, TrmH family, group 2
6232	60 kda chaperonin
6242	purine nucleoside phosphorylase (deoD)
6248	deoxyribose-phosphate aldolase (deoC)
6254	Lyme disease proteins of unknown function
6258	ribosomal large subunit pseudouridine synthase, RluD subfamily (rluD)
6262	penicillin-binding protein 2A (pbp2A)
6266	pathenogenicity protein
6268	transcription antitermination protein NusG (nusG)
6272	glycosyl transferase, family 8
6276	glycosyl transferase, family 8
6284	sugar transporter family protein
6292	sensory box histidine kinase
6306	homocysteine S-methyltransferase (metH)
6310	glycerol dehydrogenase
6312	DNA topology modulation protein FlaR
6316	translation initiation factor IF-1 (infA)
6320	adenylate kinase (adk)
6326	ribosomal protein L15 (rplO)
6330	ribosomal protein L30 (rpmD)
6336	ribosomal protein S5 (rpsE)
6344	ribosomal protein L6 (rplF)
6348	ribosomal protein S8 (rpsH)
6352	ribosomal protein S14 (rpsN)
6356	ribosomal protein L5 (rplE)
6360	ribosomal protein L24 (rplX)
6366	ribosomal protein L14 (rplN)
6368	ribosomal protein S17 (rpsQ)
6372	ribosomal protein L29 (rpmC)
6374	ribosomal protein L16 (rplP)
6378	ribosomal protein S3 (rpsC)
6382	ribosomal protein L22 (rplV)
6386	ribosomal protein S19 (rpsS)
6390	ribosomal protein L2 (rplB)
6394	ribosomal protein L23 (rplW)
6398	ribosomal protein L4/L1 family (rplD)
6402	ribosomal protein L3 (rplC)
6408	ribosomal protein S10 (rpsJ)
6414	MATE efflux family protein
6418	threonine synthase (thrC)
6428	Uncharacterized BCR, COG1636 superfamily
6436	4-alpha-glucanotransferase (malQ)
6440	glycogen phosphorylase family protein (malP)
6444	glycerol-3-phosphate transporter (glpT)
6452	rhodanese family protein
6458	ammonium transporter
6464	DNA repair protein RadA (radA)
6472	oxidoreductase, pyridine nucleotide-disulfide, class I
6478	ribose ABC transporter, periplasmic D-ribose-binding protein (rbsB)
6484	ribose ABC transporter, ATP-binding protein (rbsA)
6486	ribose ABC transporter protein (rbsD)
6488	ribokinase (rbsK)

6498	ABC transporter, ATP-binding protein
6502	DNA-binding response regulator (vicR)
6506	argininosuccinate synthase (argG)
6508	argininosuccinate lyase (argH)
6514	bacteriophage L54a, repressor protein
6528	soluble transducer HtrXIII
6542	probable transposase (insertion sequence IS861)
6544	ABC transporter, ATP-binding/permease protein
6550	ABC transporter, ATP-binding/permease protein
6560	Serine hydroxymethyltransferase
6568	HemK protein (hemK)
6572	peptide chain release factor 1 (prfA)
6576	thymidine kinases
6580	4-oxalocrotonate tautomerase (dmpI)
6588	oxidoreductase
6594	oxidoreductase
6600	formate/nitrite transporter family protein
6608	xanthine permease (pbuX)
6612	xanthine phosphoribosyltransferase (xpt)
6616	guanosine monophosphate reductase (guaC)
6620	drug resistance transporter, EmrB/QacA subfamily
6622	oxidoreductase
6624	Kup system potassium uptake protein (kup)
6636	O-methyltransferase
6642	oligoendopeptidase F (pepF)
6646	competence protein ComA (comA)
6650	major facilitator superfamily protein superfamily
6652	ribosomal small subunit pseudouridine synthase A (rsuA)
6658	glucosamine-6-phosphate isomerase (nagB)
6662	nodulin-related protein, truncation
6664	S-adenosylmethioninetRNA ribosyltransferase-isomerase (queA)
6674	permease, GntP family
6684	6-phospho-beta-glucosidase (bglA)
6686	PTS system, beta-glucosides-specific IIABC components
6688	transcription antiterminator LicT (licT)
6704	esterase
6706	sugar-binding transcriptional repressor, LacI family
6708	hydrolase, haloacid dehalogenase-like family
6712	DNA internalization-related competence protein ComEC/Rec2
6716	competence protein CelA (celA)
6720	acyltransferase family protein
6732	ATP-dependent RNA helicase DeaD (deaD)
6736	lipoprotein, YaeC family
6738	ABC transporter, permease protein
6752	diacylglycerol kinase (dgkA)
6768	formamidopyrimidine-DNA glycosylase (mutM)
6776	epidermin immunity protein F
6788	glycyl-tRNA synthetase, beta subunit (glyS)
6790	acyl carrier protein phosphodiesterase
6800	SsrA-binding protein (smpB)
6822	D-alanine--D-alanine ligase
6824	recombination protein RecR (recR)
6830	penicillin-binding protein 2b
6832	phosphoglycerate mutase (gpmA)
6836	triosephosphate isomerase (tpiA)
6856	phosphoglycerate mutase family protein
6860	D-alanyl-D-alanine carboxypeptidase family
6864	autolysin
6868	heat-inducible transcription repressor HrcA (hrcA)

6872	heat shock protein GrpE (grpE)
6876	chaperone protein dnak
6880	dnaJ protein (dnaJ)
6884	transcriptional regulator, gntR family domain protein
6888	tRNA pseudouridine synthase A (truA)
6892	phosphomethylpyrimidine kinase (thiD)
6910	galactose-6-phosphate isomerase, LacA subunit (lacA)
6922	tagatose 1,6-diphosphate aldolase (lacD)
6932	sugar ABC transporter, ATP-binding protein (msmK)
6936	glucan 1,6-alpha-glucosidase (dexB)
6940	UDP-glucose 4-epimerase (galE)
6942	response regulator (citB)
6950	citrate carrier protein (citS)
6954	malate oxidoreductase (tme)
6958	bacterocin transport accessory protein
6976	transposase family protein (orfA)
6980	pXO1-128
6986	adhesion lipoprotein (lmb)
6994	DNA-directed RNA polymerase, alpha subunit (rpoA)
6998	ribosomal protein L17 (rplQ)
7040	probable dna-directed rna polymerase delta subunit
7044	CTP synthase (pyrG)
7058	bacteriocin transport accessory protein
7074	translation initiation factor IF-3 (infC)
7100	adenosine deaminase
8468	preprotein translocase, SecE subunit
8476	antigen, 67 kDa
8486	Lipase/Acylhydrolase
8492	peptide ABC transporter, permease protein (oppB)
8494	competence protein CglB (cglB)
8502	peptide ABC transporter, peptide-binding protein
8504	oxidoreductase
8510	amino acid ABC transporter, permease protein (opuBB)
8522	abc transporter atp-binding protein ybhf
8530	glycerol-3-phosphate dehydrogenase (NAD(P)+) (gpsA)
8538	sugar ABC transporter, sugar-binding protein
8544	secreted 45 kd protein (usp45)
8556	phosphoglycerate mutase family protein
8566	glycosyl hydrolase, family 3
8576	N-acetylmuramoyl-L-alanine amidase
8596	sensory box histidine kinase (withHAMPandPASd)
8608	aminoglycoside 6-adenylyltransferase
8622	iron compound ABC transporter, permease protein (sirB)
8636	phosphate ABC transporter, permease protein (pstC-2)
8650	branched-chain amino acid transport system II carrier protein (brnQ)
8658	PTS system, IID component
8662	replisome organiser-related protein
8674	alkaline amylopullulanase
8676	exfoliative toxin A
8690	glycerol uptake facilitator protein (glpF)
8698	ABC transporter, ATP-binding protein
8706	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (pgs
8708	cobalt transport protein
8730	integral membrane protein
8734	yadS protein
8736	cell wall surface anchor family protein
8748	polysaccharide biosynthesis protein
8752	glycosyl transferase domain protein
8764	endopeptidase O

8770	beta-ketoacyl-acyl carrier protein synthase II
8772	ABC transporter, ATP-binding protein
8776	penicillin-binding protein
8778	cell wall surface anchor family protein
8780	cell wall surface anchor family protein
8786	LPXTG-motif cell wall anchor domain protein
8788	6-aminohexanoate-cyclic-dimer hydrolase
8796	NLP/P60 family protein
8802	DNA/RNA non-specific endonuclease
8806	hydroxyethylthiazole kinase (thiM)
8826	PTS system component
8832	sugar ABC transporter, permease protein
8836	potassium uptake protein, Trk family (trkA)
8850	lemA protein (lemA)
8856	cobalt transport protein
8882	spermidine/putrescine ABC transporter, spermidine/putrescine-binding pr
8884	spermidine/putrescine ABC transporter, permease protein (potC)
8906	ABC transporter, substrate-binding protein
8908	lipoprotein
8916	sensor histidine kinase
8930	TrsK-like protein (traK)
8936	R5 protein
8962	chromosome assembly protein homolog
8978	ribose ABC transporter, permease protein (rbsC)
8980	permease
8982	sensor histidine kinase (arlS)
8986	hydrolase, haloacid dehalogenase-like family (gph)
8994	dephospho-CoA kinase
8996	oxalateformate antiporter
9004	sensory box protein
9006	host cell surface-exposed lipoprotein
9012	PAP2 family protein
9034	GtrA family protein
9050	lipoprotein signal peptidase (lspA)
9280	alcohol dehydrogenase, zinc-containing (adh)
9284	trigger factor (tig)
9290	fructose-bisphosphate aldolase (fba)
9292	DAK2 domain protein
9296	oligopeptide ABC transporter, permease protein
9298	N-acetylglucosamine-6-phosphate deacetylase (nagA)
9300	transcriptional regulator, DeoR family (lacR)
9302	PTS system, mannose-specific IIC component (manM)
9306	Phosphoglucose isomerase
9310	aspartate--ammonia ligase (asnA)
9312	amino acid ABC transporter, ATP-binding protein
9314	DNA-binding protein HU (hup)
9316	DHH subfamily 1 protein
9318	chloride channel
9320	integrase (int)
9324	DNA/RNA non-specific endonuclease
9326	PTS system component
9328	cell division protein, FtsW/RodA/SpoVE family (ftsW)
9330	LPXTG-motif cell wall anchor domain protein
9332	peptide chain release factor 3 (prfC)
9334	ABC transporter, ATP-binding protein
9336	superoxide dismutase [mn-fe]
9340	phenylalanyl-tRNA synthetase, alpha subunit (pheS)
9342	amino acid ABC transporter, permease protein
9344	phosphate ABC transporter, phosphate-binding protein (pstS)

9346	NOL1/NOP2/sun family protein (sun)
9348	Abortive infection protein family
9350	permease
9352	N-acetylmuramoyl-L-alanine amidase domain protein (usp45)
9354	ABC transporter, ATP-binding protein
9356	phosphoglucomutase (pgm)
9358	oxidoreductase, short chain dehydrogenase/reductase family
9360	phosphate acetyltransferase
9362	gls24 protein
9364	ribosomal protein S1 (rpsA)
9368	dTDP-glucose 4,6-dehydratase (rfbB)
9370	excinuclease ABC, C subunit (uvrC)
9372	MATE efflux family protein
9378	amino acid permease (rocE)
9380	DNA-binding response regulator TrcR (trcR)
9382	16S rRNA processing protein RimM (rimM)
9384	transcriptional regulator
9388	ribosomal protein L20 (rplT)
9394	sugar-binding transcriptional repressor, LacI family (malR)
9396	proton/peptide symporter family protein
9398	amino acid permease
9400	exoribonuclease, VacB/Rnb family (vacB)
9402	multi-drug resistance efflux pump (pmrA)
9404	adhesion lipoprotein (psaA)
9406	iron-dependent transcriptional regulator (sirR)
9410	branched-chain amino acid ABC transporter, amino acid-binding protein (
9412	amino acid permease
9414	SpoU rRNA Methylase family protein
9416	sodium/dicarboxylate symporter (gltP-2)
9418	branched-chain amino acid transport system II carrier protein (brnQ)
9420	alcohol dehydrogenase, zinc-containing
9422	aminotransferase, class I (aspB)
9424	ribosomal protein S6 (rpsF)
9426	A/G-specific adenine glycosylase (mutY)
9428	acid phosphatase (olpA)
9430	ribosomal protein S12 (rpsL)
9434	microcin immunity protein MccF (mccF-1)
9436	undecaprenyl diphosphate synthase (uppS)
9438	preprotein translocase, YajC subunit (yajC)
9440	chaperonin, 10 kDa (groES)
9444	YitT family protein
9446	serine protease (htrA)
9448	ribose-phosphate pyrophosphokinase (prsA)
9450	aromatic amino acid aminotransferase (araT)
9452	Recombination protein O (recO)
9454	Abortive infection protein family
9456	fatty acid/phospholipid synthesis protein PlsX (plsX)
9458	acyl carrier protein (acpP)
9462	phosphoribosylaminoimidazole carboxylase, ATPase subunit (purK)
9464	alcohol dehydrogenase, iron-containing
9466	ribosomal protein L18 (rplR)
9468	preprotein translocase, SecY subunit
9470	transcriptional regulator ComX1 (comX1)
9472	deoxyuridine 5`-triphosphate nucleotidohydrolase (dut)
9478	sugar-binding transcriptional regulator, LacI family (rbsR)
9480	SPFH domain/Band 7 family
9488	zinc ABC transporter, permease protein (adcB)
9492	abortive infection protein
9494	hydrolase, haloacid dehalogenase-like family

9496	response regulator (lytT)
9500	transketolase, C-terminal subunit
9502	polyribonucleotide nucleotidyltransferase (pnp)
9504	serine O-acetyltransferase (cysE)
9508	ribosomal protein L13 (rplM)
9510	replication initiation protein
9518	amino acid ABC transporter, amino acid-binding protein
9522	glycyl-tRNA synthetase, alpha subunit (glyQ)
9524	NADH oxidase
9528	transketolase (tkt)
9534	penicillin-binding protein 1A (pbp1A)
9536	cell division protein DivIVA (divIVA)
9538	sensor histidine kinase
9540	serine/threonine protein kinase (pknB)
9542	transcriptional regulator
9544	PTS system, IIA component (lacF)
9546	glycerol dehydrogenase (gldA)
9548	aspartate kinase (thrA)
9550	enoyl-CoA hydratase/isomerase family protein
9552	acyl carrier protein (acpP)
9564	ABC transporter, ATP-binding protein
9566	N utilization substance protein A (nusA)
9568	ribosome-binding factor A (rbfA)
9570	Cof family protein
9572	CoA binding domain protein (b0965)
9574	transcriptional regulator, Fur family
9578	queuine tRNA-ribosyltransferase (tgt)
9580	ribonuclease P protein component (rnpA)
9582	serine protease, subtilase family
9584	glycosyl transferase domain protein
9586	transcriptional activator, AraC family
9588	transcriptional regulator, TetR family
9590	transcriptional regulator, AraC family
9594	surface protein Rib
9596	transposase, mutator family
9600	acetyltransferase, GNAT family
9602	Transposase, Mutator family
9606	UDP-sugar hydrolase
9610	anthranilate synthase component II (trpG)
9612	biotin synthetase (bioB)
9616	UDP-N-acetylmuramoylalanine--D-glutamate ligase (murD)
9618	ylmF protein (ylmF)
9620	amino acid ABC transporter, permease protein
9622	phosphoglucomutase (pgm)
9624	YjeF-related protein, C-terminus
9626	FemAB family protein (fibA)
9628	Cof family protein
9630	cell division ABC transporter, permease protein FtsX (ftsX)
9632	oxidoreductase, short-chain dehydrogenase/reductase family (fabG)
9634	aspartate aminotransferase (aspC)
9638	ribosomal protein L31 (rpmE)
9640	nrdI protein (nrdI)
9642	ribosomal protein L19 (rplS)
9644	bacteriophage L54a, repressor protein
9646	bacteriophage L54a, antirepressor
9652	single-strand binding protein (ssb)
9660	pneumococcal surface protein A
9666	DNA-binding response regulator (vncR)
9668	transposase OrfAB, subunit B

9670	cell division protein, FtsW/RodA/SpoVE family (rodA)
9672	DNA gyrase, B subunit (gyrB)
9674	3-phosphoshikimate 1-carboxyvinyltransferase (aroA)
9676	RNA methyltransferase, TrmA family
9680	transcriptional regulator, AraC family
9682	ABC transporter, ATP-binding protein
9690	CylJ
9696	permease
9698	regulatory protein
9700	carbohydrate kinase, pfkB family
9702	beta-glucuronidase
9704	2-deydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldo
9706	3-oxoacyl-(acyl-carrier-protein) reductase
9708	catabolite control protein A (ccpA)
9712	ribonuclease III (rnc)
9714	SMC family, C-terminal domain family
9718	S1 RNA binding domain protein
9722	prolipoprotein diacylglyceryl transferase (lgt)
9724	riboflavin synthase, alpha subunit (ribE)
9726	3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II (ri
9728	lysyl-tRNA synthetase (lysS)
9734	Transposase subfamily
9738	translation elongation factor Tu (tuf)
9740	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-a
9746	Glutathione S-transferases domain protein
9754	Ribonucleotide reductases
9756	biotin--acetyl-CoA-carboxylase ligase
9760	Uncharacterized protein family SNZ family
9762	methionine aminopeptidase, type I (map)
9764	DNA ligase, NAD-dependent (ligA)
9766	glucose-1-phosphate adenylyltransferase (glgC)
9768	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (murA)
9770	acetyltransferase, GNAT family
9772	exonuclease RexA (rexA)
9774	tRNA modification GTPase TrmE (trmE)
9776	ABC transporter, ATP-binding protein
9778	pyruvate dehydrogenase complex, E1 component, pyruvate dehydrogenase al
9782	Mur ligase family protein
9786	HD domain protein
9788	translation elongation factor G (fusA)
9796	pyruvate kinase (pyk)
9798	Signal peptidase I
9802	cytidine deaminase (cdd)
9804	sugar ABC transporter, ATP-binding protein
9806	sugar ABC transporter, permease protein
9808	acetyltransferase, GNAT family
9810	ABC transporter, permease protein
9812	SatD
9814	Helix-turn-helix domain, fis-type protein
9816	phosphate ABC transporter, ATP-binding protein (pstB-1)
9818	tRNA pseudouridine synthase B (truB)
9820	Acetyltransferase (GNAT) family
9822	DNA topoisomerase I (topA)
9824	ribonuclease HII (rnhB)
9830	orotidine 5'-phosphate decarboxylase (pyrF)
9832	aspartate-semialdehyde dehydrogenase (asd)
9836	pantothenate metabolism flavoprotein (dfp)
9840	Sua5/YciO/YrdC/YwlC family protein
9844	thiamine biosynthesis protein ApbE

9846	Domain of unknown function
9848	DNA repair protein RadC (radC)
9850	glycosyl hydrolase, family 1 (bglA)
9852	Cof family protein (b0844)
9854	spermidine/putrescine ABC transporter, permease protein (potH)
9856	folylpolyglutamate synthase (folC)
9858	homoserine dehydrogenase (hom)
9860	succinate-semialdehyde dehydrogenase (gabD-1)
9862	membrane protein
9864	ATP-dependent DNA helicase PcrA (pcrA)
9866	uracil permease (uraA)
9868	sodiumalanine symporter family protein
9878	capsular polysaccharide biosynthesis protein Cps4B (cps4B)
9880	transcriptional regulator, LysR family
9882	CpslaS
9884	chloride channel protein
9886	tributyrin esterase (estA)
9888	ABC transporter, ATP-binding protein (potA)
9890	alpha-acetolactate decarboxylase (budA)
9892	TPR domain protein
9896	metallo-beta-lactamase superfamily protein
9898	tRNA delta(2)-isopentenylpyrophosphate transferase (miaA)
9902	glycerophosphoryl diester phosphodiesterase
9904	transposase OrfAB, subunit B
9906	IS3-Spn1, transposase
9908	transposase OrfAB, subunit B (orfB)
9910	reverse transcriptase
9916	transposase OrfAB, subunit B
9918	integrase, phage family (int)
9920	transcription regulator
9922	TnpA
9926	structural gene for ultraviolet resistance (uvra)
9930	Helicases conserved C-terminal domain protein
9932	abortive infection bacteriophage resistance protein (abiEi)
9944	ribosomal protein L7/L12 (rplL)
9948	ATP-dependent Clp protease, ATP-binding subunit ClpX (clpX)
9950	dihydrofolate reductase (folA)
9952	hemolysin
9954	transcriptional regulator, MarR family
9958	polyA polymerase family protein
9960	PTS system, fructose specific IIABC components (fruA-1)
9962	lactose phosphotransferase system repressor (lacR)
9964	choline binding protein D (cbpD)
9968	pyrimidine operon regulatory protein (pyrR)
9970	ribosomal large subunit pseudouridine synthase D (rluD)
9972	thiamine biosynthesis protein ThiI (thiI)
9974	3-dehydroquinate dehydratase, type I (aroD)
9976	iron compound ABC transporter, ATP-binding protein (fepC)
9980	transcriptional regulator
9982	glycosyl transferase domain protein
9984	Cps9H
9988	4-diphosphocytidyl-2C-methyl-D-erythritol synthase (ispD)
9990	licD1 protein (licD1)
9996	large conductance mechanosensitive channel protein (mscL)
10000	maltose ABC transporter, maltose-binding protein
10004	nucleotide sugar synthetase-like protein
10006	transcriptional regulator
10008	oxidoreductase, aldo/keto reductase family
10010	NAD(P)H-flavin oxidoreductase

10016	transcriptional regulator MutR
10018	GTP-binding protein Era (era)
10022	peptide methionine sulfoxide reductase (msrA)
10026	peptide ABC transporter, ATP-binding protein
10028	peptide ABC transporter, ATP-binding protein (amiE)
10030	peptide ABC transporter, peptide-binding protein
10032	transposase, IS30 family
10034	transcriptional regulator, LysR family
10036	spoE family protein (ftsK)
10044	methionyl-tRNA synthetase (metG)
10046	D-isomer specific 2-hydroxyacid dehydrogenase family protein (serA)
10048	acetyltransferase, GNAT family
10050	phosphoserine aminotransferase (serC)
10054	thymidylate kinase (tmk)
10060	branched-chain amino acid ABC transporter, permease protein (livH)
10062	ATP-dependent Clp protease, proteolytic subunit ClpP (clpP)
10064	uracil phosphoribosyltransferase (upp)
10066	potassium uptake protein, Trk family (trkH)
10068	glutamate racemase (murI)
10070	membrane protein
10072	HD domain protein
10074	Acylphosphatase
10076	spoIIIJ family protein
10078	acetyltransferase, GNAT family
10080	glucose-inhibited division protein B (gidB)
10082	potassium uptake protein, Trk family
10084	ABC transporter, permease protein
10088	isochorismatase family protein
10092	haloacid dehalogenase-like hydrolase superfamily
10094	membrane protein
10096	glutamyl-tRNA(Gln) amidotransferase, B subunit (gatB)
10098	CBS domain protein protein
10100	transcriptional regulator (codY)
10102	universal stress protein family
10104	L-asparaginase (ansA)
10106	oxidoreductase, aldo/keto reductase 2 family
10108	preprotein translocase, SecA subunit (secA)
10112	excinuclease ABC, A subunit (uvrA)
10114	magnesium transporter, CorA family (corA)
10116	thioredoxin (trx)
10118	MutS2 family protein (mutS2)
10122	DNA-damage inducible protein P (dinP)
10124	formate acetyltransferase (pfl)
10126	transcriptional regulator, Crp family
10128	transport ATP-binding protein CydC
10138	ribosomal-protein-alanine acetyltransferase (rimI)
10140	hydrolase
10144	D-alanine-activating enzyme (dltA)
10148	carbohydrate kinase, FGGY family
10150	transaldolase
10160	Helix-turn-helix domain protein
10164	single-strand binding protein (ssb)
10166	type II DNA modification methyltransferase Spn5252IP (spn5252IMP)
10174	integrase, phage family
10178	Cyclic nucleotide-binding domain protein
10180	transcriptional regulator, MarR family
10182	prolyl-tRNA synthetase (proS)
10184	leucine-rich protein
10186	lacX protein, truncation (lacX)

10188	tagatose-6-phosphate kinase (lacC)
10190	galactose-6-phosphate isomerase, LacB subunit (lacB)
10192	neuraminidase
10198	Histidine kinase-, DNA gyrase B-, phytochrome-like ATPase domain protei
10200	ABC transporter, ATP-binding protein
10202	PTS system, IIABC components (ptsG)
10204	phosphate regulon response regulator PhoB (phoB)
10212	Uncharacterized ACR, COG2161 subfamily
10216	abortive phage resistance protein
10222	TnpA
10226	acetyltransferase, GNAT family
10230	ABC transporter domain protein
10234	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (
10236	branched-chain amino acid transport protein AzlC (azlC)
10240	DNA-binding response regulator (srrA)
10242	leucyl-tRNA synthetase (leuS)
10246	NupC family protein
10248	transcriptional regulator, GntR family
10252	glyoxalase family protein
10254	anaerobic ribonucleoside-triphosphate reductase (nrdD)
10256	competence-induced protein Ccs4
10262	competence/damage-inducible protein CinA (cinA)
10264	DNA-3-methyladenine glycosylase I (tag)
10268	DNA mismatch repair protein HexB (hexB)
10270	arginine repressor (argR)
10272	arginyl-tRNA synthetase (argS)
10274	aspartyl-tRNA synthetase (aspS)
10276	histidyl-tRNA synthetase (hisS)
10280	AGR_pAT_51p
10286	hydrolase, alpha/beta hydrolase fold family
10288	phage infection protein
10290	Glucose inhibited division protein A (gidA)
10292	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase (trmU)
10296	arginine/ornithine antiporter (arcD)
10298	chromosomal replication initiator protein DnaA (dnaA)
10302	peptidyl-tRNA hydrolase (pth)
10310	phosphotyrosine protein phosphatase
10316	ribosomal protein L36 (rpmJ)
10318	ribosomal protein S13/S18 (rpsM)
10328	L-lactate dehydrogenase (ldh)
10330	ribosomal protein L28 (rpmB)
10362	RNA polymerase sigma-70 factor, ECF subfamily
10384	BioY family protein
10386	AtsA/ElaC family protein
10388	cytidine/deoxycytidylate deaminase family protein
10394	phosphorylase, Pnp/Udp family
10396	transcriptional regulator, MerR family
10402	methyltransferase (ubiE)
10412	type IV prepilin peptidase
10416	ylmG protein (ylmG)
10444	transposase OrfAB, subunit B
10446	IS150-like transposase
10452	Bacterial regulatory proteins, tetR family domain protein
10454	cell wall surface anchor family protein, authentic frameshift (clfB)
10456	transposase OrfAB, subunit A (orfA)
10460	chaperonin, 33 kDa (hslO)
10472	(3R)-hydroxymyristoyl-(acyl-carrier-protein) dehydratase (fabZ)
10482	sprT protein
10490	transcriptional regulator, MarR family

10498	transcriptional regulator
10504	glycogen biosynthesis protein GlgD (glgD)
10536	ribonucleoside-diphosphate reductase, alpha subunit, truncation (nrdD)
10538	LPXTG-motif cell wall anchor domain
10550	membrane protein
10554	arsenate reductase (arsC)
10564	transposase, authentic frameshift
10570	transposase OrfAB, subunit A (orfA)
10574	Tn5252, Orf 9 protein
10580	IS3-Spn1, transposase
10584	transcriptional regulator, ArsR family
10628	ribosomal protein L35 (rpmI)
10630	cytidylate kinase (cmk)
10636	MutT/nudix family protein
10644	preprotein translocase, SecG subunit
10680	ribosomal protein S18 (rpsR)
10682	single-strand binding protein (ssb)
10692	glyceraldehyde 3-phosphate dehydrogenase (gap)
10694	translation elongation factor G (fusA)
10696	ribosomal protein S7 (rpsG)
10704	phosphinothricin N-acetyltransferase (pat)
10730	nrdI protein (nrdI)
10732	accessory gene regulator protein C (blpH)
10744	rhodanese family protein (pspE)
10746	cAMP factor
10758	competence/damage-inducible protein CinA (cinA)
10770	transcriptional regulator, ArgR family (argR)
10772	FliP family family
10794	peptide ABC transporter, peptide-binding protein
10800	ribosomal protein S21 (rpsU)
10802	transposase, IS30 family
10816	mucin 2 precursor, intestinal
10854	SV40-transformed marker protein pG1-related protein
10856	SV40-transformed marker protein pG1-related protein
10858	SV40-transformed marker protein pG1-related protein
10860	SV40-transformed marker protein pG1-related protein
10862	SV40-transformed marker protein pG1-related protein
10864	SV40-transformed marker protein pG1-related protein
10866	SV40-transformed marker protein pG1-related protein
10910	transcriptional regulator
10920	ribosomal protein S11 (rpsK)
10922	elaA protein
10926	5-formyltetrahydrofolate cyclo-ligase family protein
10938	inositol monophosphatase family protein
10940	amino acid ABC transporter, amino acid-binding protein (artI)
10944	Holliday junction DNA helicase RuvB (ruvB)
10946	D-alanyl-D-alanine carboxypeptidase (dacA)
10948	lipoprotein (bmpD)
10950	peptidase, U32 family family
10952	protease maturation protein
10954	glutamyl-tRNA synthetase (gltX)
10956	GTP-binding protein LepA (lepA)
10960	translation initiation factor if-2
10962	phosphoenolpyruvate carboxylase (ppc)
10964	calcium E1-E2-type ATPase
10966	serine protease, subtilase family

CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, 892, 894, 896, 898, 900, 902, 904, 906, 908, 910, 912, 914, 916, 918, 920, 922, 924, 926, 928, 930, 932, 934, 936, 938, 940, 942, 944, 946, 948, 950, 952, 954, 956, 958, 960, 962, 964, 966, 968, 970, 972, 974, 976, 978, 980, 982, 984, 986, 988, 990, 992, 994, 996, 998, 1000, 1002, 1004, 1006, 1008, 1010, 1012, 1014, 1016, 1018, 1020, 1022, 1024, 1026, 1028, 1030, 1032, 1034, 1036, 1038, 1040, 1042, 1044, 1046, 1048, 1050, 1052, 1054, 1056, 1058, 1060, 1062, 1064, 1066, 1068, 1070, 1072, 1074, 1076, 1078, 1080, 1082, 1084, 1086, 1088, 1090, 1092, 1094, 1096, 1098, 1100, 1102, 1104, 1106, 1108, 1110, 1112, 1114, 1116, 1118, 1120, 1122, 1124, 1126, 1128, 1130, 1132, 1134, 1136, 1138, 1140, 1142, 1144, 1146, 1148, 1150, 1152, 1154, 1156, 1158, 1160, 1162, 1164, 1166, 1168, 1170, 1172, 1174, 1176, 1178, 1180, 1182, 1184, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1222, 1224, 1226, 1228, 1230, 1232, 1234, 1236, 1238, 1240, 1242, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, 1272, 1274, 1276, 1278, 1280, 1282, 1284, 1286, 1288, 1290, 1292, 1294, 1296, 1298, 1300, 1302, 1304, 1306, 1308, 1310, 1312, 1314, 1316, 1318, 1320, 1322, 1324, 1326, 1328, 1330, 1332, 1334, 1336, 1338, 1340, 1342, 1344, 1346, 1348, 1350, 1352, 1354, 1356, 1358, 1360, 1362, 1364, 1366, 1368, 1370, 1372, 1374, 1376, 1378, 1380, 1382, 1384, 1386, 1388, 1390, 1392, 1394, 1396, 1398, 1400, 1402, 1404, 1406, 1408, 1410, 1412, 1414, 1416, 1418, 1420, 1422, 1424, 1426, 1428, 1430, 1432, 1434, 1436, 1438, 1440, 1442, 1444, 1446, 1448, 1450, 1452, 1454, 1456, 1458, 1460, 1462, 1464, 1466, 1468, 1470, 1472, 1474, 1476, 1478, 1480, 1482, 1484, 1486, 1488, 1490, 1492, 1494, 1496, 1498, 1500, 1502, 1504, 1506, 1508, 1510, 1512, 1514, 1516, 1518, 1520, 1522, 1524, 1526, 1528, 1530, 1532, 1534, 1536, 1538, 1540, 1542, 1544, 1546, 1548, 1550, 1552, 1554, 1556, 1558, 1560, 1562, 1564, 1566, 1568, 1570, 1572, 1574, 1576, 1578, 1580, 1582, 1584, 1586, 1588, 1590, 1592, 1594, 1596, 1598, 1600, 1602, 1604, 1606, 1608, 1610, 1612, 1614, 1616, 1618, 1620, 1622, 1624, 1626, 1628, 1630, 1632, 1634, 1636, 1638, 1640, 1642, 1644, 1646, 1648, 1650, 1652, 1654, 1656, 1658, 1660, 1662, 1664, 1666, 1668, 1670, 1672, 1674, 1676, 1678, 1680, 1682, 1684, 1686, 1688, 1690, 1692, 1694, 1696, 1698, 1700, 1702, 1704, 1706, 1708, 1710, 1712, 1714, 1716, 1718, 1720,

1722, 1724, 1726, 1728, 1730, 1732, 1734, 1736, 1738, 1740, 1742, 1744, 1746, 1748, 1750, 1752, 1754, 1756, 1758, 1760, 1762, 1764, 1766, 1768, 1770, 1772, 1774, 1776, 1778, 1780, 1782, 1784, 1786, 1788, 1790, 1792, 1794, 1796, 1798, 1800, 1802, 1804, 1806, 1808, 1810, 1812, 1814, 1816, 1818, 1820, 1822, 1824, 1826, 1828, 1830, 1832, 1834, 1836, 1838, 1840, 1842, 1844, 1846, 1848, 1850, 1852, 1854, 1856, 1858, 1860, 1862, 1864, 1866, 1868, 1870, 1872, 1874, 1876, 1878, 1880, 1882, 1884, 1886, 1888, 1890, 1892, 1894, 1896, 1898, 1900, 1902, 1904, 1906, 1908, 1910, 1912, 1914, 1916, 1918, 1920, 1922, 1924, 1926, 1928, 1930, 1932, 1934, 1936, 1938, 1940, 1942, 1944, 1946, 1948, 1950, 1952, 1954, 1956, 1958, 1960, 1962, 1964, 1966, 1968, 1970, 1972, 1974, 1976, 1978, 1980, 1982, 1984, 1986, 1988, 1990, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024, 2026, 2028, 2030, 2032, 2034, 2036, 2038, 2040, 2042, 2044, 2046, 2048, 2050, 2052, 2054, 2056, 2058, 2060, 2062, 2064, 2066, 2068, 2070, 2072, 2074, 2076, 2078, 2080, 2082, 2084, 2086, 2088, 2090, 2092, 2094, 2096, 2098, 2100, 2102, 2104, 2106, 2108, 2110, 2112, 2114, 2116, 2118, 2120, 2122, 2124, 2126, 2128, 2130, 2132, 2134, 2136, 2138, 2140, 2142, 2144, 2146, 2148, 2150, 2152, 2154, 2156, 2158, 2160, 2162, 2164, 2166, 2168, 2170, 2172, 2174, 2176, 2178, 2180, 2182, 2184, 2186, 2188, 2190, 2192, 2194, 2196, 2198, 2200, 2202, 2204, 2206, 2208, 2210, 2212, 2214, 2216, 2218, 2220, 2222, 2224, 2226, 2228, 2230, 2232, 2234, 2236, 2238, 2240, 2242, 2244, 2246, 2248, 2250, 2252, 2254, 2256, 2258, 2260, 2262, 2264, 2266, 2268, 2270, 2272, 2274, 2276, 2278, 2280, 2282, 2284, 2286, 2288, 2290, 2292, 2294, 2296, 2298, 2300, 2302, 2304, 2306, 2308, 2310, 2312, 2314, 2316, 2318, 2320, 2322, 2324, 2326, 2328, 2330, 2332, 2334, 2336, 2338, 2340, 2342, 2344, 2346, 2348, 2350, 2352, 2354, 2356, 2358, 2360, 2362, 2364, 2366, 2368, 2370, 2372, 2374, 2376, 2378, 2380, 2382, 2384, 2386, 2388, 2390, 2392, 2394, 2396, 2398, 2400, 2402, 2404, 2406, 2408, 2410, 2412, 2414, 2416, 2418, 2420, 2422, 2424, 2426, 2428, 2430, 2432, 2434, 2436, 2438, 2440, 2442, 2444, 2446, 2448, 2450, 2452, 2454, 2456, 2458, 2460, 2462, 2464, 2466, 2468, 2470, 2472, 2474, 2476, 2478, 2480, 2482, 2484, 2486, 2488, 2490, 2492, 2494, 2496, 2498, 2500, 2502, 2504, 2506, 2508, 2510, 2512, 2514, 2516, 2518, 2520, 2522, 2524, 2526, 2528, 2530, 2532, 2534, 2536, 2538, 2540, 2542, 2544, 2546, 2548, 2550, 2552, 2554, 2556, 2558, 2560, 2562, 2564, 2566, 2568, 2570, 2572, 2574, 2576, 2578, 2580, 2582, 2584, 2586, 2588, 2590, 2592, 2594, 2596, 2598, 2600, 2602, 2604, 2606, 2608, 2610, 2612, 2614, 2616, 2618, 2620, 2622, 2624, 2626, 2628, 2630, 2632, 2634, 2636, 2638, 2640, 2642, 2644, 2646, 2648, 2650, 2652, 2654, 2656, 2658, 2660, 2662, 2664, 2666, 2668, 2670, 2672, 2674, 2676, 2678, 2680, 2682, 2684, 2686, 2688, 2690, 2692, 2694, 2696, 2698, 2700, 2702, 2704, 2706, 2708, 2710, 2712, 2714, 2716, 2718, 2720, 2722, 2724, 2726, 2728, 2730, 2732, 2734, 2736, 2738, 2740, 2742, 2744, 2746, 2748, 2750, 2752, 2754, 2756, 2758, 2760, 2762, 2764, 2766, 2768, 2770, 2772, 2774, 2776, 2778, 2780, 2782, 2784, 2786, 2788, 2790, 2792, 2794, 2796, 2798, 2800, 2802, 2804, 2806, 2808, 2810, 2812, 2814, 2816, 2818, 2820, 2822, 2824, 2826, 2828, 2830, 2832, 2834, 2836, 2838, 2840, 2842, 2844, 2846, 2848, 2850, 2852, 2854, 2856, 2858, 2860, 2862, 2864, 2866, 2868, 2870, 2872, 2874, 2876, 2878, 2880, 2882, 2884, 2886, 2888, 2890, 2892, 2894, 2896, 2898, 2900, 2902, 2904, 2906, 2908, 2910, 2912, 2914, 2916, 2918, 2920, 2922, 2924, 2926, 2928, 2930, 2932, 2934, 2936, 2938, 2940, 2942, 2944, 2946, 2948, 2950, 2952, 2954, 2956, 2958, 2960, 2962, 2964, 2966, 2968, 2970, 2972, 2974, 2976, 2978, 2980, 2982, 2984, 2986, 2988, 2990, 2992, 2994, 2996, 2998, 3000, 3002, 3004, 3006, 3008, 3010, 3012, 3014, 3016, 3018, 3020, 3022, 3024, 3026, 3028, 3030, 3032, 3034, 3036, 3038, 3040, 3042, 3044, 3046, 3048, 3050, 3052, 3054, 3056, 3058, 3060, 3062, 3064, 3066, 3068, 3070, 3072, 3074, 3076, 3078, 3080, 3082, 3084, 3086, 3088, 3090, 3092, 3094, 3096, 3098, 3100, 3102, 3104, 3106, 3108, 3110, 3112, 3114, 3116, 3118, 3120, 3122, 3124, 3126, 3128, 3130, 3132, 3134, 3136, 3138, 3140, 3142, 3144, 3146, 3148, 3150, 3152, 3154, 3156, 3158, 3160, 3162, 3164, 3166, 3168, 3170, 3172, 3174, 3176, 3178, 3180, 3182, 3184, 3186, 3188, 3190, 3192, 3194, 3196, 3198, 3200, 3202, 3204, 3206, 3208, 3210, 3212, 3214, 3216, 3218, 3220, 3222, 3224, 3226, 3228, 3230, 3232, 3234, 3236, 3238, 3240, 3242, 3244, 3246, 3248, 3250, 3252, 3254, 3256, 3258, 3260, 3262, 3264, 3266, 3268, 3270, 3272, 3274, 3276, 3278,

3280, 3282, 3284, 3286, 3288, 3290, 3292, 3294, 3296, 3298, 3300, 3302, 3304, 3306, 3308, 3310, 3312, 3314, 3316, 3318, 3320, 3322, 3324, 3326, 3328, 3330, 3332, 3334, 3336, 3338, 3340, 3342, 3344, 3346, 3348, 3350, 3352, 3354, 3356, 3358, 3360, 3362, 3364, 3366, 3368, 3370, 3372, 3374, 3376, 3378, 3380, 3382, 3384, 3386, 3388, 3390, 3392, 3394, 3396, 3398, 3400, 3402, 3404, 3406, 3408, 3410, 3412, 3414, 3416, 3418, 3420, 3422, 3424, 3426, 3428, 3430, 3432, 3434, 3436, 3438, 3440, 3442, 3444, 3446, 3448, 3450, 3452, 3454, 3456, 3458, 3460, 3462, 3464, 3466, 3468, 3470, 3472, 3474, 3476, 3478, 3480, 3482, 3484, 3486, 3488, 3490, 3492, 3494, 3496, 3498, 3500, 3502, 3504, 3506, 3508, 3510, 3512, 3514, 3516, 3518, 3520, 3522, 3524, 3526, 3528, 3530, 3532, 3534, 3536, 3538, 3540, 3542, 3544, 3546, 3548, 3550, 3552, 3554, 3556, 3558, 3560, 3562, 3564, 3566, 3568, 3570, 3572, 3574, 3576, 3578, 3580, 3582, 3584, 3586, 3588, 3590, 3592, 3594, 3596, 3598, 3600, 3602, 3604, 3606, 3608, 3610, 3612, 3614, 3616, 3618, 3620, 3622, 3624, 3626, 3628, 3630, 3632, 3634, 3636, 3638, 3640, 3642, 3644, 3646, 3648, 3650, 3652, 3654, 3656, 3658, 3660, 3662, 3664, 3666, 3668, 3670, 3672, 3674, 3676, 3678, 3680, 3682, 3684, 3686, 3688, 3690, 3692, 3694, 3696, 3698, 3700, 3702, 3704, 3706, 3708, 3710, 3712, 3714, 3716, 3718, 3720, 3722, 3724, 3726, 3728, 3730, 3732, 3734, 3736, 3738, 3740, 3742, 3744, 3748, 3750, 3752, 3754, 3756, 3758, 3760, 3762, 3764, 3766, 3768, 3770, 3772, 3774, 3776, 3778, 3780, 3782, 3784, 3786, 3788, 3790, 3792, 3794, 3796, 3798, 3800, 3802, 3804, 3806, 3808, 3810, 3812, 3814, 3816, 3818, 3820, 3822, 3824, 3826, 3828, 3830, 3832, 3834, 3836, 3838, 3840, 3842, 3844, 3846, 3848, 3850, 3852, 3854, 3856, 3858, 3860, 3862, 3864, 3866, 3868, 3870, 3872, 3874, 3876, 3878, 3880, 3882, 3884, 3886, 3888, 3890, 3892, 3894, 3896, 3898, 3900, 3902, 3904, 3906, 3908, 3910, 3912, 3914, 3916, 3918, 3920, 3924, 3926, 3928, 3930, 3932, 3934, 3936, 3938, 3940, 3942, 3944, 3946, 3948, 3950, 3952, 3954, 3956, 3958, 3960, 3962, 3964, 3966, 3968, 3970, 3972, 3974, 3976, 3978, 3980, 3982, 3984, 3986, 3988, 3990, 3992, 3994, 3996, 3998, 4000, 4002, 4004, 4006, 4008, 4010, 4012, 4014, 4016, 4018, 4020, 4022, 4024, 4026, 4028, 4030, 4032, 4034, 4036, 4038, 4040, 4042, 4044, 4046, 4048, 4050, 4052, 4054, 4056, 4058, 4060, 4062, 4064, 4066, 4068, 4070, 4072, 4074, 4076, 4078, 4080, 4082, 4084, 4086, 4088, 4090, 4092, 4094, 4096, 4098, 4100, 4102, 4104, 4106, 4108, 4110, 4112, 4114, 4116, 4118, 4120, 4122, 4124, 4126, 4128, 4130, 4132, 4134, 4136, 4138, 4140, 4142, 4144, 4146, 4148, 4150, 4152, 4154, 4156, 4158, 4160, 4162, 4164, 4166, 4168, 4170, 4172, 4174, 4176, 4178, 4180, 4182, 4184, 4186, 4188, 4190, 4192, 4194, 4196, 4198, 4200, 4202, 4204, 4206, 4208, 4210, 4212, 4214, 4216, 4218, 4220, 4222, 4224, 4226, 4228, 4230, 4232, 4234, 4236, 4238, 4240, 4242, 4244, 4246, 4248, 4250, 4252, 4254, 4256, 4258, 4260, 4262, 4264, 4266, 4268, 4270, 4272, 4274, 4276, 4278, 4280, 4282, 4284, 4286, 4288, 4290, 4292, 4294, 4296, 4298, 4300, 4302, 4304, 4306, 4308, 4310, 4312, 4314, 4316, 4318, 4320, 4322, 4324, 4326, 4328, 4330, 4332, 4334, 4336, 4338, 4340, 4342, 4344, 4346, 4348, 4350, 4352, 4354, 4356, 4358, 4360, 4362, 4364, 4366, 4368, 4370, 4372, 4374, 4376, 4378, 4380, 4382, 4384, 4386, 4388, 4390, 4392, 4394, 4396, 4398, 4400, 4402, 4404, 4406, 4408, 4410, 4412, 4414, 4416, 4418, 4420, 4422, 4424, 4426, 4428, 4430, 4432, 4434, 4436, 4438, 4440, 4442, 4444, 4446, 4448, 4450, 4452, 4454, 4456, 4458, 4460, 4462, 4464, 4466, 4468, 4470, 4472, 4474, 4476, 4478, 4480, 4482, 4484, 4486, 4488, 4490, 4492, 4494, 4496, 4498, 4500, 4502, 4504, 4506, 4508, 4510, 4512, 4514, 4516, 4518, 4520, 4522, 4524, 4526, 4528, 4530, 4532, 4534, 4536, 4538, 4540, 4542, 4544, 4546, 4548, 4550, 4552, 4554, 4556, 4558, 4560, 4562, 4564, 4566, 4568, 4570, 4572, 4574, 4576, 4578, 4580, 4582, 4584, 4586, 4588, 4590, 4592, 4594, 4596, 4598, 4600, 4602, 4604, 4606, 4608, 4610, 4612, 4614, 4616, 4618, 4620, 4622, 4624, 4626, 4628, 4630, 4632, 4634, 4636, 4638, 4640, 4642, 4644, 4646, 4648, 4650, 4652, 4654, 4656, 4658, 4660, 4662, 4664, 4666, 4668, 4670, 4672, 4674, 4676, 4678, 4680, 4682, 4684, 4686, 4688, 4690, 4692, 4694, 4696, 4698, 4700, 4702, 4704, 4706, 4708, 4710, 4712, 4714, 4716, 4718, 4720, 4722, 4724, 4726, 4728, 4730, 4732, 4734, 4736, 4738, 4740, 4742, 4744, 4746, 4748, 4750, 4752, 4754, 4756, 4758, 4760, 4762, 4764, 4766, 4768, 4770, 4772, 4774, 4776, 4778, 4780, 4782, 4784, 4786, 4788, 4790, 4792, 4794, 4796, 4798, 4800, 4802, 4804, 4806, 4808, 4810, 4812, 4814, 4816, 4818, 4820, 4822, 4824, 4826, 4828, 4830, 4832, 4834, 4836,

4838, 4840, 4842, 4844, 4846, 4848, 4850, 4852, 4854, 4856, 4858, 4860, 4862, 4864, 4866, 4868, 4870, 4872, 4874, 4876, 4878, 4880, 4882, 4884, 4886, 4888, 4890, 4892, 4894, 4896, 4898, 4900, 4902, 4904, 4906, 4908, 4910, 4912, 4914, 4916, 4918, 4920, 4922, 4924, 4926, 4928, 4930, 4932, 4934, 4936, 4938, 4940, 4942, 4944, 4946, 4948, 4950, 4952, 4954, 4956, 4958, 4960, 4962, 4964, 4966, 4968, 4970, 4972, 4974, 4976, 4978, 4980, 4982, 4984, 4986, 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5002, 5004, 5006, 5008, 5010, 5012, 5014, 5016, 5018, 5020, 5022, 5024, 5026, 5028, 5030, 5032, 5034, 5036, 5038, 5040, 5042, 5044, 5046, 5048, 5050, 5052, 5054, 5056, 5058, 5060, 5062, 5064, 5066, 5068, 5070, 5072, 5074, 5076, 5078, 5080, 5082, 5084, 5086, 5088, 5090, 5092, 5094, 5096, 5098, 5100, 5102, 5104, 5106, 5108, 5110, 5112, 5114, 5116, 5118, 5120, 5122, 5124, 5126, 5128, 5130, 5132, 5134, 5136, 5138, 5140, 5142, 5144, 5146, 5148, 5150, 5152, 5154, 5156, 5158, 5160, 5162, 5164, 5166, 5168, 5170, 5172, 5174, 5176, 5178, 5180, 5182, 5184, 5186, 5188, 5190, 5192, 5194, 5196, 5198, 5200, 5202, 5204, 5206, 5208, 5210, 5212, 5214, 5216, 5218, 5220, 5222, 5224, 5226, 5228, 5230, 5232, 5234, 5236, 5238, 5240, 5242, 5244, 5246, 5248, 5250, 5252, 5254, 5256, 5258, 5260, 5262, 5264, 5266, 5268, 5270, 5272, 5274, 5276, 5278, 5280, 5282, 5284, 5286, 5288, 5290, 5292, 5294, 5296, 5298, 5300, 5302, 5304, 5306, 5308, 5310, 5312, 5314, 5316, 5318, 5320, 5322, 5324, 5326, 5328, 5330, 5332, 5334, 5336, 5338, 5340, 5342, 5344, 5346, 5348, 5350, 5352, 5354, 5356, 5358, 5360, 5362, 5364, 5366, 5368, 5370, 5372, 5374, 5376, 5378, 5380, 5382, 5384, 5386, 5388, 5390, 5392, 5394, 5396, 5398, 5400, 5402, 5404, 5406, 5408, 5410, 5412, 5414, 5416, 5418, 5420, 5422, 5424, 5426, 5428, 5430, 5432, 5434, 5436, 5438, 5440, 5442, 5444, 5446, 5448, 5450, 5452, 5454, 5456, 5458, 5460, 5462, 5464, 5466, 5468, 5470, 5472, 5474, 5476, 5478, 5480, 5482, 5484, 5486, 5488, 5490, 5492, 5494, 5496, 5498, 5500, 5502, 5504, 5506, 5508, 5510, 5512, 5514, 5516, 5518, 5520, 5522, 5524, 5526, 5528, 5530, 5532, 5534, 5536, 5538, 5540, 5542, 5544, 5546, 5548, 5550, 5552, 5554, 5556, 5558, 5560, 5562, 5564, 5566, 5568, 5570, 5572, 5574, 5576, 5578, 5580, 5582, 5584, 5586, 5588, 5590, 5592, 5594, 5596, 5598, 5600, 5602, 5604, 5606, 5608, 5610, 5612, 5614, 5616, 5618, 5620, 5622, 5624, 5626, 5628, 5630, 5632, 5634, 5636, 5638, 5640, 5642, 5644, 5646, 5648, 5650, 5652, 5654, 5656, 5658, 5660, 5662, 5664, 5666, 5668, 5670, 5672, 5674, 5676, 5678, 5680, 5682, 5684, 5686, 5688, 5690, 5692, 5694, 5696, 5698, 5700, 5702, 5704, 5706, 5708, 5710, 5712, 5714, 5716, 5718, 5720, 5722, 5724, 5726, 5728, 5730, 5732, 5734, 5736, 5738, 5740, 5742, 5744, 5746, 5748, 5750, 5752, 5754, 5756, 5758, 5760, 5762, 5764, 5766, 5768, 5770, 5772, 5774, 5776, 5778, 5780, 5782, 5784, 5786, 5788, 5790, 5792, 5794, 5796, 5798, 5800, 5802, 5804, 5806, 5808, 5810, 5812, 5814, 5816, 5818, 5820, 5822, 5824, 5826, 5828, 5830, 5832, 5834, 5836, 5838, 5840, 5842, 5844, 5846, 5848, 5850, 5852, 5854, 5856, 5858, 5860, 5862, 5864, 5866, 5868, 5870, 5872, 5874, 5876, 5878, 5880, 5882, 5884, 5886, 5888, 5890, 5892, 5894, 5896, 5898, 5900, 5902, 5904, 5906, 5908, 5910, 5912, 5914, 5916, 5918, 5920, 5922, 5924, 5926, 5928, 5930, 5932, 5934, 5936, 5938, 5940, 5942, 5944, 5946, 5948, 5950, 5952, 5954, 5956, 5958, 5960, 5962, 5964, 5966, 5968, 5970, 5972, 5974, 5976, 5978, 5980, 5982, 5984, 5986, 5988, 5990, 5992, 5994, 5996, 5998, 6000, 6002, 6004, 6006, 6008, 6010, 6012, 6014, 6016, 6018, 6020, 6022, 6024, 6026, 6028, 6030, 6032, 6034, 6036, 6038, 6040, 6042, 6044, 6046, 6048, 6050, 6052, 6054, 6056, 6058, 6060, 6062, 6064, 6066, 6068, 6070, 6072, 6074, 6076, 6078, 6080, 6082, 6084, 6086, 6088, 6090, 6092, 6094, 6096, 6098, 6100, 6102, 6104, 6106, 6108, 6110, 6112, 6114, 6116, 6118, 6120, 6122, 6124, 6126, 6128, 6130, 6132, 6134, 6136, 6138, 6140, 6142, 6144, 6146, 6148, 6150, 6152, 6154, 6156, 6158, 6160, 6162, 6164, 6166, 6168, 6170, 6172, 6174, 6176, 6178, 6180, 6182, 6184, 6186, 6188, 6190, 6192, 6194, 6196, 6198, 6200, 6202, 6204, 6206, 6208, 6210, 6212, 6214, 6216, 6218, 6220, 6222, 6224, 6226, 6228, 6230, 6232, 6234, 6236, 6238, 6240, 6242, 6244, 6246, 6248, 6250, 6252, 6254, 6256, 6258, 6260, 6262, 6264, 6266, 6268, 6270, 6272, 6274, 6276, 6278, 6280, 6282, 6284, 6286, 6288, 6290, 6292, 6294, 6296, 6298, 6300, 6302, 6304, 6306, 6308, 6310, 6312, 6314, 6316, 6318, 6320, 6322, 6324, 6326, 6328, 6330, 6332, 6334, 6336, 6338, 6340, 6342, 6344, 6346, 6348, 6350, 6352, 6354, 6356, 6358, 6360, 6362, 6364, 6366, 6368, 6370, 6372, 6374, 6376, 6378, 6380, 6382, 6384, 6386, 6388, 6390, 6392, 6394,

6396, 6398, 6400, 6402, 6404, 6406, 6408, 6410, 6412, 6414, 6416, 6418, 6420, 6422, 6424, 6426, 6428, 6430, 6432, 6434, 6436, 6438, 6440, 6442, 6444, 6446, 6448, 6450, 6452, 6454, 6456, 6458, 6460, 6462, 6464, 6466, 6468, 6470, 6472, 6474, 6476, 6478, 6480, 6482, 6484, 6486, 6488, 6490, 6492, 6494, 6496, 6498, 6500, 6502, 6504, 6506, 6508, 6510, 6512, 6514, 6516, 6518, 6520, 6522, 6524, 6526, 6528, 6530, 6532, 6534, 6536, 6538, 6540, 6542, 6544, 6546, 6548, 6550, 6552, 6554, 6556, 6558, 6560, 6562, 6564, 6566, 6568, 6570, 6572, 6574, 6576, 6578, 6580, 6582, 6584, 6586, 6588, 6590, 6592, 6594, 6596, 6598, 6600, 6602, 6604, 6606, 6608, 6610, 6612, 6614, 6616, 6618, 6620, 6622, 6624, 6626, 6628, 6630, 6632, 6634, 6636, 6638, 6640, 6642, 6644, 6646, 6648, 6650, 6652, 6654, 6656, 6658, 6660, 6662, 6664, 6666, 6668, 6670, 6672, 6674, 6676, 6678, 6680, 6682, 6684, 6686, 6688, 6690, 6692, 6694, 6696, 6698, 6700, 6702, 6704, 6706, 6708, 6710, 6712, 6714, 6716, 6718, 6720, 6722, 6724, 6726, 6728, 6730, 6732, 6734, 6736, 6738, 6740, 6742, 6744, 6746, 6748, 6750, 6752, 6754, 6756, 6758, 6760, 6762, 6764, 6766, 6768, 6770, 6772, 6774, 6776, 6778, 6780, 6782, 6784, 6786, 6788, 6790, 6792, 6794, 6796, 6798, 6800, 6802, 6804, 6806, 6808, 6810, 6812, 6814, 6816, 6818, 6820, 6822, 6824, 6826, 6828, 6830, 6832, 6834, 6836, 6838, 6840, 6842, 6844, 6846, 6848, 6850, 6852, 6854, 6856, 6858, 6860, 6862, 6864, 6866, 6868, 6870, 6872, 6874, 6876, 6878, 6880, 6882, 6884, 6886, 6888, 6890, 6892, 6894, 6896, 6898, 6900, 6902, 6904, 6906, 6908, 6910, 6912, 6914, 6916, 6918, 6920, 6922, 6924, 6926, 6928, 6930, 6932, 6934, 6936, 6938, 6940, 6942, 6944, 6946, 6948, 6950, 6952, 6954, 6956, 6958, 6960, 6962, 6964, 6966, 6968, 6970, 6972, 6974, 6976, 6978, 6980, 6982, 6984, 6986, 6988, 6990, 6992, 6994, 6996, 6998, 7000, 7002, 7004, 7006, 7008, 7010, 7012, 7014, 7016, 7018, 7020, 7022, 7024, 7026, 7028, 7030, 7032, 7034, 7036, 7038, 7040, 7042, 7044, 7046, 7048, 7050, 7052, 7054, 7056, 7058, 7060, 7062, 7064, 7066, 7068, 7070, 7072, 7074, 7076, 7078, 7080, 7082, 7084, 7086, 7088, 7090, 7092, 7094, 7096, 7098, 7100, 7102, 7104, 7106, 7108, 7110, 7112, 7114, 7116, 7118, 7120, 7122, 7124, 7126, 7128, 7130, 7132, 7134, 7136, 7138, 7140, 7142, 7144, 7146, 7148, 7150, 7152, 7154, 7156, 7158, 7160, 7162, 7164, 7166, 7168, 7170, 7172, 7174, 7176, 7178, 7180, 7182, 7184, 7186, 7188, 7190, 7192, 7194, 7196, 7198, 7200, 7202, 7204, 7206, 7208, 7210, 7212, 7214, 7216, 7218, 7220, 7222, 7224, 7226, 7228, 7230, 7232, 7234, 7236, 7238, 7240, 7242, 7244, 7246, 7248, 7250, 7252, 7254, 7256, 7258, 7260, 7262, 7264, 7266, 7268, 7270, 7272, 7274, 7276, 7278, 7280, 7282, 7284, 7286, 7288, 7290, 7292, 7294, 7296, 7298, 7300, 7302, 7304, 7306, 7308, 7310, 7312, 7314, 7316, 7318, 7320, 7322, 7324, 7326, 7328, 7330, 7332, 7334, 7336, 7338, 7340, 7342, 7344, 7346, 7348, 7350, 7352, 7354, 7356, 7358, 7360, 7362, 7364, 7366, 7368, 7370, 7372, 7374, 7376, 7378, 7380, 7382, 7384, 7386, 7388, 7390, 7392, 7394, 7396, 7398, 7400, 7402, 7404, 7406, 7408, 7410, 7412, 7414, 7416, 7418, 7420, 7422, 7424, 7426, 7428, 7430, 7432, 7434, 7436, 7438, 7440, 7442, 7444, 7446, 7448, 7450, 7452, 7454, 7456, 7458, 7460, 7462, 7464, 7468, 7470, 7472, 7474, 7476, 7478, 7480, 7482, 7484, 7486, 7488, 7490, 7492, 7494, 7496, 7498, 7500, 7502, 7504, 7506, 7508, 7510, 7512, 7514, 7516, 7518, 7520, 7522, 7524, 7526, 7528, 7530, 7532, 7534, 7536, 7538, 7540, 7542, 7544, 7546, 7548, 7550, 7552, 7554, 7556, 7558, 7560, 7562, 7564, 7566, 7568, 7570, 7572, 7574, 7576, 7578, 7580, 7582, 7584, 7586, 7588, 7590, 7592, 7594, 7596, 7598, 7600, 7602, 7604, 7606, 7608, 7610, 7612, 7614, 7616, 7618, 7620, 7622, 7624, 7626, 7628, 7630, 7632, 7634, 7636, 7638, 7640, 7642, 7644, 7646, 7648, 7650, 7652, 7654, 7656, 7658, 7660, 7662, 7664, 7666, 7668, 7670, 7672, 7674, 7676, 7678, 7680, 7682, 7684, 7686, 7688, 7690, 7692, 7694, 7696, 7698, 7700, 7702, 7704, 7706, 7708, 7710, 7712, 7714, 7716, 7718, 7720, 7722, 7724, 7726, 7728, 7730, 7732, 7734, 7736, 7738, 7740, 7742, 7744, 7746, 7748, 7750, 7752, 7754, 7756, 7758, 7760, 7762, 7764, 7766, 7768, 7770, 7772, 7774, 7776, 7778, 7780, 7782, 7784, 7786, 7788, 7790, 7792, 7794, 7796, 7798, 7800, 7802, 7804, 7806, 7808, 7810, 7812, 7814, 7816, 7818, 7820, 7822, 7824, 7826, 7828, 7830, 7832, 7834, 7836, 7838, 7840, 7842, 7844, 7846, 7848, 7850, 7852, 7854, 7856, 7858, 7860, 7862, 7864, 7866, 7868, 7870, 7872, 7874, 7876, 7878, 7880, 7882, 7884, 7886, 7888, 7890, 7892, 7894, 7896, 7898, 7900, 7902, 7904, 7906, 7908, 7910, 7912, 7914, 7916, 7918, 7920, 7922, 7924, 7926, 7928, 7930, 7932, 7934, 7936, 7938, 7940, 7942, 7944, 7946, 7948, 7950, 7952,

7954, 7956, 7958, 7960, 7962, 7964, 7966, 7968, 7970, 7972, 7974, 7976, 7978, 7980, 7982, 7984, 7986, 7988, 7990, 7992, 7994, 7996, 7998, 8000, 8002, 8004, 8006, 8008, 8010, 8012, 8014, 8016, 8018, 8020, 8022, 8024, 8026, 8028, 8030, 8032, 8034, 8036, 8038, 8040, 8042, 8044, 8046, 8048, 8050, 8052, 8054, 8056, 8058, 8060, 8062, 8064, 8066, 8068, 8070, 8072, 8074, 8076, 8078, 8080, 8082, 8084, 8086, 8088, 8090, 8092, 8094, 8096, 8098, 8100, 8102, 8104, 8106, 8108, 8110, 8112, 8114, 8116, 8118, 8120, 8122, 8124, 8126, 8128, 8130, 8132, 8134, 8136, 8138, 8140, 8142, 8144, 8146, 8148, 8150, 8152, 8154, 8156, 8158, 8160, 8162, 8164, 8166, 8168, 8170, 8172, 8174, 8176, 8178, 8180, 8182, 8184, 8186, 8188, 8190, 8192, 8194, 8196, 8198, 8200, 8202, 8204, 8206, 8208, 8210, 8212, 8214, 8216, 8218, 8220, 8222, 8224, 8226, 8228, 8230, 8232, 8234, 8236, 8238, 8240, 8242, 8244, 8246, 8248, 8250, 8252, 8254, 8256, 8258, 8260, 8262, 8264, 8266, 8268, 8270, 8272, 8274, 8276, 8278, 8280, 8282, 8284, 8286, 8288, 8290, 8292, 8294, 8296, 8298, 8300, 8302, 8304, 8306, 8308, 8310, 8312, 8314, 8316, 8318, 8320, 8322, 8324, 8326, 8328, 8330, 8332, 8334, 8336, 8338, 8340, 8342, 8344, 8346, 8348, 8350, 8352, 8354, 8356, 8358, 8360, 8362, 8364, 8366, 8368, 8370, 8372, 8374, 8376, 8378, 8380, 8382, 8384, 8386, 8388, 8390, 8392, 8394, 8396, 8398, 8400, 8402, 8404, 8406, 8408, 8410, 8412, 8414, 8416, 8418, 8420, 8422, 8424, 8426, 8428, 8430, 8432, 8434, 8436, 8438, 8440, 8442, 8444, 8446, 8448, 8450, 8452, 8454, 8456, 8458, 8460, 8462, 8464, 8466, 8468, 8470, 8472, 8474, 8476, 8478, 8480, 8482, 8484, 8486, 8488, 8490, 8492, 8494, 8496, 8498, 8500, 8502, 8504, 8506, 8508, 8510, 8512, 8514, 8516, 8518, 8520, 8522, 8524, 8526, 8528, 8530, 8532, 8534, 8536, 8538, 8540, 8542, 8544, 8546, 8548, 8550, 8552, 8554, 8556, 8558, 8560, 8562, 8564, 8566, 8568, 8570, 8572, 8574, 8576, 8578, 8580, 8582, 8584, 8586, 8588, 8590, 8592, 8594, 8596, 8598, 8600, 8602, 8604, 8606, 8608, 8610, 8612, 8614, 8616, 8618, 8620, 8622, 8624, 8626, 8628, 8630, 8632, 8634, 8636, 8638, 8640, 8642, 8644, 8646, 8648, 8650, 8652, 8654, 8656, 8658, 8660, 8662, 8664, 8666, 8668, 8670, 8672, 8674, 8676, 8678, 8680, 8682, 8684, 8686, 8688, 8690, 8692, 8694, 8696, 8698, 8700, 8702, 8704, 8706, 8708, 8710, 8712, 8714, 8716, 8718, 8720, 8722, 8724, 8726, 8728, 8730, 8732, 8734, 8736, 8738, 8740, 8742, 8744, 8746, 8748, 8750, 8752, 8754, 8756, 8758, 8760, 8762, 8764, 8766, 8768, 8770, 8772, 8774, 8776, 8778, 8782, 8784, 8786, 8788, 8790, 8792, 8794, 8796, 8798, 8800, 8802, 8804, 8806, 8808, 8810, 8812, 8814, 8816, 8818, 8820, 8822, 8824, 8826, 8828, 8830, 8832, 8834, 8836, 8838, 8840, 8842, 8844, 8846, 8848, 8850, 8852, 8854, 8856, 8858, 8860, 8862, 8864, 8866, 8868, 8870, 8872, 8874, 8876, 8878, 8880, 8882, 8884, 8886, 8888, 8890, 8892, 8894, 8896, 8898, 8900, 8902, 8904, 8906, 8908, 8910, 8912, 8914, 8916, 8918, 8920, 8922, 8924, 8926, 8928, 8930, 8932, 8934, 8936, 8938, 8940, 8942, 8944, 8946, 8948, 8950, 8952, 8954, 8956, 8958, 8960, 8962, 8964, 8966, 8968, 8970, 8972, 8974, 8976, 8978, 8980, 8982, 8984, 8986, 8988, 8990, 8992, 8994, 8996, 8998, 9000, 9002, 9004, 9006, 9008, 9010, 9012, 9014, 9016, 9018, 9020, 9022, 9024, 9026, 9028, 9030, 9032, 9034, 9036, 9038, 9040, 9042, 9044, 9046, 9048, 9050, 9052, 9054, 9056, 9058, 9060, 9062, 9064, 9066, 9068, 9070, 9072, 9074, 9076, 9078, 9080, 9082, 9084, 9086, 9088, 9090, 9092, 9094, 9096, 9098, 9100, 9102, 9104, 9106, 9108, 9110, 9112, 9114, 9116, 9118, 9120, 9122, 9124, 9126, 9128, 9130, 9132, 9134, 9136, 9138, 9140, 9142, 9144, 9146, 9148, 9150, 9152, 9154, 9156, 9158, 9160, 9162, 9164, 9166, 9168, 9170, 9172, 9174, 9176, 9178, 9180, 9182, 9184, 9186, 9190, 9192, 9194, 9196, 9198, 9200, 9202, 9204, 9206, 9208, 9210, 9212, 9214, 9216, 9218, 9220, 9222, 9224, 9226, 9228, 9230, 9232, 9234, 9236, 9238, 9240, 9242, 9244, 9246, 9248, 9250, 9252, 9254, 9256, 9258, 9260, 9262, 9264, 9266, 9268, 9270, 9272, 9274, 9276, 9278, 9280, 9282, 9284, 9286, 9288, 9290, 9292, 9294, 9296, 9298, 9300, 9302, 9304, 9306, 9308, 9310, 9312, 9314, 9316, 9318, 9320, 9322, 9324, 9326, 9328, 9330, 9332, 9334, 9336, 9338, 9340, 9342, 9344, 9346, 9348, 9350, 9352, 9354, 9356, 9358, 9360, 9362, 9364, 9366, 9368, 9370, 9372, 9374, 9376, 9378, 9380, 9382, 9384, 9386, 9388, 9390, 9392, 9394, 9396, 9398, 9400, 9402, 9404, 9406, 9408, 9410, 9412, 9414, 9416, 9418, 9420, 9422, 9424, 9426, 9428, 9430, 9432, 9434, 9436, 9438, 9440, 9442, 9444, 9446, 9448, 9450, 9452, 9454, 9456, 9458, 9460, 9462, 9464, 9466, 9468, 9470, 9472, 9474, 9476, 9478, 9480, 9482, 9484, 9486, 9488, 9490, 9492, 9494, 9496, 9498, 9500, 9502, 9504, 9506, 9508, 9510,

9512, 9514, 9516, 9518, 9520, 9522, 9524, 9526, 9528, 9530, 9532, 9534, 9536, 9538, 9540, 9542, 9544, 9546, 9548, 9550, 9552, 9554, 9556, 9558, 9560, 9562, 9564, 9566, 9568, 9570, 9572, 9574, 9576, 9578, 9580, 9582, 9584, 9586, 9588, 9590, 9592, 9594, 9596, 9598, 9600, 9602, 9604, 9606, 9608, 9610, 9612, 9614, 9616, 9618, 9620, 9622, 9624, 9626, 9628, 9630, 9632, 9634, 9636, 9638, 9640, 9642, 9644, 9646, 9648, 9650, 9652, 9654, 9656, 9658, 9660, 9662, 9664, 9666, 9668, 9670, 9672, 9674, 9676, 9678, 9680, 9682, 9684, 9686, 9688, 9690, 9692, 9694, 9696, 9698, 9700, 9702, 9704, 9706, 9708, 9710, 9712, 9714, 9716, 9718, 9720, 9722, 9724, 9726, 9728, 9730, 9732, 9734, 9736, 9738, 9740, 9742, 9744, 9746, 9748, 9750, 9752, 9754, 9756, 9758, 9760, 9762, 9764, 9766, 9768, 9770, 9772, 9774, 9776, 9778, 9780, 9782, 9784, 9786, 9788, 9790, 9792, 9794, 9796, 9798, 9800, 9802, 9804, 9806, 9808, 9810, 9812, 9814, 9816, 9818, 9820, 9822, 9824, 9826, 9828, 9830, 9832, 9834, 9836, 9838, 9840, 9842, 9844, 9846, 9848, 9850, 9852, 9854, 9856, 9858, 9860, 9862, 9864, 9866, 9868, 9870, 9872, 9874, 9876, 9878, 9880, 9882, 9884, 9886, 9888, 9890, 9892, 9894, 9896, 9898, 9900, 9902, 9904, 9906, 9908, 9910, 9912, 9914, 9916, 9918, 9920, 9922, 9924, 9926, 9928, 9930, 9932, 9934, 9936, 9938, 9940, 9942, 9944, 9946, 9948, 9950, 9952, 9954, 9956, 9958, 9960, 9962, 9964, 9966, 9968, 9970, 9972, 9974, 9976, 9978, 9980, 9982, 9984, 9986, 9988, 9990, 9992, 9994, 9996, 9998, 10000, 10002, 10004, 10006, 10008, 10010, 10012, 10014, 10016, 10018, 10020, 10022, 10024, 10026, 10028, 10030, 10032, 10034, 10036, 10038, 10040, 10042, 10044, 10046, 10048, 10050, 10052, 10054, 10056, 10058, 10060, 10062, 10064, 10066, 10068, 10070, 10072, 10074, 10076, 10078, 10080, 10082, 10084, 10086, 10088, 10090, 10092, 10094, 10096, 10098, 10100, 10102, 10104, 10106, 10108, 10110, 10112, 10114, 10116, 10118, 10120, 10122, 10124, 10126, 10128, 10130, 10132, 10134, 10136, 10138, 10140, 10142, 10144, 10146, 10148, 10150, 10152, 10154, 10156, 10158, 10160, 10162, 10164, 10166, 10168, 10170, 10172, 10174, 10176, 10178, 10180, 10182, 10184, 10186, 10188, 10190, 10192, 10194, 10196, 10198, 10200, 10202, 10204, 10206, 10208, 10210, 10212, 10214, 10216, 10218, 10220, 10222, 10224, 10226, 10228, 10230, 10232, 10234, 10236, 10238, 10240, 10242, 10244, 10246, 10248, 10250, 10252, 10254, 10256, 10258, 10260, 10262, 10264, 10266, 10268, 10270, 10272, 10274, 10276, 10278, 10280, 10282, 10284, 10286, 10288, 10290, 10292, 10294, 10296, 10298, 10300, 10302, 10304, 10306, 10308, 10310, 10312, 10314, 10316, 10318, 10320, 10322, 10324, 10326, 10328, 10330, 10332, 10334, 10336, 10338, 10340, 10342, 10344, 10346, 10348, 10350, 10352, 10354, 10356, 10358, 10360, 10362, 10364, 10366, 10368, 10370, 10372, 10374, 10376, 10378, 10380, 10382, 10384, 10386, 10388, 10390, 10392, 10394, 10396, 10398, 10400, 10402, 10404, 10406, 10408, 10410, 10412, 10414, 10416, 10418, 10420, 10422, 10424, 10426, 10428, 10430, 10432, 10434, 10436, 10438, 10440, 10442, 10444, 10446, 10448, 10450, 10452, 10454, 10456, 10458, 10460, 10462, 10464, 10466, 10468, 10470, 10472, 10474, 10476, 10478, 10480, 10482, 10484, 10486, 10488, 10490, 10492, 10494, 10496, 10498, 10500, 10502, 10504, 10506, 10508, 10510, 10512, 10514, 10516, 10518, 10520, 10522, 10524, 10526, 10528, 10530, 10532, 10534, 10536, 10538, 10540, 10542, 10544, 10546, 10548, 10550, 10552, 10554, 10556, 10558, 10560, 10562, 10564, 10566, 10568, 10570, 10572, 10574, 10576, 10578, 10580, 10582, 10584, 10586, 10588, 10590, 10592, 10594, 10596, 10598, 10600, 10602, 10604, 10606, 10608, 10610, 10612, 10614, 10616, 10618, 10620, 10622, 10624, 10626, 10628, 10630, 10632, 10634, 10636, 10638, 10640, 10642, 10644, 10646, 10648, 10650, 10652, 10654, 10656, 10658, 10660, 10662, 10664, 10666, 10668, 10670, 10672, 10674, 10676, 10678, 10680, 10682, 10684, 10686, 10688, 10690, 10692, 10694, 10696, 10698, 10700, 10702, 10704, 10706, 10708, 10710, 10712, 10714, 10716, 10718, 10720, 10722, 10724, 10726, 10728, 10730, 10732, 10734, 10736, 10738, 10740, 10742, 10744, 10746, 10748, 10750, 10752, 10754, 10756, 10758, 10760, 10762, 10764, 10766, 10768, 10770, 10772, 10774, 10776, 10778, 10780, 10782, 10784, 10786, 10788, 10790, 10792, 10794, 10796, 10798, 10800, 10802, 10804, 10806, 10808, 10810, 10812, 10814, 10816, 10818, 10820, 10822, 10824, 10826, 10828, 10830, 10832, 10834, 10836, 10838, 10840, 10842, 10844, 10846, 10848, 10850, 10852, 10854, 10856, 10858, 10860, 10862, 10864, 10866, 10868, 10870, 10872, 10874, 10876, 10878, 10880, 10882, 10884, 10886, 10888, 10890, 10892, 10894, 10896, 10898,

10900, 10902, 10904, 10906, 10908, 10910, 10912, 10914, 10916, 10918, 10920, 10922, 10924, 10926, 10928, 10930, 10932, 10934, 10936, 10938, 10940, 10942, 10944, 10946, 10948, 10950, 10952, 10954, 10956, 10958, 10960, 10962, 10964 and 10966.

2. A protein having 60% or greater sequence identity to a protein according to claim 1.

3. A protein comprising a fragment of 12 or more consecutive amino acids from an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, 892, 894, 896, 898, 900, 902, 904, 906, 908, 910, 912, 914, 916, 918, 920, 922, 924, 926, 928, 930, 932, 934, 936, 938, 940, 942, 944, 946, 948, 950, 952, 954, 956, 958, 960, 962, 964, 966, 968, 970, 972, 974, 976, 978, 980, 982, 984, 986, 988, 990, 992, 994, 996, 998, 1000, 1002, 1004, 1006, 1008, 1010, 1012, 1014, 1016, 1018, 1020, 1022, 1024, 1026, 1028, 1030, 1032, 1034, 1036, 1038, 1040, 1042, 1044, 1046, 1048, 1050, 1052, 1054, 1056, 1058, 1060, 1062, 1064, 1066, 1068, 1070, 1072, 1074, 1076, 1078, 1080, 1082, 1084, 1086, 1088, 1090, 1092, 1094, 1096, 1098, 1100, 1102, 1104, 1106, 1108, 1110, 1112, 1114, 1116, 1118, 1120, 1122, 1124, 1126, 1128, 1130, 1132, 1134, 1136, 1138, 1140, 1142, 1144, 1146, 1148, 1150, 1152, 1154, 1156, 1158, 1160, 1162, 1164, 1166, 1168, 1170, 1172, 1174, 1176, 1178, 1180, 1182, 1184, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1222, 1224, 1226, 1228, 1230, 1232, 1234, 1236, 1238, 1240, 1242, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, 1272, 1274, 1276, 1278, 1280, 1282, 1284, 1286, 1288, 1290, 1292, 1294, 1296, 1298, 1300, 1302, 1304, 1306, 1308, 1310, 1312, 1314, 1316, 1318, 1320, 1322, 1324, 1326, 1328, 1330, 1332, 1334, 1336, 1338, 1340, 1342, 1344, 1346, 1348, 1350, 1352, 1354, 1356, 1358, 1360, 1362, 1364, 1366, 1368, 1370, 1372, 1374, 1376, 1378, 1380, 1382, 1384, 1386, 1388, 1390, 1392, 1394, 1396, 1398, 1400, 1402, 1404, 1406, 1408, 1410, 1412, 1414, 1416, 1418, 1420, 1422, 1424, 1426, 1428, 1430, 1432, 1434, 1436, 1438, 1440, 1442, 1444, 1446, 1448, 1450, 1452, 1454, 1456, 1458, 1460, 1462, 1464, 1466, 1468, 1470, 1472, 1474, 1476, 1478, 1480, 1482, 1484, 1486, 1488, 1490, 1492, 1494, 1496, 1498, 1500, 1502, 1504, 1506, 1508, 1510, 1512, 1514, 1516, 1518, 1520, 1522, 1524, 1526, 1528, 1530, 1532, 1534, 1536, 1538, 1540, 1542, 1544, 1546, 1548, 1550, 1552, 1554, 1556, 1558, 1560, 1562, 1564, 1566, 1568, 1570, 1572, 1574, 1576, 1578, 1580, 1582, 1584, 1586, 1588, 1590,

1592, 1594, 1596, 1598, 1600, 1602, 1604, 1606, 1608, 1610, 1612, 1614, 1616, 1618, 1620, 1622, 1624, 1626, 1628, 1630, 1632, 1634, 1636, 1638, 1640, 1642, 1644, 1646, 1648, 1650, 1652, 1654, 1656, 1658, 1660, 1662, 1664, 1666, 1668, 1670, 1672, 1674, 1676, 1678, 1680, 1682, 1684, 1686, 1688, 1690, 1692, 1694, 1696, 1698, 1700, 1702, 1704, 1706, 1708, 1710, 1712, 1714, 1716, 1718, 1720, 1722, 1724, 1726, 1728, 1730, 1732, 1734, 1736, 1738, 1740, 1742, 1744, 1746, 1748, 1750, 1752, 1754, 1756, 1758, 1760, 1762, 1764, 1766, 1768, 1770, 1772, 1774, 1776, 1778, 1780, 1782, 1784, 1786, 1788, 1790, 1792, 1794, 1796, 1798, 1800, 1802, 1804, 1806, 1808, 1810, 1812, 1814, 1816, 1818, 1820, 1822, 1824, 1826, 1828, 1830, 1832, 1834, 1836, 1838, 1840, 1842, 1844, 1846, 1848, 1850, 1852, 1854, 1856, 1858, 1860, 1862, 1864, 1866, 1868, 1870, 1872, 1874, 1876, 1878, 1880, 1882, 1884, 1886, 1888, 1890, 1892, 1894, 1896, 1898, 1900, 1902, 1904, 1906, 1908, 1910, 1912, 1914, 1916, 1918, 1920, 1922, 1924, 1926, 1928, 1930, 1932, 1934, 1936, 1938, 1940, 1942, 1944, 1946, 1948, 1950, 1952, 1954, 1956, 1958, 1960, 1962, 1964, 1966, 1968, 1970, 1972, 1974, 1976, 1978, 1980, 1982, 1984, 1986, 1988, 1990, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022, 2024, 2026, 2028, 2030, 2032, 2034, 2036, 2038, 2040, 2042, 2044, 2046, 2048, 2050, 2052, 2054, 2056, 2058, 2060, 2062, 2064, 2066, 2068, 2070, 2072, 2074, 2076, 2078, 2080, 2082, 2084, 2086, 2088, 2090, 2092, 2094, 2096, 2098, 2100, 2102, 2104, 2106, 2108, 2110, 2112, 2114, 2116, 2118, 2120, 2122, 2124, 2126, 2128, 2130, 2132, 2134, 2136, 2138, 2140, 2142, 2144, 2146, 2148, 2150, 2152, 2154, 2156, 2158, 2160, 2162, 2164, 2166, 2168, 2170, 2172, 2174, 2176, 2178, 2180, 2182, 2184, 2186, 2188, 2190, 2192, 2194, 2196, 2198, 2200, 2202, 2204, 2206, 2208, 2210, 2212, 2214, 2216, 2218, 2220, 2222, 2224, 2226, 2228, 2230, 2232, 2234, 2236, 2238, 2240, 2242, 2244, 2246, 2248, 2250, 2252, 2254, 2256, 2258, 2260, 2262, 2264, 2266, 2268, 2270, 2272, 2274, 2276, 2278, 2280, 2282, 2284, 2286, 2288, 2290, 2292, 2294, 2296, 2298, 2300, 2302, 2304, 2306, 2308, 2310, 2312, 2314, 2316, 2318, 2320, 2322, 2324, 2326, 2328, 2330, 2332, 2334, 2336, 2338, 2340, 2342, 2344, 2346, 2348, 2350, 2352, 2354, 2356, 2358, 2360, 2362, 2364, 2366, 2368, 2370, 2372, 2374, 2376, 2378, 2380, 2382, 2384, 2386, 2388, 2390, 2392, 2394, 2396, 2398, 2400, 2402, 2404, 2406, 2408, 2410, 2412, 2414, 2416, 2418, 2420, 2422, 2424, 2426, 2428, 2430, 2432, 2434, 2436, 2438, 2440, 2442, 2444, 2446, 2448, 2450, 2452, 2454, 2456, 2458, 2460, 2462, 2464, 2466, 2468, 2470, 2472, 2474, 2476, 2478, 2480, 2482, 2484, 2486, 2488, 2490, 2492, 2494, 2496, 2498, 2500, 2502, 2504, 2506, 2508, 2510, 2512, 2514, 2516, 2518, 2520, 2522, 2524, 2526, 2528, 2530, 2532, 2534, 2536, 2538, 2540, 2542, 2544, 2546, 2548, 2550, 2552, 2554, 2556, 2558, 2560, 2562, 2564, 2566, 2568, 2570, 2572, 2574, 2576, 2578, 2580, 2582, 2584, 2586, 2588, 2590, 2592, 2594, 2596, 2598, 2600, 2602, 2604, 2606, 2608, 2610, 2612, 2614, 2616, 2618, 2620, 2622, 2624, 2626, 2628, 2630, 2632, 2634, 2636, 2638, 2640, 2642, 2644, 2646, 2648, 2650, 2652, 2654, 2656, 2658, 2660, 2662, 2664, 2666, 2668, 2670, 2672, 2674, 2676, 2678, 2680, 2682, 2684, 2686, 2688, 2690, 2692, 2694, 2696, 2698, 2700, 2702, 2704, 2706, 2708, 2710, 2712, 2714, 2716, 2718, 2720, 2722, 2724, 2726, 2728, 2730, 2732, 2734, 2736, 2738, 2740, 2742, 2744, 2746, 2748, 2750, 2752, 2754, 2756, 2758, 2760, 2762, 2764, 2766, 2768, 2770, 2772, 2774, 2776, 2778, 2780, 2782, 2784, 2786, 2788, 2790, 2792, 2794, 2796, 2798, 2800, 2802, 2804, 2806, 2808, 2810, 2812, 2814, 2816, 2818, 2820, 2822, 2824, 2826, 2828, 2830, 2832, 2834, 2836, 2838, 2840, 2842, 2844, 2846, 2848, 2850, 2852, 2854, 2856, 2858, 2860, 2862, 2864, 2866, 2868, 2870, 2872, 2874, 2876, 2878, 2880, 2882, 2884, 2886, 2888, 2890, 2892, 2894, 2896, 2898, 2900, 2902, 2904, 2906, 2908, 2910, 2912, 2914, 2916, 2918, 2920, 2922, 2924, 2926, 2928, 2930, 2932, 2934, 2936, 2938, 2940, 2942, 2944, 2946, 2948, 2950, 2952, 2954, 2956, 2958, 2960, 2962, 2964, 2966, 2968, 2970, 2972, 2974, 2976, 2978, 2980, 2982, 2984, 2986, 2988, 2990, 2992, 2994, 2996, 2998, 3000, 3002, 3004, 3006, 3008, 3010, 3012, 3014, 3016, 3018, 3020, 3022, 3024, 3026, 3028, 3030, 3032, 3034, 3036, 3038, 3040, 3042, 3044, 3046, 3048, 3050, 3052, 3054, 3056, 3058, 3060, 3062, 3064, 3066, 3068, 3070, 3072, 3074, 3076, 3078, 3080, 3082, 3084, 3086, 3088, 3090, 3092, 3094, 3096, 3098, 3100, 3102, 3104, 3106, 3108, 3110, 3112, 3114, 3116, 3118, 3120, 3122, 3124, 3126, 3128, 3130, 3132, 3134, 3136, 3138, 3140, 3142, 3144, 3146, 3148,

3150, 3152, 3154, 3156, 3158, 3160, 3162, 3164, 3166, 3168, 3170, 3172, 3174, 3176, 3178, 3180, 3182, 3184, 3186, 3188, 3190, 3192, 3194, 3196, 3198, 3200, 3202, 3204, 3206, 3208, 3210, 3212, 3214, 3216, 3218, 3220, 3222, 3224, 3226, 3228, 3230, 3232, 3234, 3236, 3238, 3240, 3242, 3244, 3246, 3248, 3250, 3252, 3254, 3256, 3258, 3260, 3262, 3264, 3266, 3268, 3270, 3272, 3274, 3276, 3278, 3280, 3282, 3284, 3286, 3288, 3290, 3292, 3294, 3296, 3298, 3300, 3302, 3304, 3306, 3308, 3310, 3312, 3314, 3316, 3318, 3320, 3322, 3324, 3326, 3328, 3330, 3332, 3334, 3336, 3338, 3340, 3342, 3344, 3346, 3348, 3350, 3352, 3354, 3356, 3358, 3360, 3362, 3364, 3366, 3368, 3370, 3372, 3374, 3376, 3378, 3380, 3382, 3384, 3386, 3388, 3390, 3392, 3394, 3396, 3398, 3400, 3402, 3404, 3406, 3408, 3410, 3412, 3414, 3416, 3418, 3420, 3422, 3424, 3426, 3428, 3430, 3432, 3434, 3436, 3438, 3440, 3442, 3444, 3446, 3448, 3450, 3452, 3454, 3456, 3458, 3460, 3462, 3464, 3466, 3468, 3470, 3472, 3474, 3476, 3478, 3480, 3482, 3484, 3486, 3488, 3490, 3492, 3494, 3496, 3498, 3500, 3502, 3504, 3506, 3508, 3510, 3512, 3514, 3516, 3518, 3520, 3522, 3524, 3526, 3528, 3530, 3532, 3534, 3536, 3538, 3540, 3542, 3544, 3546, 3548, 3550, 3552, 3554, 3556, 3558, 3560, 3562, 3564, 3566, 3568, 3570, 3572, 3574, 3576, 3578, 3580, 3582, 3584, 3586, 3588, 3590, 3592, 3594, 3596, 3598, 3600, 3602, 3604, 3606, 3608, 3610, 3612, 3614, 3616, 3618, 3620, 3622, 3624, 3626, 3628, 3630, 3632, 3634, 3636, 3638, 3640, 3642, 3644, 3646, 3648, 3650, 3652, 3654, 3656, 3658, 3660, 3662, 3664, 3666, 3668, 3670, 3672, 3674, 3676, 3678, 3680, 3682, 3684, 3686, 3688, 3690, 3692, 3694, 3696, 3698, 3700, 3702, 3704, 3706, 3708, 3710, 3712, 3714, 3716, 3718, 3720, 3722, 3724, 3726, 3728, 3730, 3732, 3734, 3736, 3738, 3740, 3742, 3744, 3748, 3750, 3752, 3754, 3756, 3758, 3760, 3762, 3764, 3766, 3768, 3770, 3772, 3774, 3776, 3778, 3780, 3782, 3784, 3786, 3788, 3790, 3792, 3794, 3796, 3798, 3800, 3802, 3804, 3806, 3808, 3810, 3812, 3814, 3816, 3818, 3820, 3822, 3824, 3826, 3828, 3830, 3832, 3834, 3836, 3838, 3840, 3842, 3844, 3846, 3848, 3850, 3852, 3854, 3856, 3858, 3860, 3862, 3864, 3866, 3868, 3870, 3872, 3874, 3876, 3878, 3880, 3882, 3884, 3886, 3888, 3890, 3892, 3894, 3896, 3898, 3900, 3902, 3904, 3906, 3908, 3910, 3912, 3914, 3916, 3918, 3920, 3924, 3926, 3928, 3930, 3932, 3934, 3936, 3938, 3940, 3942, 3944, 3946, 3948, 3950, 3952, 3954, 3956, 3958, 3960, 3962, 3964, 3966, 3968, 3970, 3972, 3974, 3976, 3978, 3980, 3982, 3984, 3986, 3988, 3990, 3992, 3994, 3996, 3998, 4000, 4002, 4004, 4006, 4008, 4010, 4012, 4014, 4016, 4018, 4020, 4022, 4024, 4026, 4028, 4030, 4032, 4034, 4036, 4038, 4040, 4042, 4044, 4046, 4048, 4050, 4052, 4054, 4056, 4058, 4060, 4062, 4064, 4066, 4068, 4070, 4072, 4074, 4076, 4078, 4080, 4082, 4084, 4086, 4088, 4090, 4092, 4094, 4096, 4098, 4100, 4102, 4104, 4106, 4108, 4110, 4112, 4114, 4116, 4118, 4120, 4122, 4124, 4126, 4128, 4130, 4132, 4134, 4136, 4138, 4140, 4142, 4144, 4146, 4148, 4150, 4152, 4154, 4156, 4158, 4160, 4162, 4164, 4166, 4168, 4170, 4172, 4174, 4176, 4178, 4180, 4182, 4184, 4186, 4188, 4190, 4192, 4194, 4196, 4198, 4200, 4202, 4204, 4206, 4208, 4210, 4212, 4214, 4216, 4218, 4220, 4222, 4224, 4226, 4228, 4230, 4232, 4234, 4236, 4238, 4240, 4242, 4244, 4246, 4248, 4250, 4252, 4254, 4256, 4258, 4260, 4262, 4264, 4266, 4268, 4270, 4272, 4274, 4276, 4278, 4280, 4282, 4284, 4286, 4288, 4290, 4292, 4294, 4296, 4298, 4300, 4302, 4304, 4306, 4308, 4310, 4312, 4314, 4316, 4318, 4320, 4322, 4324, 4326, 4328, 4330, 4332, 4334, 4336, 4338, 4340, 4342, 4344, 4346, 4348, 4350, 4352, 4354, 4356, 4358, 4360, 4362, 4364, 4366, 4368, 4370, 4372, 4374, 4376, 4378, 4380, 4382, 4384, 4386, 4388, 4390, 4392, 4394, 4396, 4398, 4400, 4402, 4404, 4406, 4408, 4410, 4412, 4414, 4416, 4418, 4420, 4422, 4424, 4426, 4428, 4430, 4432, 4434, 4436, 4438, 4440, 4442, 4444, 4446, 4448, 4450, 4452, 4454, 4456, 4458, 4460, 4462, 4464, 4466, 4468, 4470, 4472, 4474, 4476, 4478, 4480, 4482, 4484, 4486, 4488, 4490, 4492, 4494, 4496, 4498, 4500, 4502, 4504, 4506, 4508, 4510, 4512, 4514, 4516, 4518, 4520, 4522, 4524, 4526, 4528, 4530, 4532, 4534, 4536, 4538, 4540, 4542, 4544, 4546, 4548, 4550, 4552, 4554, 4556, 4558, 4560, 4562, 4564, 4566, 4568, 4570, 4572, 4574, 4576, 4578, 4580, 4582, 4584, 4586, 4588, 4590, 4592, 4594, 4596, 4598, 4600, 4602, 4604, 4606, 4608, 4610, 4612, 4614, 4616, 4618, 4620, 4622, 4624, 4626, 4628, 4630, 4632, 4634, 4636, 4638, 4640, 4642, 4644, 4646, 4648, 4650, 4652, 4654, 4656, 4658, 4660, 4662, 4664, 4666, 4668, 4670, 4672, 4674, 4676, 4678, 4680, 4682, 4684, 4686, 4688, 4690, 4692, 4694, 4696, 4698, 4700, 4702, 4704, 4706,

4708, 4710, 4712, 4714, 4716, 4718, 4720, 4722, 4724, 4726, 4728, 4730, 4732, 4734, 4736, 4738, 4740, 4742, 4744, 4746, 4748, 4750, 4752, 4754, 4756, 4758, 4760, 4762, 4764, 4766, 4768, 4770, 4772, 4774, 4776, 4778, 4780, 4782, 4784, 4786, 4788, 4790, 4792, 4794, 4796, 4798, 4800, 4802, 4804, 4806, 4808, 4810, 4812, 4814, 4816, 4818, 4820, 4822, 4824, 4826, 4828, 4830, 4832, 4834, 4836, 4838, 4840, 4842, 4844, 4846, 4848, 4850, 4852, 4854, 4856, 4858, 4860, 4862, 4864, 4866, 4868, 4870, 4872, 4874, 4876, 4878, 4880, 4882, 4884, 4886, 4888, 4890, 4892, 4894, 4896, 4898, 4900, 4902, 4904, 4906, 4908, 4910, 4912, 4914, 4916, 4918, 4920, 4922, 4924, 4926, 4928, 4930, 4932, 4934, 4936, 4938, 4940, 4942, 4944, 4946, 4948, 4950, 4952, 4954, 4956, 4958, 4960, 4962, 4964, 4966, 4968, 4970, 4972, 4974, 4976, 4978, 4980, 4982, 4984, 4986, 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5002, 5004, 5006, 5008, 5010, 5012, 5014, 5016, 5018, 5020, 5022, 5024, 5026, 5028, 5030, 5032, 5034, 5036, 5038, 5040, 5042, 5044, 5046, 5048, 5050, 5052, 5054, 5056, 5058, 5060, 5062, 5064, 5066, 5068, 5070, 5072, 5074, 5076, 5078, 5080, 5082, 5084, 5086, 5088, 5090, 5092, 5094, 5096, 5098, 5100, 5102, 5104, 5106, 5108, 5110, 5112, 5114, 5116, 5118, 5120, 5122, 5124, 5126, 5128, 5130, 5132, 5134, 5136, 5138, 5140, 5142, 5144, 5146, 5148, 5150, 5152, 5154, 5156, 5158, 5160, 5162, 5164, 5166, 5168, 5170, 5172, 5174, 5176, 5178, 5180, 5182, 5184, 5186, 5188, 5190, 5192, 5194, 5196, 5198, 5200, 5202, 5204, 5206, 5208, 5210, 5212, 5214, 5216, 5218, 5220, 5222, 5224, 5226, 5228, 5230, 5232, 5234, 5236, 5238, 5240, 5242, 5244, 5246, 5248, 5250, 5252, 5254, 5256, 5258, 5260, 5262, 5264, 5266, 5268, 5270, 5272, 5274, 5276, 5278, 5280, 5282, 5284, 5286, 5288, 5290, 5292, 5294, 5296, 5298, 5300, 5302, 5304, 5306, 5308, 5310, 5312, 5314, 5316, 5318, 5320, 5322, 5324, 5326, 5328, 5330, 5332, 5334, 5336, 5338, 5340, 5342, 5344, 5346, 5348, 5350, 5352, 5354, 5356, 5358, 5360, 5362, 5364, 5366, 5368, 5370, 5372, 5374, 5376, 5378, 5380, 5382, 5384, 5386, 5388, 5390, 5392, 5394, 5396, 5398, 5400, 5402, 5404, 5406, 5408, 5410, 5412, 5414, 5416, 5418, 5420, 5422, 5424, 5426, 5428, 5430, 5432, 5434, 5436, 5438, 5440, 5442, 5444, 5446, 5448, 5450, 5452, 5454, 5456, 5458, 5460, 5462, 5464, 5466, 5468, 5470, 5472, 5474, 5476, 5478, 5480, 5482, 5484, 5486, 5488, 5490, 5492, 5494, 5496, 5498, 5500, 5502, 5504, 5506, 5508, 5510, 5512, 5514, 5516, 5518, 5520, 5522, 5524, 5526, 5528, 5530, 5532, 5534, 5536, 5538, 5540, 5542, 5544, 5546, 5548, 5550, 5552, 5554, 5556, 5558, 5560, 5562, 5564, 5566, 5568, 5570, 5572, 5574, 5576, 5578, 5580, 5582, 5584, 5586, 5588, 5590, 5592, 5594, 5596, 5598, 5600, 5602, 5604, 5606, 5608, 5610, 5612, 5614, 5616, 5618, 5620, 5622, 5624, 5626, 5628, 5630, 5632, 5634, 5636, 5638, 5640, 5642, 5644, 5646, 5648, 5650, 5652, 5654, 5656, 5658, 5660, 5662, 5664, 5666, 5668, 5670, 5672, 5674, 5676, 5678, 5680, 5682, 5684, 5686, 5688, 5690, 5692, 5694, 5696, 5698, 5700, 5702, 5704, 5706, 5708, 5710, 5712, 5714, 5716, 5718, 5720, 5722, 5724, 5726, 5728, 5730, 5732, 5734, 5736, 5738, 5740, 5742, 5744, 5746, 5748, 5750, 5752, 5754, 5756, 5758, 5760, 5762, 5764, 5766, 5768, 5770, 5772, 5774, 5776, 5778, 5780, 5782, 5784, 5786, 5788, 5790, 5792, 5794, 5796, 5798, 5800, 5802, 5804, 5806, 5808, 5810, 5812, 5814, 5816, 5818, 5820, 5822, 5824, 5826, 5828, 5830, 5832, 5834, 5836, 5838, 5840, 5842, 5844, 5846, 5848, 5850, 5852, 5854, 5856, 5858, 5860, 5862, 5864, 5866, 5868, 5870, 5872, 5874, 5876, 5878, 5880, 5882, 5884, 5886, 5888, 5890, 5892, 5894, 5896, 5898, 5900, 5902, 5904, 5906, 5908, 5910, 5912, 5914, 5916, 5918, 5920, 5922, 5924, 5926, 5928, 5930, 5932, 5934, 5936, 5938, 5940, 5942, 5944, 5946, 5948, 5950, 5952, 5954, 5956, 5958, 5960, 5962, 5964, 5966, 5968, 5970, 5972, 5974, 5976, 5978, 5980, 5982, 5984, 5986, 5988, 5990, 5992, 5994, 5996, 5998, 6000, 6002, 6004, 6006, 6008, 6010, 6012, 6014, 6016, 6018, 6020, 6022, 6024, 6026, 6028, 6030, 6032, 6034, 6036, 6038, 6040, 6042, 6044, 6046, 6048, 6050, 6052, 6054, 6056, 6058, 6060, 6062, 6064, 6066, 6068, 6070, 6072, 6074, 6076, 6078, 6080, 6082, 6084, 6086, 6088, 6090, 6092, 6094, 6096, 6098, 6100, 6102, 6104, 6106, 6108, 6110, 6112, 6114, 6116, 6118, 6120, 6122, 6124, 6126, 6128, 6130, 6132, 6134, 6136, 6138, 6140, 6142, 6144, 6146, 6148, 6150, 6152, 6154, 6156, 6158, 6160, 6162, 6164, 6166, 6168, 6170, 6172, 6174, 6176, 6178, 6180, 6182, 6184, 6186, 6188, 6190, 6192, 6194, 6196, 6198, 6200, 6202, 6204, 6206, 6208, 6210, 6212, 6214, 6216, 6218, 6220, 6222, 6224, 6226, 6228, 6230, 6232, 6234, 6236, 6238, 6240, 6242, 6244, 6246, 6248, 6250, 6252, 6254, 6256, 6258, 6260, 6262, 6264,

6266, 6268, 6270, 6272, 6274, 6276, 6278, 6280, 6282, 6284, 6286, 6288, 6290, 6292, 6294, 6296, 6298, 6300, 6302, 6304, 6306, 6308, 6310, 6312, 6314, 6316, 6318, 6320, 6322, 6324, 6326, 6328, 6330, 6332, 6334, 6336, 6338, 6340, 6342, 6344, 6346, 6348, 6350, 6352, 6354, 6356, 6358, 6360, 6362, 6364, 6366, 6368, 6370, 6372, 6374, 6376, 6378, 6380, 6382, 6384, 6386, 6388, 6390, 6392, 6394, 6396, 6398, 6400, 6402, 6404, 6406, 6408, 6410, 6412, 6414, 6416, 6418, 6420, 6422, 6424, 6426, 6428, 6430, 6432, 6434, 6436, 6438, 6440, 6442, 6444, 6446, 6448, 6450, 6452, 6454, 6456, 6458, 6460, 6462, 6464, 6466, 6468, 6470, 6472, 6474, 6476, 6478, 6480, 6482, 6484, 6486, 6488, 6490, 6492, 6494, 6496, 6498, 6500, 6502, 6504, 6506, 6508, 6510, 6512, 6514, 6516, 6518, 6520, 6522, 6524, 6526, 6528, 6530, 6532, 6534, 6536, 6538, 6540, 6542, 6544, 6546, 6548, 6550, 6552, 6554, 6556, 6558, 6560, 6562, 6564, 6566, 6568, 6570, 6572, 6574, 6576, 6578, 6580, 6582, 6584, 6586, 6588, 6590, 6592, 6594, 6596, 6598, 6600, 6602, 6604, 6606, 6608, 6610, 6612, 6614, 6616, 6618, 6620, 6622, 6624, 6626, 6628, 6630, 6632, 6634, 6636, 6638, 6640, 6642, 6644, 6646, 6648, 6650, 6652, 6654, 6656, 6658, 6660, 6662, 6664, 6666, 6668, 6670, 6672, 6674, 6676, 6678, 6680, 6682, 6684, 6686, 6688, 6690, 6692, 6694, 6696, 6698, 6700, 6702, 6704, 6706, 6708, 6710, 6712, 6714, 6716, 6718, 6720, 6722, 6724, 6726, 6728, 6730, 6732, 6734, 6736, 6738, 6740, 6742, 6744, 6746, 6748, 6750, 6752, 6754, 6756, 6758, 6760, 6762, 6764, 6766, 6768, 6770, 6772, 6774, 6776, 6778, 6780, 6782, 6784, 6786, 6788, 6790, 6792, 6794, 6796, 6798, 6800, 6802, 6804, 6806, 6808, 6810, 6812, 6814, 6816, 6818, 6820, 6822, 6824, 6826, 6828, 6830, 6832, 6834, 6836, 6838, 6840, 6842, 6844, 6846, 6848, 6850, 6852, 6854, 6856, 6858, 6860, 6862, 6864, 6866, 6868, 6870, 6872, 6874, 6876, 6878, 6880, 6882, 6884, 6886, 6888, 6890, 6892, 6894, 6896, 6898, 6900, 6902, 6904, 6906, 6908, 6910, 6912, 6914, 6916, 6918, 6920, 6922, 6924, 6926, 6928, 6930, 6932, 6934, 6936, 6938, 6940, 6942, 6944, 6946, 6948, 6950, 6952, 6954, 6956, 6958, 6960, 6962, 6964, 6966, 6968, 6970, 6972, 6974, 6976, 6978, 6980, 6982, 6984, 6986, 6988, 6990, 6992, 6994, 6996, 6998, 7000, 7002, 7004, 7006, 7008, 7010, 7012, 7014, 7016, 7018, 7020, 7022, 7024, 7026, 7028, 7030, 7032, 7034, 7036, 7038, 7040, 7042, 7044, 7046, 7048, 7050, 7052, 7054, 7056, 7058, 7060, 7062, 7064, 7066, 7068, 7070, 7072, 7074, 7076, 7078, 7080, 7082, 7084, 7086, 7088, 7090, 7092, 7094, 7096, 7098, 7100, 7102, 7104, 7106, 7108, 7110, 7112, 7114, 7116, 7118, 7120, 7122, 7124, 7126, 7128, 7130, 7132, 7134, 7136, 7138, 7140, 7142, 7144, 7146, 7148, 7150, 7152, 7154, 7156, 7158, 7160, 7162, 7164, 7166, 7168, 7170, 7172, 7174, 7176, 7178, 7180, 7182, 7184, 7186, 7188, 7190, 7192, 7194, 7196, 7198, 7200, 7202, 7204, 7206, 7208, 7210, 7212, 7214, 7216, 7218, 7220, 7222, 7224, 7226, 7228, 7230, 7232, 7234, 7236, 7238, 7240, 7242, 7244, 7246, 7248, 7250, 7252, 7254, 7256, 7258, 7260, 7262, 7264, 7266, 7268, 7270, 7272, 7274, 7276, 7278, 7280, 7282, 7284, 7286, 7288, 7290, 7292, 7294, 7296, 7298, 7300, 7302, 7304, 7306, 7308, 7310, 7312, 7314, 7316, 7318, 7320, 7322, 7324, 7326, 7328, 7330, 7332, 7334, 7336, 7338, 7340, 7342, 7344, 7346, 7348, 7350, 7352, 7354, 7356, 7358, 7360, 7362, 7364, 7366, 7368, 7370, 7372, 7374, 7376, 7378, 7380, 7382, 7384, 7386, 7388, 7390, 7392, 7394, 7396, 7398, 7400, 7402, 7404, 7406, 7408, 7410, 7412, 7414, 7416, 7418, 7420, 7422, 7424, 7426, 7428, 7430, 7432, 7434, 7436, 7438, 7440, 7442, 7444, 7446, 7448, 7450, 7452, 7454, 7456, 7458, 7460, 7462, 7464, 7468, 7470, 7472, 7474, 7476, 7478, 7480, 7482, 7484, 7486, 7488, 7490, 7492, 7494, 7496, 7498, 7500, 7502, 7504, 7506, 7508, 7510, 7512, 7514, 7516, 7518, 7520, 7522, 7524, 7526, 7528, 7530, 7532, 7534, 7536, 7538, 7540, 7542, 7544, 7546, 7548, 7550, 7552, 7554, 7556, 7558, 7560, 7562, 7564, 7566, 7568, 7570, 7572, 7574, 7576, 7578, 7580, 7582, 7584, 7586, 7588, 7590, 7592, 7594, 7596, 7598, 7600, 7602, 7604, 7606, 7608, 7610, 7612, 7614, 7616, 7618, 7620, 7622, 7624, 7626, 7628, 7630, 7632, 7634, 7636, 7638, 7640, 7642, 7644, 7646, 7648, 7650, 7652, 7654, 7656, 7658, 7660, 7662, 7664, 7666, 7668, 7670, 7672, 7674, 7676, 7678, 7680, 7682, 7684, 7686, 7688, 7690, 7692, 7694, 7696, 7698, 7700, 7702, 7704, 7706, 7708, 7710, 7712, 7714, 7716, 7718, 7720, 7722, 7724, 7726, 7728, 7730, 7732, 7734, 7736, 7738, 7740, 7742, 7744, 7746, 7748, 7750, 7752, 7754, 7756, 7758, 7760, 7762, 7764, 7766, 7768, 7770, 7772, 7774, 7776, 7778, 7780, 7782, 7784, 7786, 7788, 7790, 7792, 7794, 7796, 7798, 7800, 7802, 7804, 7806, 7808, 7810, 7812, 7814, 7816, 7818, 7820, 7822,

7824, 7826, 7828, 7830, 7832, 7834, 7836, 7838, 7840, 7842, 7844, 7846, 7848, 7850, 7852, 7854, 7856, 7858, 7860, 7862, 7864, 7866, 7868, 7870, 7872, 7874, 7876, 7878, 7880, 7882, 7884, 7886, 7888, 7890, 7892, 7894, 7896, 7898, 7900, 7902, 7904, 7906, 7908, 7910, 7912, 7914, 7916, 7918, 7920, 7922, 7924, 7926, 7928, 7930, 7932, 7934, 7936, 7938, 7940, 7942, 7944, 7946, 7948, 7950, 7952, 7954, 7956, 7958, 7960, 7962, 7964, 7966, 7968, 7970, 7972, 7974, 7976, 7978, 7980, 7982, 7984, 7986, 7988, 7990, 7992, 7994, 7996, 7998, 8000, 8002, 8004, 8006, 8008, 8010, 8012, 8014, 8016, 8018, 8020, 8022, 8024, 8026, 8028, 8030, 8032, 8034, 8036, 8038, 8040, 8042, 8044, 8046, 8048, 8050, 8052, 8054, 8056, 8058, 8060, 8062, 8064, 8066, 8068, 8070, 8072, 8074, 8076, 8078, 8080, 8082, 8084, 8086, 8088, 8090, 8092, 8094, 8096, 8098, 8100, 8102, 8104, 8106, 8108, 8110, 8112, 8114, 8116, 8118, 8120, 8122, 8124, 8126, 8128, 8130, 8132, 8134, 8136, 8138, 8140, 8142, 8144, 8146, 8148, 8150, 8152, 8154, 8156, 8158, 8160, 8162, 8164, 8166, 8168, 8170, 8172, 8174, 8176, 8178, 8180, 8182, 8184, 8186, 8188, 8190, 8192, 8194, 8196, 8198, 8200, 8202, 8204, 8206, 8208, 8210, 8212, 8214, 8216, 8218, 8220, 8222, 8224, 8226, 8228, 8230, 8232, 8234, 8236, 8238, 8240, 8242, 8244, 8246, 8248, 8250, 8252, 8254, 8256, 8258, 8260, 8262, 8264, 8266, 8268, 8270, 8272, 8274, 8276, 8278, 8280, 8282, 8284, 8286, 8288, 8290, 8292, 8294, 8296, 8298, 8300, 8302, 8304, 8306, 8308, 8310, 8312, 8314, 8316, 8318, 8320, 8322, 8324, 8326, 8328, 8330, 8332, 8334, 8336, 8338, 8340, 8342, 8344, 8346, 8348, 8350, 8352, 8354, 8356, 8358, 8360, 8362, 8364, 8366, 8368, 8370, 8372, 8374, 8376, 8378, 8380, 8382, 8384, 8386, 8388, 8390, 8392, 8394, 8396, 8398, 8400, 8402, 8404, 8406, 8408, 8410, 8412, 8414, 8416, 8418, 8420, 8422, 8424, 8426, 8428, 8430, 8432, 8434, 8436, 8438, 8440, 8442, 8444, 8446, 8448, 8450, 8452, 8454, 8456, 8458, 8460, 8462, 8464, 8466, 8468, 8470, 8472, 8474, 8476, 8478, 8480, 8482, 8484, 8486, 8488, 8490, 8492, 8494, 8496, 8498, 8500, 8502, 8504, 8506, 8508, 8510, 8512, 8514, 8516, 8518, 8520, 8522, 8524, 8526, 8528, 8530, 8532, 8534, 8536, 8538, 8540, 8542, 8544, 8546, 8548, 8550, 8552, 8554, 8556, 8558, 8560, 8562, 8564, 8566, 8568, 8570, 8572, 8574, 8576, 8578, 8580, 8582, 8584, 8586, 8588, 8590, 8592, 8594, 8596, 8598, 8600, 8602, 8604, 8606, 8608, 8610, 8612, 8614, 8616, 8618, 8620, 8622, 8624, 8626, 8628, 8630, 8632, 8634, 8636, 8638, 8640, 8642, 8644, 8646, 8648, 8650, 8652, 8654, 8656, 8658, 8660, 8662, 8664, 8666, 8668, 8670, 8672, 8674, 8676, 8678, 8680, 8682, 8684, 8686, 8688, 8690, 8692, 8694, 8696, 8698, 8700, 8702, 8704, 8706, 8708, 8710, 8712, 8714, 8716, 8718, 8720, 8722, 8724, 8726, 8728, 8730, 8732, 8734, 8736, 8738, 8740, 8742, 8744, 8746, 8748, 8750, 8752, 8754, 8756, 8758, 8760, 8762, 8764, 8766, 8768, 8770, 8772, 8774, 8776, 8778, 8782, 8784, 8786, 8788, 8790, 8792, 8794, 8796, 8798, 8800, 8802, 8804, 8806, 8808, 8810, 8812, 8814, 8816, 8818, 8820, 8822, 8824, 8826, 8828, 8830, 8832, 8834, 8836, 8838, 8840, 8842, 8844, 8846, 8848, 8850, 8852, 8854, 8856, 8858, 8860, 8862, 8864, 8866, 8868, 8870, 8872, 8874, 8876, 8878, 8880, 8882, 8884, 8886, 8888, 8890, 8892, 8894, 8896, 8898, 8900, 8902, 8904, 8906, 8908, 8910, 8912, 8914, 8916, 8918, 8920, 8922, 8924, 8926, 8928, 8930, 8932, 8934, 8936, 8938, 8940, 8942, 8944, 8946, 8948, 8950, 8952, 8954, 8956, 8958, 8960, 8962, 8964, 8966, 8968, 8970, 8972, 8974, 8976, 8978, 8980, 8982, 8984, 8986, 8988, 8990, 8992, 8994, 8996, 8998, 9000, 9002, 9004, 9006, 9008, 9010, 9012, 9014, 9016, 9018, 9020, 9022, 9024, 9026, 9028, 9030, 9032, 9034, 9036, 9038, 9040, 9042, 9044, 9046, 9048, 9050, 9052, 9054, 9056, 9058, 9060, 9062, 9064, 9066, 9068, 9070, 9072, 9074, 9076, 9078, 9080, 9082, 9084, 9086, 9088, 9090, 9092, 9094, 9096, 9098, 9100, 9102, 9104, 9106, 9108, 9110, 9112, 9114, 9116, 9118, 9120, 9122, 9124, 9126, 9128, 9130, 9132, 9134, 9136, 9138, 9140, 9142, 9144, 9146, 9148, 9150, 9152, 9154, 9156, 9158, 9160, 9162, 9164, 9166, 9168, 9170, 9172, 9174, 9176, 9178, 9180, 9182, 9184, 9186, 9190, 9192, 9194, 9196, 9198, 9200, 9202, 9204, 9206, 9208, 9210, 9212, 9214, 9216, 9218, 9220, 9222, 9224, 9226, 9228, 9230, 9232, 9234, 9236, 9238, 9240, 9242, 9244, 9246, 9248, 9250, 9252, 9254, 9256, 9258, 9260, 9262, 9264, 9266, 9268, 9270, 9272, 9274, 9276, 9278, 9280, 9282, 9284, 9286, 9288, 9290, 9292, 9294, 9296, 9298, 9300, 9302, 9304, 9306, 9308, 9310, 9312, 9314, 9316, 9318, 9320, 9322, 9324, 9326, 9328, 9330, 9332, 9334, 9336, 9338, 9340, 9342, 9344, 9346, 9348, 9350, 9352, 9354, 9356, 9358, 9360, 9362, 9364, 9366, 9368, 9370, 9372, 9374, 9376, 9378, 9380,

9382, 9384, 9386, 9388, 9390, 9392, 9394, 9396, 9398, 9400, 9402, 9404, 9406, 9408, 9410, 9412, 9414, 9416, 9418, 9420, 9422, 9424, 9426, 9428, 9430, 9432, 9434, 9436, 9438, 9440, 9442, 9444, 9446, 9448, 9450, 9452, 9454, 9456, 9458, 9460, 9462, 9464, 9466, 9468, 9470, 9472, 9474, 9476, 9478, 9480, 9482, 9484, 9486, 9488, 9490, 9492, 9494, 9496, 9498, 9500, 9502, 9504, 9506, 9508, 9510, 9512, 9514, 9516, 9518, 9520, 9522, 9524, 9526, 9528, 9530, 9532, 9534, 9536, 9538, 9540, 9542, 9544, 9546, 9548, 9550, 9552, 9554, 9556, 9558, 9560, 9562, 9564, 9566, 9568, 9570, 9572, 9574, 9576, 9578, 9580, 9582, 9584, 9586, 9588, 9590, 9592, 9594, 9596, 9598, 9600, 9602, 9604, 9606, 9608, 9610, 9612, 9614, 9616, 9618, 9620, 9622, 9624, 9626, 9628, 9630, 9632, 9634, 9636, 9638, 9640, 9642, 9644, 9646, 9648, 9650, 9652, 9654, 9656, 9658, 9660, 9662, 9664, 9666, 9668, 9670, 9672, 9674, 9676, 9678, 9680, 9682, 9684, 9686, 9688, 9690, 9692, 9694, 9696, 9698, 9700, 9702, 9704, 9706, 9708, 9710, 9712, 9714, 9716, 9718, 9720, 9722, 9724, 9726, 9728, 9730, 9732, 9734, 9736, 9738, 9740, 9742, 9744, 9746, 9748, 9750, 9752, 9754, 9756, 9758, 9760, 9762, 9764, 9766, 9768, 9770, 9772, 9774, 9776, 9778, 9780, 9782, 9784, 9786, 9788, 9790, 9792, 9794, 9796, 9798, 9800, 9802, 9804, 9806, 9808, 9810, 9812, 9814, 9816, 9818, 9820, 9822, 9824, 9826, 9828, 9830, 9832, 9834, 9836, 9838, 9840, 9842, 9844, 9846, 9848, 9850, 9852, 9854, 9856, 9858, 9860, 9862, 9864, 9866, 9868, 9870, 9872, 9874, 9876, 9878, 9880, 9882, 9884, 9886, 9888, 9890, 9892, 9894, 9896, 9898, 9900, 9902, 9904, 9906, 9908, 9910, 9912, 9914, 9916, 9918, 9920, 9922, 9924, 9926, 9928, 9930, 9932, 9934, 9936, 9938, 9940, 9942, 9944, 9946, 9948, 9950, 9952, 9954, 9956, 9958, 9960, 9962, 9964, 9966, 9968, 9970, 9972, 9974, 9976, 9978, 9980, 9982, 9984, 9986, 9988, 9990, 9992, 9994, 9996, 9998, 10000, 10002, 10004, 10006, 10008, 10010, 10012, 10014, 10016, 10018, 10020, 10022, 10024, 10026, 10028, 10030, 10032, 10034, 10036, 10038, 10040, 10042, 10044, 10046, 10048, 10050, 10052, 10054, 10056, 10058, 10060, 10062, 10064, 10066, 10068, 10070, 10072, 10074, 10076, 10078, 10080, 10082, 10084, 10086, 10088, 10090, 10092, 10094, 10096, 10098, 10100, 10102, 10104, 10106, 10108, 10110, 10112, 10114, 10116, 10118, 10120, 10122, 10124, 10126, 10128, 10130, 10132, 10134, 10136, 10138, 10140, 10142, 10144, 10146, 10148, 10150, 10152, 10154, 10156, 10158, 10160, 10162, 10164, 10166, 10168, 10170, 10172, 10174, 10176, 10178, 10180, 10182, 10184, 10186, 10188, 10190, 10192, 10194, 10196, 10198, 10200, 10202, 10204, 10206, 10208, 10210, 10212, 10214, 10216, 10218, 10220, 10222, 10224, 10226, 10228, 10230, 10232, 10234, 10236, 10238, 10240, 10242, 10244, 10246, 10248, 10250, 10252, 10254, 10256, 10258, 10260, 10262, 10264, 10266, 10268, 10270, 10272, 10274, 10276, 10278, 10280, 10282, 10284, 10286, 10288, 10290, 10292, 10294, 10296, 10298, 10300, 10302, 10304, 10306, 10308, 10310, 10312, 10314, 10316, 10318, 10320, 10322, 10324, 10326, 10328, 10330, 10332, 10334, 10336, 10338, 10340, 10342, 10344, 10346, 10348, 10350, 10352, 10354, 10356, 10358, 10360, 10362, 10364, 10366, 10368, 10370, 10372, 10374, 10376, 10378, 10380, 10382, 10384, 10386, 10388, 10390, 10392, 10394, 10396, 10398, 10400, 10402, 10404, 10406, 10408, 10410, 10412, 10414, 10416, 10418, 10420, 10422, 10424, 10426, 10428, 10430, 10432, 10434, 10436, 10438, 10440, 10442, 10444, 10446, 10448, 10450, 10452, 10454, 10456, 10458, 10460, 10462, 10464, 10466, 10468, 10470, 10472, 10474, 10476, 10478, 10480, 10482, 10484, 10486, 10488, 10490, 10492, 10494, 10496, 10498, 10500, 10502, 10504, 10506, 10508, 10510, 10512, 10514, 10516, 10518, 10520, 10522, 10524, 10526, 10528, 10530, 10532, 10534, 10536, 10538, 10540, 10542, 10544, 10546, 10548, 10550, 10552, 10554, 10556, 10558, 10560, 10562, 10564, 10566, 10568, 10570, 10572, 10574, 10576, 10578, 10580, 10582, 10584, 10586, 10588, 10590, 10592, 10594, 10596, 10598, 10600, 10602, 10604, 10606, 10608, 10610, 10612, 10614, 10616, 10618, 10620, 10622, 10624, 10626, 10628, 10630, 10632, 10634, 10636, 10638, 10640, 10642, 10644, 10646, 10648, 10650, 10652, 10654, 10656, 10658, 10660, 10662, 10664, 10666, 10668, 10670, 10672, 10674, 10676, 10678, 10680, 10682, 10684, 10686, 10688, 10690, 10692, 10694, 10696, 10698, 10700, 10702, 10704, 10706, 10708, 10710, 10712, 10714, 10716, 10718, 10720, 10722, 10724, 10726, 10728, 10730, 10732, 10734, 10736, 10738, 10740, 10742, 10744, 10746, 10748, 10750, 10752, 10754, 10756, 10758, 10760, 10762, 10764, 10766, 10768, 10770, 10772, 10774, 10776, 10778, 10780, 10782, 10784, 10786, 10788, 10790,

10792, 10794, 10796, 10798, 10800, 10802, 10804, 10806, 10808, 10810, 10812, 10814, 10816, 10818, 10820, 10822, 10824, 10826, 10828, 10830, 10832, 10834, 10836, 10838, 10840, 10842, 10844, 10846, 10848, 10850, 10852, 10854, 10856, 10858, 10860, 10862, 10864, 10866, 10868, 10870, 10872, 10874, 10876, 10878, 10880, 10882, 10884, 10886, 10888, 10890, 10892, 10894, 10896, 10898, 10900, 10902, 10904, 10906, 10908, 10910, 10912, 10914, 10916, 10918, 10920, 10922, 10924, 10926, 10928, 10930, 10932, 10934, 10936, 10938, 10940, 10942, 10944, 10946, 10948, 10950, 10952, 10954, 10956, 10958, 10960, 10962, 10964 and 10966, wherein said protein confers protective immunity.

4. An antibody which binds to a protein according to any one of claims 1 to 3.

5. The antibody of claim 4, wherein said antibody is a monoclonal antibody, a chimeric antibody, a humanised antibody, or a fully human antibody.

6. A nucleic acid molecule which encodes a protein according to any one of claims 1 to 3.

7. A nucleic acid molecule according to claim 6, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, 891, 893, 895, 897, 899, 901, 903, 905, 907, 909, 911, 913, 915, 917, 919, 921, 923, 925, 927, 929, 931, 933, 935, 937, 939, 941, 943, 945, 947, 949, 951, 953, 955, 957, 959, 961, 963, 965, 967, 969, 971, 973, 975, 977, 979, 981, 983, 985, 987, 989, 991, 993, 995, 997, 999, 1001, 1003, 1005, 1007, 1009, 1011, 1013, 1015, 1017, 1019, 1021, 1023, 1025, 1027, 1029, 1031, 1033, 1035, 1037, 1039, 1041, 1043, 1045, 1047, 1049, 1051, 1053, 1055, 1057, 1059, 1061, 1063, 1065, 1067, 1069, 1071, 1073, 1075, 1077, 1079, 1081, 1083, 1085, 1087, 1089, 1091, 1093, 1095, 1097, 1099, 1101, 1103, 1105, 1107, 1109, 1111, 1113, 1115, 1117, 1119, 1121, 1123, 1125, 1127, 1129, 1131, 1133, 1135, 1137, 1139, 1141, 1143, 1145, 1147, 1149, 1151, 1153, 1155, 1157, 1159, 1161, 1163, 1165, 1167, 1169, 1171, 1173, 1175, 1177, 1179, 1181, 1183, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1221, 1223, 1225, 1227, 1229, 1231, 1233, 1235, 1237, 1239, 1241, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, 1273, 1275, 1277, 1279, 1281, 1283, 1285, 1287, 1289, 1291, 1293, 1295, 1297, 1299, 1301, 1303, 1305, 1307, 1309, 1311, 1313, 1315, 1317, 1319, 1321, 1323, 1325, 1327, 1329,

1331, 1333, 1335, 1337, 1339, 1341, 1343, 1345, 1347, 1349, 1351, 1353, 1355, 1357, 1359, 1361, 1363, 1365, 1367, 1369, 1371, 1373, 1375, 1377, 1379, 1381, 1383, 1385, 1387, 1389, 1391, 1393, 1395, 1397, 1399, 1401, 1403, 1405, 1407, 1409, 1411, 1413, 1415, 1417, 1419, 1421, 1423, 1425, 1427, 1429, 1431, 1433, 1435, 1437, 1439, 1441, 1443, 1445, 1447, 1449, 1451, 1453, 1455, 1457, 1459, 1461, 1463, 1465, 1467, 1469, 1471, 1473, 1475, 1477, 1479, 1481, 1483, 1485, 1487, 1489, 1491, 1493, 1495, 1497, 1499, 1501, 1503, 1505, 1507, 1509, 1511, 1513, 1515, 1517, 1519, 1521, 1523, 1525, 1527, 1529, 1531, 1533, 1535, 1537, 1539, 1541, 1543, 1545, 1547, 1549, 1551, 1553, 1555, 1557, 1559, 1561, 1563, 1565, 1567, 1569, 1571, 1573, 1575, 1577, 1579, 1581, 1583, 1585, 1587, 1589, 1591, 1593, 1595, 1597, 1599, 1601, 1603, 1605, 1607, 1609, 1611, 1613, 1615, 1617, 1619, 1621, 1623, 1625, 1627, 1629, 1631, 1633, 1635, 1637, 1639, 1641, 1643, 1645, 1647, 1649, 1651, 1653, 1655, 1657, 1659, 1661, 1663, 1665, 1667, 1669, 1671, 1673, 1675, 1677, 1679, 1681, 1683, 1685, 1687, 1689, 1691, 1693, 1695, 1697, 1699, 1701, 1703, 1705, 1707, 1709, 1711, 1713, 1715, 1717, 1719, 1721, 1723, 1725, 1727, 1729, 1731, 1733, 1735, 1737, 1739, 1741, 1743, 1745, 1747, 1749, 1751, 1753, 1755, 1757, 1759, 1761, 1763, 1765, 1767, 1769, 1771, 1773, 1775, 1777, 1779, 1781, 1783, 1785, 1787, 1789, 1791, 1793, 1795, 1797, 1799, 1801, 1803, 1805, 1807, 1809, 1811, 1813, 1815, 1817, 1819, 1821, 1823, 1825, 1827, 1829, 1831, 1833, 1835, 1837, 1839, 1841, 1843, 1845, 1847, 1849, 1851, 1853, 1855, 1857, 1859, 1861, 1863, 1865, 1867, 1869, 1871, 1873, 1875, 1877, 1879, 1881, 1883, 1885, 1887, 1889, 1891, 1893, 1895, 1897, 1899, 1901, 1903, 1905, 1907, 1909, 1911, 1913, 1915, 1917, 1919, 1921, 1923, 1925, 1927, 1929, 1931, 1933, 1935, 1937, 1939, 1941, 1943, 1945, 1947, 1949, 1951, 1953, 1955, 1957, 1959, 1961, 1963, 1965, 1967, 1969, 1971, 1973, 1975, 1977, 1979, 1981, 1983, 1985, 1987, 1989, 1991, 1993, 1995, 1997, 1999, 2001, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017, 2019, 2021, 2023, 2025, 2027, 2029, 2031, 2033, 2035, 2037, 2039, 2041, 2043, 2045, 2047, 2049, 2051, 2053, 2055, 2057, 2059, 2061, 2063, 2065, 2067, 2069, 2071, 2073, 2075, 2077, 2079, 2081, 2083, 2085, 2087, 2089, 2091, 2093, 2095, 2097, 2099, 2101, 2103, 2105, 2107, 2109, 2111, 2113, 2115, 2117, 2119, 2121, 2123, 2125, 2127, 2129, 2131, 2133, 2135, 2137, 2139, 2141, 2143, 2145, 2147, 2149, 2151, 2153, 2155, 2157, 2159, 2161, 2163, 2165, 2167, 2169, 2171, 2173, 2175, 2177, 2179, 2181, 2183, 2185, 2187, 2189, 2191, 2193, 2195, 2197, 2199, 2201, 2203, 2205, 2207, 2209, 2211, 2213, 2215, 2217, 2219, 2221, 2223, 2225, 2227, 2229, 2231, 2233, 2235, 2237, 2239, 2241, 2243, 2245, 2247, 2249, 2251, 2253, 2255, 2257, 2259, 2261, 2263, 2265, 2267, 2269, 2271, 2273, 2275, 2277, 2279, 2281, 2283, 2285, 2287, 2289, 2291, 2293, 2295, 2297, 2299, 2301, 2303, 2305, 2307, 2309, 2311, 2313, 2315, 2317, 2319, 2321, 2323, 2325, 2327, 2329, 2331, 2333, 2335, 2337, 2339, 2341, 2343, 2345, 2347, 2349, 2351, 2353, 2355, 2357, 2359, 2361, 2363, 2365, 2367, 2369, 2371, 2373, 2375, 2377, 2379, 2381, 2383, 2385, 2387, 2389, 2391, 2393, 2395, 2397, 2399, 2401, 2403, 2405, 2407, 2409, 2411, 2413, 2415, 2417, 2419, 2421, 2423, 2425, 2427, 2429, 2431, 2433, 2435, 2437, 2439, 2441, 2443, 2445, 2447, 2449, 2451, 2453, 2455, 2457, 2459, 2461, 2463, 2465, 2467, 2469, 2471, 2473, 2475, 2477, 2479, 2481, 2483, 2485, 2487, 2489, 2491, 2493, 2495, 2497, 2499, 2501, 2503, 2505, 2507, 2509, 2511, 2513, 2515, 2517, 2519, 2521, 2523, 2525, 2527, 2529, 2531, 2533, 2535, 2537, 2539, 2541, 2543, 2545, 2547, 2549, 2551, 2553, 2555, 2557, 2559, 2561, 2563, 2565, 2567, 2569, 2571, 2573, 2575, 2577, 2579, 2581, 2583, 2585, 2587, 2589, 2591, 2593, 2595, 2597, 2599, 2601, 2603, 2605, 2607, 2609, 2611, 2613, 2615, 2617, 2619, 2621, 2623, 2625, 2627, 2629, 2631, 2633, 2635, 2637, 2639, 2641, 2643, 2645, 2647, 2649, 2651, 2653, 2655, 2657, 2659, 2661, 2663, 2665, 2667, 2669, 2671, 2673, 2675, 2677, 2679, 2681, 2683, 2685, 2687, 2689, 2691, 2693, 2695, 2697, 2699, 2701, 2703, 2705, 2707, 2709, 2711, 2713, 2715, 2717, 2719, 2721, 2723, 2725, 2727, 2729, 2731, 2733, 2735, 2737, 2739, 2741, 2743, 2745, 2747, 2749, 2751, 2753, 2755, 2757, 2759, 2761, 2763, 2765, 2767, 2769, 2771, 2773, 2775, 2777, 2779, 2781, 2783, 2785, 2787, 2789, 2791, 2793, 2795, 2797, 2799, 2801, 2803, 2805, 2807, 2809, 2811, 2813, 2815, 2817, 2819, 2821, 2823, 2825, 2827, 2829, 2831, 2833, 2835, 2837, 2839, 2841, 2843, 2845, 2847, 2849, 2851, 2853, 2855, 2857, 2859, 2861, 2863, 2865, 2867, 2869, 2871, 2873, 2875, 2877, 2879, 2881, 2883, 2885, 2887,

2889, 2891, 2893, 2895, 2897, 2899, 2901, 2903, 2905, 2907, 2909, 2911, 2913, 2915, 2917, 2919, 2921, 2923, 2925, 2927, 2929, 2931, 2933, 2935, 2937, 2939, 2941, 2943, 2945, 2947, 2949, 2951, 2953, 2955, 2957, 2959, 2961, 2963, 2965, 2967, 2969, 2971, 2973, 2975, 2977, 2979, 2981, 2983, 2985, 2987, 2989, 2991, 2993, 2995, 2997, 2999, 3001, 3003, 3005, 3007, 3009, 3011, 3013, 3015, 3017, 3019, 3021, 3023, 3025, 3027, 3029, 3031, 3033, 3035, 3037, 3039, 3041, 3043, 3045, 3047, 3049, 3051, 3053, 3055, 3057, 3059, 3061, 3063, 3065, 3067, 3069, 3071, 3073, 3075, 3077, 3079, 3081, 3083, 3085, 3087, 3089, 3091, 3093, 3095, 3097, 3099, 3101, 3103, 3105, 3107, 3109, 3111, 3113, 3115, 3117, 3119, 3121, 3123, 3125, 3127, 3129, 3131, 3133, 3135, 3137, 3139, 3141, 3143, 3145, 3147, 3149, 3151, 3153, 3155, 3157, 3159, 3161, 3163, 3165, 3167, 3169, 3171, 3173, 3175, 3177, 3179, 3181, 3183, 3185, 3187, 3189, 3191, 3193, 3195, 3197, 3199, 3201, 3203, 3205, 3207, 3209, 3211, 3213, 3215, 3217, 3219, 3221, 3223, 3225, 3227, 3229, 3231, 3233, 3235, 3237, 3239, 3241, 3243, 3245, 3247, 3249, 3251, 3253, 3255, 3257, 3259, 3261, 3263, 3265, 3267, 3269, 3271, 3273, 3275, 3277, 3279, 3281, 3283, 3285, 3287, 3289, 3291, 3293, 3295, 3297, 3299, 3301, 3303, 3305, 3307, 3309, 3311, 3313, 3315, 3317, 3319, 3321, 3323, 3325, 3327, 3329, 3331, 3333, 3335, 3337, 3339, 3341, 3343, 3345, 3347, 3349, 3351, 3353, 3355, 3357, 3359, 3361, 3363, 3365, 3367, 3369, 3371, 3373, 3375, 3377, 3379, 3381, 3383, 3385, 3387, 3389, 3391, 3393, 3395, 3397, 3399, 3401, 3403, 3405, 3407, 3409, 3411, 3413, 3415, 3417, 3419, 3421, 3423, 3425, 3427, 3429, 3431, 3433, 3435, 3437, 3439, 3441, 3443, 3445, 3447, 3449, 3451, 3453, 3455, 3457, 3459, 3461, 3463, 3465, 3467, 3469, 3471, 3473, 3475, 3477, 3479, 3481, 3483, 3485, 3487, 3489, 3491, 3493, 3495, 3497, 3499, 3501, 3503, 3505, 3507, 3509, 3511, 3513, 3515, 3517, 3519, 3521, 3523, 3525, 3527, 3529, 3531, 3533, 3535, 3537, 3539, 3541, 3543, 3545, 3547, 3549, 3551, 3553, 3555, 3557, 3559, 3561, 3563, 3565, 3567, 3569, 3571, 3573, 3575, 3577, 3579, 3581, 3583, 3585, 3587, 3589, 3591, 3593, 3595, 3597, 3599, 3601, 3603, 3605, 3607, 3609, 3611, 3613, 3615, 3617, 3619, 3621, 3623, 3625, 3627, 3629, 3631, 3633, 3635, 3637, 3639, 3641, 3643, 3645, 3647, 3649, 3651, 3653, 3655, 3657, 3659, 3661, 3663, 3665, 3667, 3669, 3671, 3673, 3675, 3677, 3679, 3681, 3683, 3685, 3687, 3689, 3691, 3693, 3695, 3697, 3699, 3701, 3703, 3705, 3707, 3709, 3711, 3713, 3715, 3717, 3719, 3721, 3723, 3725, 3727, 3729, 3731, 3733, 3735, 3737, 3739, 3741, 3743, 3747, 3749, 3751, 3753, 3755, 3757, 3759, 3761, 3763, 3765, 3767, 3769, 3771, 3773, 3775, 3777, 3779, 3781, 3783, 3785, 3787, 3789, 3791, 3793, 3795, 3797, 3799, 3801, 3803, 3805, 3807, 3809, 3811, 3813, 3815, 3817, 3819, 3821, 3823, 3825, 3827, 3829, 3831, 3833, 3835, 3837, 3839, 3841, 3843, 3845, 3847, 3849, 3851, 3853, 3855, 3857, 3859, 3861, 3863, 3865, 3867, 3869, 3871, 3873, 3875, 3877, 3879, 3881, 3883, 3885, 3887, 3889, 3891, 3893, 3895, 3897, 3899, 3901, 3903, 3905, 3907, 3909, 3911, 3913, 3915, 3917, 3919, 3923, 3925, 3927, 3929, 3931, 3933, 3935, 3937, 3939, 3941, 3943, 3945, 3947, 3949, 3951, 3953, 3955, 3957, 3959, 3961, 3963, 3965, 3967, 3969, 3971, 3973, 3975, 3977, 3979, 3981, 3983, 3985, 3987, 3989, 3991, 3993, 3995, 3997, 3999, 4001, 4003, 4005, 4007, 4009, 4011, 4013, 4015, 4017, 4019, 4021, 4023, 4025, 4027, 4029, 4031, 4033, 4035, 4037, 4039, 4041, 4043, 4045, 4047, 4049, 4051, 4053, 4055, 4057, 4059, 4061, 4063, 4065, 4067, 4069, 4071, 4073, 4075, 4077, 4079, 4081, 4083, 4085, 4087, 4089, 4091, 4093, 4095, 4097, 4099, 4101, 4103, 4105, 4107, 4109, 4111, 4113, 4115, 4117, 4119, 4121, 4123, 4125, 4127, 4129, 4131, 4133, 4135, 4137, 4139, 4141, 4143, 4145, 4147, 4149, 4151, 4153, 4155, 4157, 4159, 4161, 4163, 4165, 4167, 4169, 4171, 4173, 4175, 4177, 4179, 4181, 4183, 4185, 4187, 4189, 4191, 4193, 4195, 4197, 4199, 4201, 4203, 4205, 4207, 4209, 4211, 4213, 4215, 4217, 4219, 4221, 4223, 4225, 4227, 4229, 4231, 4233, 4235, 4237, 4239, 4241, 4243, 4245, 4247, 4249, 4251, 4253, 4255, 4257, 4259, 4261, 4263, 4265, 4267, 4269, 4271, 4273, 4275, 4277, 4279, 4281, 4283, 4285, 4287, 4289, 4291, 4293, 4295, 4297, 4299, 4301, 4303, 4305, 4307, 4309, 4311, 4313, 4315, 4317, 4319, 4321, 4323, 4325, 4327, 4329, 4331, 4333, 4335, 4337, 4339, 4341, 4343, 4345, 4347, 4349, 4351, 4353, 4355, 4357, 4359, 4361, 4363, 4365, 4367, 4369, 4371, 4373, 4375, 4377, 4379, 4381, 4383, 4385, 4387, 4389, 4391, 4393, 4395, 4397, 4399, 4401, 4403, 4405, 4407, 4409, 4411, 4413, 4415, 4417, 4419, 4421, 4423, 4425, 4427, 4429, 4431, 4433, 4435, 4437, 4439, 4441, 4443, 4445,

4447, 4449, 4451, 4453, 4455, 4457, 4459, 4461, 4463, 4465, 4467, 4469, 4471, 4473, 4475, 4477, 4479, 4481, 4483, 4485, 4487, 4489, 4491, 4493, 4495, 4497, 4499, 4501, 4503, 4505, 4507, 4509, 4511, 4513, 4515, 4517, 4519, 4521, 4523, 4525, 4527, 4529, 4531, 4533, 4535, 4537, 4539, 4541, 4543, 4545, 4547, 4549, 4551, 4553, 4555, 4557, 4559, 4561, 4563, 4565, 4567, 4569, 4571, 4573, 4575, 4577, 4579, 4581, 4583, 4585, 4587, 4589, 4591, 4593, 4595, 4597, 4599, 4601, 4603, 4605, 4607, 4609, 4611, 4613, 4615, 4617, 4619, 4621, 4623, 4625, 4627, 4629, 4631, 4633, 4635, 4637, 4639, 4641, 4643, 4645, 4647, 4649, 4651, 4653, 4655, 4657, 4659, 4661, 4663, 4665, 4667, 4669, 4671, 4673, 4675, 4677, 4679, 4681, 4683, 4685, 4687, 4689, 4691, 4693, 4695, 4697, 4699, 4701, 4703, 4705, 4707, 4709, 4711, 4713, 4715, 4717, 4719, 4721, 4723, 4725, 4727, 4729, 4731, 4733, 4735, 4737, 4739, 4741, 4743, 4745, 4747, 4749, 4751, 4753, 4755, 4757, 4759, 4761, 4763, 4765, 4767, 4769, 4771, 4773, 4775, 4777, 4779, 4781, 4783, 4785, 4787, 4789, 4791, 4793, 4795, 4797, 4799, 4801, 4803, 4805, 4807, 4809, 4811, 4813, 4815, 4817, 4819, 4821, 4823, 4825, 4827, 4829, 4831, 4833, 4835, 4837, 4839, 4841, 4843, 4845, 4847, 4849, 4851, 4853, 4855, 4857, 4859, 4861, 4863, 4865, 4867, 4869, 4871, 4873, 4875, 4877, 4879, 4881, 4883, 4885, 4887, 4889, 4891, 4893, 4895, 4897, 4899, 4901, 4903, 4905, 4907, 4909, 4911, 4913, 4915, 4917, 4919, 4921, 4923, 4925, 4927, 4929, 4931, 4933, 4935, 4937, 4939, 4941, 4943, 4945, 4947, 4949, 4951, 4953, 4955, 4957, 4959, 4961, 4963, 4965, 4967, 4969, 4971, 4973, 4975, 4977, 4979, 4981, 4983, 4985, 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5001, 5003, 5005, 5007, 5009, 5011, 5013, 5015, 5017, 5019, 5021, 5023, 5025, 5027, 5029, 5031, 5033, 5035, 5037, 5039, 5041, 5043, 5045, 5047, 5049, 5051, 5053, 5055, 5057, 5059, 5061, 5063, 5065, 5067, 5069, 5071, 5073, 5075, 5077, 5079, 5081, 5083, 5085, 5087, 5089, 5091, 5093, 5095, 5097, 5099, 5101, 5103, 5105, 5107, 5109, 5111, 5113, 5115, 5117, 5119, 5121, 5123, 5125, 5127, 5129, 5131, 5133, 5135, 5137, 5139, 5141, 5143, 5145, 5147, 5149, 5151, 5153, 5155, 5157, 5159, 5161, 5163, 5165, 5167, 5169, 5171, 5173, 5175, 5177, 5179, 5181, 5183, 5185, 5187, 5189, 5191, 5193, 5195, 5197, 5199, 5201, 5203, 5205, 5207, 5209, 5211, 5213, 5215, 5217, 5219, 5221, 5223, 5225, 5227, 5229, 5231, 5233, 5235, 5237, 5239, 5241, 5243, 5245, 5247, 5249, 5251, 5253, 5255, 5257, 5259, 5261, 5263, 5265, 5267, 5269, 5271, 5273, 5275, 5277, 5279, 5281, 5283, 5285, 5287, 5289, 5291, 5293, 5295, 5297, 5299, 5301, 5303, 5305, 5307, 5309, 5311, 5313, 5315, 5317, 5319, 5321, 5323, 5325, 5327, 5329, 5331, 5333, 5335, 5337, 5339, 5341, 5343, 5345, 5347, 5349, 5351, 5353, 5355, 5357, 5359, 5361, 5363, 5365, 5367, 5369, 5371, 5373, 5375, 5377, 5379, 5381, 5383, 5385, 5387, 5389, 5391, 5393, 5395, 5397, 5399, 5401, 5403, 5405, 5407, 5409, 5411, 5413, 5415, 5417, 5419, 5421, 5423, 5425, 5427, 5429, 5431, 5433, 5435, 5437, 5439, 5441, 5443, 5445, 5447, 5449, 5451, 5453, 5455, 5457, 5459, 5461, 5463, 5465, 5467, 5469, 5471, 5473, 5475, 5477, 5479, 5481, 5483, 5485, 5487, 5489, 5491, 5493, 5495, 5497, 5499, 5501, 5503, 5505, 5507, 5509, 5511, 5513, 5515, 5517, 5519, 5521, 5523, 5525, 5527, 5529, 5531, 5533, 5535, 5537, 5539, 5541, 5543, 5545, 5547, 5549, 5551, 5553, 5555, 5557, 5559, 5561, 5563, 5565, 5567, 5569, 5571, 5573, 5575, 5577, 5579, 5581, 5583, 5585, 5587, 5589, 5591, 5593, 5595, 5597, 5599, 5601, 5603, 5605, 5607, 5609, 5611, 5613, 5615, 5617, 5619, 5621, 5623, 5625, 5627, 5629, 5631, 5633, 5635, 5637, 5639, 5641, 5643, 5645, 5647, 5649, 5651, 5653, 5655, 5657, 5659, 5661, 5663, 5665, 5667, 5669, 5671, 5673, 5675, 5677, 5679, 5681, 5683, 5685, 5687, 5689, 5691, 5693, 5695, 5697, 5699, 5701, 5703, 5705, 5707, 5709, 5711, 5713, 5715, 5717, 5719, 5721, 5723, 5725, 5727, 5729, 5731, 5733, 5735, 5737, 5739, 5741, 5743, 5745, 5747, 5749, 5751, 5753, 5755, 5757, 5759, 5761, 5763, 5765, 5767, 5769, 5771, 5773, 5775, 5777, 5779, 5781, 5783, 5785, 5787, 5789, 5791, 5793, 5795, 5797, 5799, 5801, 5803, 5805, 5807, 5809, 5811, 5813, 5815, 5817, 5819, 5821, 5823, 5825, 5827, 5829, 5831, 5833, 5835, 5837, 5839, 5841, 5843, 5845, 5847, 5849, 5851, 5853, 5855, 5857, 5859, 5861, 5863, 5865, 5867, 5869, 5871, 5873, 5875, 5877, 5879, 5881, 5883, 5885, 5887, 5889, 5891, 5893, 5895, 5897, 5899, 5901, 5903, 5905, 5907, 5909, 5911, 5913, 5915, 5917, 5919, 5921, 5923, 5925, 5927, 5929, 5931, 5933, 5935, 5937, 5939, 5941, 5943, 5945, 5947, 5949, 5951, 5953, 5955, 5957, 5959, 5961, 5963, 5965, 5967, 5969, 5971, 5973, 5975, 5977, 5979, 5981, 5983, 5985, 5987, 5989, 5991, 5993, 5995, 5997, 5999, 6001, 6003,

6005, 6007, 6009, 6011, 6013, 6015, 6017, 6019, 6021, 6023, 6025, 6027, 6029, 6031, 6033, 6035, 6037, 6039, 6041, 6043, 6045, 6047, 6049, 6051, 6053, 6055, 6057, 6059, 6061, 6063, 6065, 6067, 6069, 6071, 6073, 6075, 6077, 6079, 6081, 6083, 6085, 6087, 6089, 6091, 6093, 6095, 6097, 6099, 6101, 6103, 6105, 6107, 6109, 6111, 6113, 6115, 6117, 6119, 6121, 6123, 6125, 6127, 6129, 6131, 6133, 6135, 6137, 6139, 6141, 6143, 6145, 6147, 6149, 6151, 6153, 6155, 6157, 6159, 6161, 6163, 6165, 6167, 6169, 6171, 6173, 6175, 6177, 6179, 6181, 6183, 6185, 6187, 6189, 6191, 6193, 6195, 6197, 6199, 6201, 6203, 6205, 6207, 6209, 6211, 6213, 6215, 6217, 6219, 6221, 6223, 6225, 6227, 6229, 6231, 6233, 6235, 6237, 6239, 6241, 6243, 6245, 6247, 6249, 6251, 6253, 6255, 6257, 6259, 6261, 6263, 6265, 6267, 6269, 6271, 6273, 6275, 6277, 6279, 6281, 6283, 6285, 6287, 6289, 6291, 6293, 6295, 6297, 6299, 6301, 6303, 6305, 6307, 6309, 6311, 6313, 6315, 6317, 6319, 6321, 6323, 6325, 6327, 6329, 6331, 6333, 6335, 6337, 6339, 6341, 6343, 6345, 6347, 6349, 6351, 6353, 6355, 6357, 6359, 6361, 6363, 6365, 6367, 6369, 6371, 6373, 6375, 6377, 6379, 6381, 6383, 6385, 6387, 6389, 6391, 6393, 6395, 6397, 6399, 6401, 6403, 6405, 6407, 6409, 6411, 6413, 6415, 6417, 6419, 6421, 6423, 6425, 6427, 6429, 6431, 6433, 6435, 6437, 6439, 6441, 6443, 6445, 6447, 6449, 6451, 6453, 6455, 6457, 6459, 6461, 6463, 6465, 6467, 6469, 6471, 6473, 6475, 6477, 6479, 6481, 6483, 6485, 6487, 6489, 6491, 6493, 6495, 6497, 6499, 6501, 6503, 6505, 6507, 6509, 6511, 6513, 6515, 6517, 6519, 6521, 6523, 6525, 6527, 6529, 6531, 6533, 6535, 6537, 6539, 6541, 6543, 6545, 6547, 6549, 6551, 6553, 6555, 6557, 6559, 6561, 6563, 6565, 6567, 6569, 6571, 6573, 6575, 6577, 6579, 6581, 6583, 6585, 6587, 6589, 6591, 6593, 6595, 6597, 6599, 6601, 6603, 6605, 6607, 6609, 6611, 6613, 6615, 6617, 6619, 6621, 6623, 6625, 6627, 6629, 6631, 6633, 6635, 6637, 6639, 6641, 6643, 6645, 6647, 6649, 6651, 6653, 6655, 6657, 6659, 6661, 6663, 6665, 6667, 6669, 6671, 6673, 6675, 6677, 6679, 6681, 6683, 6685, 6687, 6689, 6691, 6693, 6695, 6697, 6699, 6701, 6703, 6705, 6707, 6709, 6711, 6713, 6715, 6717, 6719, 6721, 6723, 6725, 6727, 6729, 6731, 6733, 6735, 6737, 6739, 6741, 6743, 6745, 6747, 6749, 6751, 6753, 6755, 6757, 6759, 6761, 6763, 6765, 6767, 6769, 6771, 6773, 6775, 6777, 6779, 6781, 6783, 6785, 6787, 6789, 6791, 6793, 6795, 6797, 6799, 6801, 6803, 6805, 6807, 6809, 6811, 6813, 6815, 6817, 6819, 6821, 6823, 6825, 6827, 6829, 6831, 6833, 6835, 6837, 6839, 6841, 6843, 6845, 6847, 6849, 6851, 6853, 6855, 6857, 6859, 6861, 6863, 6865, 6867, 6869, 6871, 6873, 6875, 6877, 6879, 6881, 6883, 6885, 6887, 6889, 6891, 6893, 6895, 6897, 6899, 6901, 6903, 6905, 6907, 6909, 6911, 6913, 6915, 6917, 6919, 6921, 6923, 6925, 6927, 6929, 6931, 6933, 6935, 6937, 6939, 6941, 6943, 6945, 6947, 6949, 6951, 6953, 6955, 6957, 6959, 6961, 6963, 6965, 6967, 6969, 6971, 6973, 6975, 6977, 6979, 6981, 6983, 6985, 6987, 6989, 6991, 6993, 6995, 6997, 6999, 7001, 7003, 7005, 7007, 7009, 7011, 7013, 7015, 7017, 7019, 7021, 7023, 7025, 7027, 7029, 7031, 7033, 7035, 7037, 7039, 7041, 7043, 7045, 7047, 7049, 7051, 7053, 7055, 7057, 7059, 7061, 7063, 7065, 7067, 7069, 7071, 7073, 7075, 7077, 7079, 7081, 7083, 7085, 7087, 7089, 7091, 7093, 7095, 7097, 7099, 7101, 7103, 7105, 7107, 7109, 7111, 7113, 7115, 7117, 7119, 7121, 7123, 7125, 7127, 7129, 7131, 7133, 7135, 7137, 7139, 7141, 7143, 7145, 7147, 7149, 7151, 7153, 7155, 7157, 7159, 7161, 7163, 7165, 7167, 7169, 7171, 7173, 7175, 7177, 7179, 7181, 7183, 7185, 7187, 7189, 7191, 7193, 7195, 7197, 7199, 7201, 7203, 7205, 7207, 7209, 7211, 7213, 7215, 7217, 7219, 7221, 7223, 7225, 7227, 7229, 7231, 7233, 7235, 7237, 7239, 7241, 7243, 7245, 7247, 7249, 7251, 7253, 7255, 7257, 7259, 7261, 7263, 7265, 7267, 7269, 7271, 7273, 7275, 7277, 7279, 7281, 7283, 7285, 7287, 7289, 7291, 7293, 7295, 7297, 7299, 7301, 7303, 7305, 7307, 7309, 7311, 7313, 7315, 7317, 7319, 7321, 7323, 7325, 7327, 7329, 7331, 7333, 7335, 7337, 7339, 7341, 7343, 7345, 7347, 7349, 7351, 7353, 7355, 7357, 7359, 7361, 7363, 7365, 7367, 7369, 7371, 7373, 7375, 7377, 7379, 7381, 7383, 7385, 7387, 7389, 7391, 7393, 7395, 7397, 7399, 7401, 7403, 7405, 7407, 7409, 7411, 7413, 7415, 7417, 7419, 7421, 7423, 7425, 7427, 7429, 7431, 7433, 7435, 7437, 7439, 7441, 7443, 7445, 7447, 7449, 7451, 7453, 7455, 7457, 7459, 7461, 7463, 7467, 7469, 7471, 7473, 7475, 7477, 7479, 7481, 7483, 7485, 7487, 7489, 7491, 7493, 7495, 7497, 7499, 7501, 7503, 7505, 7507, 7509, 7511, 7513, 7515, 7517, 7519, 7521, 7523, 7525, 7527, 7529, 7531, 7533, 7535, 7537, 7539, 7541, 7543, 7545, 7547, 7549, 7551, 7553, 7555, 7557, 7559, 7561,

7563, 7565, 7567, 7569, 7571, 7573, 7575, 7577, 7579, 7581, 7583, 7585, 7587, 7589, 7591, 7593, 7595, 7597, 7599, 7601, 7603, 7605, 7607, 7609, 7611, 7613, 7615, 7617, 7619, 7621, 7623, 7625, 7627, 7629, 7631, 7633, 7635, 7637, 7639, 7641, 7643, 7645, 7647, 7649, 7651, 7653, 7655, 7657, 7659, 7661, 7663, 7665, 7667, 7669, 7671, 7673, 7675, 7677, 7679, 7681, 7683, 7685, 7687, 7689, 7691, 7693, 7695, 7697, 7699, 7701, 7703, 7705, 7707, 7709, 7711, 7713, 7715, 7717, 7719, 7721, 7723, 7725, 7727, 7729, 7731, 7733, 7735, 7737, 7739, 7741, 7743, 7745, 7747, 7749, 7751, 7753, 7755, 7757, 7759, 7761, 7763, 7765, 7767, 7769, 7771, 7773, 7775, 7777, 7779, 7781, 7783, 7785, 7787, 7789, 7791, 7793, 7795, 7797, 7799, 7801, 7803, 7805, 7807, 7809, 7811, 7813, 7815, 7817, 7819, 7821, 7823, 7825, 7827, 7829, 7831, 7833, 7835, 7837, 7839, 7841, 7843, 7845, 7847, 7849, 7851, 7853, 7855, 7857, 7859, 7861, 7863, 7865, 7867, 7869, 7871, 7873, 7875, 7877, 7879, 7881, 7883, 7885, 7887, 7889, 7891, 7893, 7895, 7897, 7899, 7901, 7903, 7905, 7907, 7909, 7911, 7913, 7915, 7917, 7919, 7921, 7923, 7925, 7927, 7929, 7931, 7933, 7935, 7937, 7939, 7941, 7943, 7945, 7947, 7949, 7951, 7953, 7955, 7957, 7959, 7961, 7963, 7965, 7967, 7969, 7971, 7973, 7975, 7977, 7979, 7981, 7983, 7985, 7987, 7989, 7991, 7993, 7995, 7997, 7999, 8001, 8003, 8005, 8007, 8009, 8011, 8013, 8015, 8017, 8019, 8021, 8023, 8025, 8027, 8029, 8031, 8033, 8035, 8037, 8039, 8041, 8043, 8045, 8047, 8049, 8051, 8053, 8055, 8057, 8059, 8061, 8063, 8065, 8067, 8069, 8071, 8073, 8075, 8077, 8079, 8081, 8083, 8085, 8087, 8089, 8091, 8093, 8095, 8097, 8099, 8101, 8103, 8105, 8107, 8109, 8111, 8113, 8115, 8117, 8119, 8121, 8123, 8125, 8127, 8129, 8131, 8133, 8135, 8137, 8139, 8141, 8143, 8145, 8147, 8149, 8151, 8153, 8155, 8157, 8159, 8161, 8163, 8165, 8167, 8169, 8171, 8173, 8175, 8177, 8179, 8181, 8183, 8185, 8187, 8189, 8191, 8193, 8195, 8197, 8199, 8201, 8203, 8205, 8207, 8209, 8211, 8213, 8215, 8217, 8219, 8221, 8223, 8225, 8227, 8229, 8231, 8233, 8235, 8237, 8239, 8241, 8243, 8245, 8247, 8249, 8251, 8253, 8255, 8257, 8259, 8261, 8263, 8265, 8267, 8269, 8271, 8273, 8275, 8277, 8279, 8281, 8283, 8285, 8287, 8289, 8291, 8293, 8295, 8297, 8299, 8301, 8303, 8305, 8307, 8309, 8311, 8313, 8315, 8317, 8319, 8321, 8323, 8325, 8327, 8329, 8331, 8333, 8335, 8337, 8339, 8341, 8343, 8345, 8347, 8349, 8351, 8353, 8355, 8357, 8359, 8361, 8363, 8365, 8367, 8369, 8371, 8373, 8375, 8377, 8379, 8381, 8383, 8385, 8387, 8389, 8391, 8393, 8395, 8397, 8399, 8401, 8403, 8405, 8407, 8409, 8411, 8413, 8415, 8417, 8419, 8421, 8423, 8425, 8427, 8429, 8431, 8433, 8435, 8437, 8439, 8441, 8443, 8445, 8447, 8449, 8451, 8453, 8455, 8457, 8459, 8461, 8463, 8465, 8467, 8469, 8471, 8473, 8475, 8477, 8479, 8481, 8483, 8485, 8487, 8489, 8491, 8493, 8495, 8497, 8499, 8501, 8503, 8505, 8507, 8509, 8511, 8513, 8515, 8517, 8519, 8521, 8523, 8525, 8527, 8529, 8531, 8533, 8535, 8537, 8539, 8541, 8543, 8545, 8547, 8549, 8551, 8553, 8555, 8557, 8559, 8561, 8563, 8565, 8567, 8569, 8571, 8573, 8575, 8577, 8579, 8581, 8583, 8585, 8587, 8589, 8591, 8593, 8595, 8597, 8599, 8601, 8603, 8605, 8607, 8609, 8611, 8613, 8615, 8617, 8619, 8621, 8623, 8625, 8627, 8629, 8631, 8633, 8635, 8637, 8639, 8641, 8643, 8645, 8647, 8649, 8651, 8653, 8655, 8657, 8659, 8661, 8663, 8665, 8667, 8669, 8671, 8673, 8675, 8677, 8679, 8681, 8683, 8685, 8687, 8689, 8691, 8693, 8695, 8697, 8699, 8701, 8703, 8705, 8707, 8709, 8711, 8713, 8715, 8717, 8719, 8721, 8723, 8725, 8727, 8729, 8731, 8733, 8735, 8737, 8739, 8741, 8743, 8745, 8747, 8749, 8751, 8753, 8755, 8757, 8759, 8761, 8763, 8765, 8767, 8769, 8771, 8773, 8775, 8777, 8781, 8783, 8785, 8787, 8789, 8791, 8793, 8795, 8797, 8799, 8801, 8803, 8805, 8807, 8809, 8811, 8813, 8815, 8817, 8819, 8821, 8823, 8825, 8827, 8829, 8831, 8833, 8835, 8837, 8839, 8841, 8843, 8845, 8847, 8849, 8851, 8853, 8855, 8857, 8859, 8861, 8863, 8865, 8867, 8869, 8871, 8873, 8875, 8877, 8879, 8881, 8883, 8885, 8887, 8889, 8891, 8893, 8895, 8897, 8899, 8901, 8903, 8905, 8907, 8909, 8911, 8913, 8915, 8917, 8919, 8921, 8923, 8925, 8927, 8929, 8931, 8933, 8935, 8937, 8939, 8941, 8943, 8945, 8947, 8949, 8951, 8953, 8955, 8957, 8959, 8961, 8963, 8965, 8967, 8969, 8971, 8973, 8975, 8977, 8979, 8981, 8983, 8985, 8987, 8989, 8991, 8993, 8995, 8997, 8999, 9001, 9003, 9005, 9007, 9009, 9011, 9013, 9015, 9017, 9019, 9021, 9023, 9025, 9027, 9029, 9031, 9033, 9035, 9037, 9039, 9041, 9043, 9045, 9047, 9049, 9051, 9053, 9055, 9057, 9059, 9061, 9063, 9065, 9067, 9069, 9071, 9073, 9075, 9077, 9079, 9081, 9083, 9085, 9087, 9089, 9091, 9093, 9095, 9097, 9099, 9101, 9103, 9105, 9107, 9109, 9111, 9113, 9115, 9117, 9119,

9121, 9123, 9125, 9127, 9129, 9131, 9133, 9135, 9137, 9139, 9141, 9143, 9145, 9147, 9149, 9151, 9153, 9155, 9157, 9159, 9161, 9163, 9165, 9167, 9169, 9171, 9173, 9175, 9177, 9179, 9181, 9183, 9185, 9189, 9191, 9193, 9195, 9197, 9199, 9201, 9203, 9205, 9207, 9209, 9211, 9213, 9215, 9217, 9219, 9221, 9223, 9225, 9227, 9229, 9231, 9233, 9235, 9237, 9239, 9241, 9243, 9245, 9247, 9249, 9251, 9253, 9255, 9257, 9259, 9261, 9263, 9265, 9267, 9269, 9271, 9273, 9275, 9277, 9279, 9281, 9283, 9285, 9287, 9289, 9291, 9293, 9295, 9297, 9299, 9301, 9303, 9305, 9307, 9309, 9311, 9313, 9315, 9317, 9319, 9321, 9323, 9325, 9327, 9329, 9331, 9333, 9335, 9337, 9339, 9341, 9343, 9345, 9347, 9349, 9351, 9353, 9355, 9357, 9359, 9361, 9363, 9365, 9367, 9369, 9371, 9373, 9375, 9377, 9379, 9381, 9383, 9385, 9387, 9389, 9391, 9393, 9395, 9397, 9399, 9401, 9403, 9405, 9407, 9409, 9411, 9413, 9415, 9417, 9419, 9421, 9423, 9425, 9427, 9429, 9431, 9433, 9435, 9437, 9439, 9441, 9443, 9445, 9447, 9449, 9451, 9453, 9455, 9457, 9459, 9461, 9463, 9465, 9467, 9469, 9471, 9473, 9475, 9477, 9479, 9481, 9483, 9485, 9487, 9489, 9491, 9493, 9495, 9497, 9499, 9501, 9503, 9505, 9507, 9509, 9511, 9513, 9515, 9517, 9519, 9521, 9523, 9525, 9527, 9529, 9531, 9533, 9535, 9537, 9539, 9541, 9543, 9545, 9547, 9549, 9551, 9553, 9555, 9557, 9559, 9561, 9563, 9565, 9567, 9569, 9571, 9573, 9575, 9577, 9579, 9581, 9583, 9585, 9587, 9589, 9591, 9593, 9595, 9597, 9599, 9601, 9603, 9605, 9607, 9609, 9611, 9613, 9615, 9617, 9619, 9621, 9623, 9625, 9627, 9629, 9631, 9633, 9635, 9637, 9639, 9641, 9643, 9645, 9647, 9649, 9651, 9653, 9655, 9657, 9659, 9661, 9663, 9665, 9667, 9669, 9671, 9673, 9675, 9677, 9679, 9681, 9683, 9685, 9687, 9689, 9691, 9693, 9695, 9697, 9699, 9701, 9703, 9705, 9707, 9709, 9711, 9713, 9715, 9717, 9719, 9721, 9723, 9725, 9727, 9729, 9731, 9733, 9735, 9737, 9739, 9741, 9743, 9745, 9747, 9749, 9751, 9753, 9755, 9757, 9759, 9761, 9763, 9765, 9767, 9769, 9771, 9773, 9775, 9777, 9779, 9781, 9783, 9785, 9787, 9789, 9791, 9793, 9795, 9797, 9799, 9801, 9803, 9805, 9807, 9809, 9811, 9813, 9815, 9817, 9819, 9821, 9823, 9825, 9827, 9829, 9831, 9833, 9835, 9837, 9839, 9841, 9843, 9845, 9847, 9849, 9851, 9853, 9855, 9857, 9859, 9861, 9863, 9865, 9867, 9869, 9871, 9873, 9875, 9877, 9879, 9881, 9883, 9885, 9887, 9889, 9891, 9893, 9895, 9897, 9899, 9901, 9903, 9905, 9907, 9909, 9911, 9913, 9915, 9917, 9919, 9921, 9923, 9925, 9927, 9929, 9931, 9933, 9935, 9937, 9939, 9941, 9943, 9945, 9947, 9949, 9951, 9953, 9955, 9957, 9959, 9961, 9963, 9965, 9967, 9969, 9971, 9973, 9975, 9977, 9979, 9981, 9983, 9985, 9987, 9989, 9991, 9993, 9995, 9997, 9999, 10001, 10003, 10005, 10007, 10009, 10011, 10013, 10015, 10017, 10019, 10021, 10023, 10025, 10027, 10029, 10031, 10033, 10035, 10037, 10039, 10041, 10043, 10045, 10047, 10049, 10051, 10053, 10055, 10057, 10059, 10061, 10063, 10065, 10067, 10069, 10071, 10073, 10075, 10077, 10079, 10081, 10083, 10085, 10087, 10089, 10091, 10093, 10095, 10097, 10099, 10101, 10103, 10105, 10107, 10109, 10111, 10113, 10115, 10117, 10119, 10121, 10123, 10125, 10127, 10129, 10131, 10133, 10135, 10137, 10139, 10141, 10143, 10145, 10147, 10149, 10151, 10153, 10155, 10157, 10159, 10161, 10163, 10165, 10167, 10169, 10171, 10173, 10175, 10177, 10179, 10181, 10183, 10185, 10187, 10189, 10191, 10193, 10195, 10197, 10199, 10201, 10203, 10205, 10207, 10209, 10211, 10213, 10215, 10217, 10219, 10221, 10223, 10225, 10227, 10229, 10231, 10233, 10235, 10237, 10239, 10241, 10243, 10245, 10247, 10249, 10251, 10253, 10255, 10257, 10259, 10261, 10263, 10265, 10267, 10269, 10271, 10273, 10275, 10277, 10279, 10281, 10283, 10285, 10287, 10289, 10291, 10293, 10295, 10297, 10299, 10301, 10303, 10305, 10307, 10309, 10311, 10313, 10315, 10317, 10319, 10321, 10323, 10325, 10327, 10329, 10331, 10333, 10335, 10337, 10339, 10341, 10343, 10345, 10347, 10349, 10351, 10353, 10355, 10357, 10359, 10361, 10363, 10365, 10367, 10369, 10371, 10373, 10375, 10377, 10379, 10381, 10383, 10385, 10387, 10389, 10391, 10393, 10395, 10397, 10399, 10401, 10403, 10405, 10407, 10409, 10411, 10413, 10415, 10417, 10419, 10421, 10423, 10425, 10427, 10429, 10431, 10433, 10435, 10437, 10439, 10441, 10443, 10445, 10447, 10449, 10451, 10453, 10455, 10457, 10459, 10461, 10463, 10465, 10467, 10469, 10471, 10473, 10475, 10477, 10479, 10481, 10483, 10485, 10487, 10489, 10491, 10493, 10495, 10497, 10499, 10501, 10503, 10505, 10507, 10509, 10511, 10513, 10515, 10517, 10519, 10521, 10523, 10525, 10527, 10529, 10531, 10533, 10535, 10537, 10539, 10541, 10543, 10545, 10547, 10549, 10551, 10553, 10555, 10557, 10559, 10561, 10563, 10565, 10567, 10569,

10571, 10573, 10575, 10577, 10579, 10581, 10583, 10585, 10587, 10589, 10591, 10593, 10595, 10597, 10599, 10601, 10603, 10605, 10607, 10609, 10611, 10613, 10615, 10617, 10619, 10621, 10623, 10625, 10627, 10629, 10631, 10633, 10635, 10637, 10639, 10641, 10643, 10645, 10647, 10649, 10651, 10653, 10655, 10657, 10659, 10661, 10663, 10665, 10667, 10669, 10671, 10673, 10675, 10677, 10679, 10681, 10683, 10685, 10687, 10689, 10691, 10693, 10695, 10697, 10699, 10701, 10703, 10705, 10707, 10709, 10711, 10713, 10715, 10717, 10719, 10721, 10723, 10725, 10727, 10729, 10731, 10733, 10735, 10737, 10739, 10741, 10743, 10745, 10747, 10749, 10751, 10753, 10755, 10757, 10759, 10761, 10763, 10765, 10767, 10769, 10771, 10773, 10775, 10777, 10779, 10781, 10783, 10785, 10787, 10789, 10791, 10793, 10795, 10797, 10799, 10801, 10803, 10805, 10807, 10809, 10811, 10813, 10815, 10817, 10819, 10821, 10823, 10825, 10827, 10829, 10831, 10833, 10835, 10837, 10839, 10841, 10843, 10845, 10847, 10849, 10851, 10853, 10855, 10857, 10859, 10861, 10863, 10865, 10867, 10869, 10871, 10873, 10875, 10877, 10879, 10881, 10883, 10885, 10887, 10889, 10891, 10893, 10895, 10897, 10899, 10901, 10903, 10905, 10907, 10909, 10911, 10913, 10915, 10917, 10919, 10921, 10923, 10925, 10927, 10929, 10931, 10933, 10935, 10937, 10939, 10941, 10943, 10945, 10947, 10949, 10951, 10953, 10955, 10957, 10959, 10961, 10963 and 10965.

8. A nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ IDs 10967, 10968, 10969, 10970, 10971, 10972, 10973, 10974, 10975, 10976, 10977, 10978, 10979, 10980, 10981, 10982, 10983, 10984, 10985, 10986, 10987, 10988, 10989, 10990, 10991, 10992, 10993, 10994, 10995, 10996, 10997, 10998, 10999, 11000, 11001, 11002, 11003, 11004, 11005, 11006, 11007, 11008, 11009, 11010, 11011, 11012, 11013, 11014, 11015, 11016, 11017, 11018, 11019, 11020, 11021, 11022, 11023, 11024, 11025, 11026, 11027, 11028, 11029, 11030, 11031, 11032, 11033, 11034, 11035, 11036, 11037, 11038, 11039, 11040, 11041, 11042, 11043, 11044, 11045, 11046, 11047, 11048, 11049, 11050, 11051, 11052, 11053, 11054, 11055, 11056, 11057, 11058, 11059, 11060, 11061, 11062, 11063, 11064, 11065, 11066, 11067, 11068, 11069, 11070, 11071, 11072, 11073, 11074, 11075, 11076, 11077, 11078, 11079, 11080, 11081, 11082, 11083, 11084, 11085, 11086, 11087, 11088, 11089, 11090, 11091, 11092, 11093, 11094, 11095, 11096, 11097, 11098, 11099, 11100, 11101, 11102, 11103, 11104, 11105, 11106, 11107, 11108, 11109, 11110, 11111, 11112, 11113, 11114, 11115, 11116, 11117, 11118, 11119, 11120, 11121, 11122, 11123, 11124, 11125, 11126, 11127, 11128, 11129, 11130, 11131, 11132, 11133, 11134, 11135, 11136, 11137, 11138, 11139, 11140, 11141, 11142, 11143, 11144, 11145, 11146, 11147, 11148, 11149, 11150, 11151, 11152, 11153, 11154, 11155, 11156, 11157, 11158, 11159, 11160, 11161, 11162, 11163, 11164, 11165, 11166, 11167, 11168, 11169, 11170, 11171, 11172, 11173, 11174, 11175, 11176, 11177, 11178, 11179, 11180, 11181, 11182, 11183, 11184, 11185, 11186, 11187, 11188, 11189, 11190, 11191, 11192, 11193, 11194, 11195, 11196, 11197, 11198, 11199, 11200, 11201, 11202, 11203, 11204, 11205, 11206, 11207, 11208, 11209, 11210, 11211, 11212, 11213, 11214, 11215, 11216, 11217, 11218, 11219, 11220, 11221, 11222, 11223, 11224, 11225, 11226, 11227, 11228, 11229, 11230, 11231, 11232, 11233, 11234, 11235, 11236, 11237, 11238, 11239, 11240, 11241, 11242, 11243, 11244, 11245, 11246, 11247, 11248, 11249, 11250, 11251, 11252, 11253, 11254, 11255, 11256, 11257, 11258, 11259, 11260, 11261, 11262, 11263, 11264, 11265, 11266, 11267, 11268, 11269, 11270, 11271, 11272, 11273, 11274, 11275, 11276, 11277, 11278, 11279, 11280, 11281, 11282, 11283, 11284, 11285, 11286, 11287, 11288, 11289, 11290, 11291, 11292, 11293, 11294, 11295, 11296, 11297, 11298, 11299, 11300, 11301, 11302, 11303, 11304, 11305, 11306, 11307, 11308, 11309, 11310, 11311, 11312, 11313, 11314, 11315, 11316, 11317, 11318, 11319, 11320, 11321, 11322, 11323, 11324, 11325, 11326, 11327, 11328, 11329, 11330, 11331, 11332, 11333, 11334, 11335, 11336, 11337, 11338, 11339, 11340, 11341, 11342, 11343, 11344, 11345, 11346, 11347, 11348, 11349, 11350, 11351, 11352, 11353, 11354, 11355, 11356, 11357, 11358, 11359, 11360, 11361, 11362, 11363, 11364, 11365, 11366, 11367, 11368, 11369, 11370, 11371, 11372, 11373, 11374, 11375, 11376, 11377, 11378, 11379, 11380, 11381, 11382, 11383, 11384,

11385, 11386, 11387, 11388, 11389, 11390, 11391, 11392, 11393, 11394, 11395, 11396, 11397, 11398, 11399, 11400, 11401, 11402, 11403, 11404, 11405, 11406, 11407, 11408, 11409, 11410, 11411, 11412, 11413, 11414, 11415, 11416, 11417, 11418, 11419, 11420, 11421, 11422, 11423, 11424, 11425, 11426, 11427, 11428, 11429, 11430, 11431, 11432, 11433, 11434, 11435, 11436, 11437, 11438, 11439, 11440, 11441, 11442, 11443, 11444, 11445, 11446, 11447, 11448, 11449, 11450, 11451, 11452, 11453, 11454, 11455, 11456, 11457, 11458, 11459, 11460, 11461, 11462, 11463, 11464, 11465, 11466, 11467, 11468, 11469, 11470, 11471, 11472, 11473, 11474, 11475, 11476, 11477, 11478, 11479, 11480, 11481, 11482, 11483, 11484, 11485, 11486, 11487, 11488, 11489, 11490, 11491, 11492, 11493, 11494, 11495, 11496, 11497, 11498, 11499, 11500, 11501, 11502, 11503, 11504, 11505, 11506, 11507, 11508, 11509, 11510, 11511, 11512, 11513, 11514, 11515, 11516, 11517, 11518, 11519, 11520, 11521, 11522, 11523, 11524, 11525, 11526, 11527, 11528, 11529, 11530, 11531, 11532, 11533, 11534, 11535, 11536, 11537, 11538, 11539, 11540, 11541, 11542, 11543, 11544, 11545, 11546, 11547, 11548, 11549, 11550, 11551, 11552, 11553, 11554, 11555, 11556, 11557, 11558, 11559, 11560, 11561, 11562, 11563, 11564, 11565, 11566, 11567, 11568, 11569, 11570, 11571, 11572, 11573, 11574, 11575, 11576, 11577, 11578, 11579, 11580, 11581, 11582, 11583, 11584, 11585, 11586, 11587, 11588, 11589, 11590, 11591, 11592, 11593, 11594, 11595, 11596, 11597, 11598, 11599, 11600, 11601, 11602, 11603, 11604, 11605, 11606, 11607, 11608, 11609, 11610, 11611, 11612, 11613, 11614, 11615, 11616, 11617, 11618, 11619, 11620, 11621, 11622, 11623, 11624, 11625, 11626, 11627, 11628, 11629, 11630, 11631, 11632, 11633, 11634, 11635, 11636, 11637, 11638, 11639, 11640, 11641, 11642, 11643, 11644, 11645, 11646, 11647, 11648, 11649, 11650, 11651, 11652, 11653, 11654, 11655, 11656, 11657, 11658, 11659, 11660, 11661, 11662, 11663, 11664, 11665, 11666, 11667, 11668, 11669, 11670, 11671, 11672, 11673, 11674, 11675, 11676, 11677, 11678, 11679, 11680, 11681, 11682, 11683, 11684, 11685, 11686, 11687, 11688, 11689, 11690, 11691, 11692, 11693, 11694, 11695, 11696, 11697, 11698, 11699, 11700, 11701, 11702, 11703, 11704, 11705, 11706, 11707, 11708, 11709, 11710, 11711, 11712, 11713, 11714, 11715, 11716, 11717, 11718, 11719, 11720, 11721, 11722, 11723, 11724, 11725, 11726, 11727, 11728, 11729, 11730, 11731, 11732, 11733, 11734, 11735, 11736, 11737, 11738, 11739, 11740, 11741, 11742, 11743, 11744, 11745, 11746, 11747, 11748, 11749, 11750, 11751, 11752, 11753, 11754, 11755, 11756, 11757, 11758, 11759, 11760, 11761, 11762, 11763, 11764, 11765, 11766, 11767, 11768, 11769, 11770, 11771, 11772, 11773, 11774, 11775, 11776, 11777, 11778, 11779, 11780, 11781, 11782, 11783, 11784, 11785, 11786, 11787, 11788, 11789, 11790, 11791, 11792, 11793, 11794, 11795, 11796, 11797, 11798, 11799, 11800, 11801, 11802, 11803, 11804, 11805, 11806, 11807, 11808, 11809, 11810, 11811, 11812, 11813, 11814, 11815, 11816, 11817, 11818, 11819, 11820, 11821, 11822, 11823, 11824, 11825, 11826, 11827, 11828, 11829, 11830, 11831, 11832, 11833, 11834, 11835, 11836, 11837, 11838, 11839, 11840, 11841, 11842, 11843, 11844, 11845, 11846, 11847, 11848, 11849, 11850, 11851, 11852, 11853, 11854, 11855, 11856, 11857, 11858, 11859, 11860, 11861, 11862, 11863, 11864, 11865, 11866, 11867, 11868, 11869, 11870, 11871, 11872, 11873, 11874, 11875, 11876, 11877, 11878, 11879, 11880, 11881, 11882, 11883, 11884, 11885, 11886, 11887, 11888, 11889, 11890, 11891, 11892, 11893, 11894, 11895, 11896, 11897, 11898, 11899, 11900, 11901, 11902, 11903, 11904, 11905, 11906, 11907, 11908, 11909, 11910, 11911, 11912, 11913, 11914, 11915, 11916, 11917, 11918, 11919, 11920, 11921, 11922, 11923, 11924, 11925, 11926, 11927, 11928, 11929, 11930, 11931, 11932, 11933, 11934, 11935, 11936, 11937, 11938, 11939, 11940, 11941, 11942, 11943, 11944, 11945, 11946, 11947, 11948, 11949, 11950, 11951, 11952, 11953, 11954, 11955, 11956, 11957, 11958, 11959, 11960, 11961, 11962, 11963, 11964, 11965, 11966, 11967, 11968, 11969, 11970, 11971, 11972, 11973, 11974, 11975, 11976, 11977, 11978, 11979, 11980, 11981, 11982, 11983, 11984, 11985, 11986, 11987, 11988, 11989, 11990, 11991, 11992, 11993, 11994, 11995, 11996, 11997, 11998, 11999, 12000, 12001, 12002, 12003, 12004, 12005, 12006, 12007, 12008,

12009, 12010, 12011, 12012, 12013, 12014, 12015, 12016, 12017, 12018, 12019, 12020, 12021, 12022, 12023 and 12024.

9. A nucleic acid molecule comprising a fragment of 12 or more consecutive nucleotides from a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, 891, 893, 895, 897, 899, 901, 903, 905, 907, 909, 911, 913, 915, 917, 919, 921, 923, 925, 927, 929, 931, 933, 935, 937, 939, 941, 943, 945, 947, 949, 951, 953, 955, 957, 959, 961, 963, 965, 967, 969, 971, 973, 975, 977, 979, 981, 983, 985, 987, 989, 991, 993, 995, 997, 999, 1001, 1003, 1005, 1007, 1009, 1011, 1013, 1015, 1017, 1019, 1021, 1023, 1025, 1027, 1029, 1031, 1033, 1035, 1037, 1039, 1041, 1043, 1045, 1047, 1049, 1051, 1053, 1055, 1057, 1059, 1061, 1063, 1065, 1067, 1069, 1071, 1073, 1075, 1077, 1079, 1081, 1083, 1085, 1087, 1089, 1091, 1093, 1095, 1097, 1099, 1101, 1103, 1105, 1107, 1109, 1111, 1113, 1115, 1117, 1119, 1121, 1123, 1125, 1127, 1129, 1131, 1133, 1135, 1137, 1139, 1141, 1143, 1145, 1147, 1149, 1151, 1153, 1155, 1157, 1159, 1161, 1163, 1165, 1167, 1169, 1171, 1173, 1175, 1177, 1179, 1181, 1183, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1221, 1223, 1225, 1227, 1229, 1231, 1233, 1235, 1237, 1239, 1241, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, 1273, 1275, 1277, 1279, 1281, 1283, 1285, 1287, 1289, 1291, 1293, 1295, 1297, 1299, 1301, 1303, 1305, 1307, 1309, 1311, 1313, 1315, 1317, 1319, 1321, 1323, 1325, 1327, 1329, 1331, 1333, 1335, 1337, 1339, 1341, 1343, 1345, 1347, 1349, 1351, 1353, 1355, 1357, 1359, 1361, 1363, 1365, 1367, 1369, 1371, 1373, 1375, 1377, 1379, 1381, 1383, 1385, 1387, 1389, 1391, 1393, 1395, 1397, 1399, 1401, 1403, 1405, 1407, 1409, 1411, 1413, 1415, 1417, 1419, 1421, 1423, 1425, 1427, 1429, 1431, 1433, 1435, 1437, 1439, 1441, 1443, 1445, 1447, 1449, 1451, 1453, 1455, 1457, 1459, 1461, 1463, 1465, 1467, 1469, 1471, 1473, 1475, 1477, 1479, 1481, 1483, 1485, 1487, 1489, 1491, 1493, 1495, 1497, 1499, 1501, 1503, 1505, 1507, 1509, 1511, 1513, 1515, 1517, 1519, 1521, 1523, 1525, 1527, 1529, 1531, 1533, 1535, 1537, 1539, 1541, 1543, 1545, 1547, 1549, 1551, 1553, 1555, 1557, 1559, 1561, 1563, 1565, 1567, 1569, 1571, 1573, 1575, 1577, 1579, 1581, 1583, 1585, 1587, 1589, 1591, 1593, 1595, 1597, 1599, 1601, 1603, 1605, 1607, 1609, 1611, 1613, 1615, 1617, 1619, 1621, 1623,

1625, 1627, 1629, 1631, 1633, 1635, 1637, 1639, 1641, 1643, 1645, 1647, 1649, 1651, 1653, 1655, 1657, 1659, 1661, 1663, 1665, 1667, 1669, 1671, 1673, 1675, 1677, 1679, 1681, 1683, 1685, 1687, 1689, 1691, 1693, 1695, 1697, 1699, 1701, 1703, 1705, 1707, 1709, 1711, 1713, 1715, 1717, 1719, 1721, 1723, 1725, 1727, 1729, 1731, 1733, 1735, 1737, 1739, 1741, 1743, 1745, 1747, 1749, 1751, 1753, 1755, 1757, 1759, 1761, 1763, 1765, 1767, 1769, 1771, 1773, 1775, 1777, 1779, 1781, 1783, 1785, 1787, 1789, 1791, 1793, 1795, 1797, 1799, 1801, 1803, 1805, 1807, 1809, 1811, 1813, 1815, 1817, 1819, 1821, 1823, 1825, 1827, 1829, 1831, 1833, 1835, 1837, 1839, 1841, 1843, 1845, 1847, 1849, 1851, 1853, 1855, 1857, 1859, 1861, 1863, 1865, 1867, 1869, 1871, 1873, 1875, 1877, 1879, 1881, 1883, 1885, 1887, 1889, 1891, 1893, 1895, 1897, 1899, 1901, 1903, 1905, 1907, 1909, 1911, 1913, 1915, 1917, 1919, 1921, 1923, 1925, 1927, 1929, 1931, 1933, 1935, 1937, 1939, 1941, 1943, 1945, 1947, 1949, 1951, 1953, 1955, 1957, 1959, 1961, 1963, 1965, 1967, 1969, 1971, 1973, 1975, 1977, 1979, 1981, 1983, 1985, 1987, 1989, 1991, 1993, 1995, 1997, 1999, 2001, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017, 2019, 2021, 2023, 2025, 2027, 2029, 2031, 2033, 2035, 2037, 2039, 2041, 2043, 2045, 2047, 2049, 2051, 2053, 2055, 2057, 2059, 2061, 2063, 2065, 2067, 2069, 2071, 2073, 2075, 2077, 2079, 2081, 2083, 2085, 2087, 2089, 2091, 2093, 2095, 2097, 2099, 2101, 2103, 2105, 2107, 2109, 2111, 2113, 2115, 2117, 2119, 2121, 2123, 2125, 2127, 2129, 2131, 2133, 2135, 2137, 2139, 2141, 2143, 2145, 2147, 2149, 2151, 2153, 2155, 2157, 2159, 2161, 2163, 2165, 2167, 2169, 2171, 2173, 2175, 2177, 2179, 2181, 2183, 2185, 2187, 2189, 2191, 2193, 2195, 2197, 2199, 2201, 2203, 2205, 2207, 2209, 2211, 2213, 2215, 2217, 2219, 2221, 2223, 2225, 2227, 2229, 2231, 2233, 2235, 2237, 2239, 2241, 2243, 2245, 2247, 2249, 2251, 2253, 2255, 2257, 2259, 2261, 2263, 2265, 2267, 2269, 2271, 2273, 2275, 2277, 2279, 2281, 2283, 2285, 2287, 2289, 2291, 2293, 2295, 2297, 2299, 2301, 2303, 2305, 2307, 2309, 2311, 2313, 2315, 2317, 2319, 2321, 2323, 2325, 2327, 2329, 2331, 2333, 2335, 2337, 2339, 2341, 2343, 2345, 2347, 2349, 2351, 2353, 2355, 2357, 2359, 2361, 2363, 2365, 2367, 2369, 2371, 2373, 2375, 2377, 2379, 2381, 2383, 2385, 2387, 2389, 2391, 2393, 2395, 2397, 2399, 2401, 2403, 2405, 2407, 2409, 2411, 2413, 2415, 2417, 2419, 2421, 2423, 2425, 2427, 2429, 2431, 2433, 2435, 2437, 2439, 2441, 2443, 2445, 2447, 2449, 2451, 2453, 2455, 2457, 2459, 2461, 2463, 2465, 2467, 2469, 2471, 2473, 2475, 2477, 2479, 2481, 2483, 2485, 2487, 2489, 2491, 2493, 2495, 2497, 2499, 2501, 2503, 2505, 2507, 2509, 2511, 2513, 2515, 2517, 2519, 2521, 2523, 2525, 2527, 2529, 2531, 2533, 2535, 2537, 2539, 2541, 2543, 2545, 2547, 2549, 2551, 2553, 2555, 2557, 2559, 2561, 2563, 2565, 2567, 2569, 2571, 2573, 2575, 2577, 2579, 2581, 2583, 2585, 2587, 2589, 2591, 2593, 2595, 2597, 2599, 2601, 2603, 2605, 2607, 2609, 2611, 2613, 2615, 2617, 2619, 2621, 2623, 2625, 2627, 2629, 2631, 2633, 2635, 2637, 2639, 2641, 2643, 2645, 2647, 2649, 2651, 2653, 2655, 2657, 2659, 2661, 2663, 2665, 2667, 2669, 2671, 2673, 2675, 2677, 2679, 2681, 2683, 2685, 2687, 2689, 2691, 2693, 2695, 2697, 2699, 2701, 2703, 2705, 2707, 2709, 2711, 2713, 2715, 2717, 2719, 2721, 2723, 2725, 2727, 2729, 2731, 2733, 2735, 2737, 2739, 2741, 2743, 2745, 2747, 2749, 2751, 2753, 2755, 2757, 2759, 2761, 2763, 2765, 2767, 2769, 2771, 2773, 2775, 2777, 2779, 2781, 2783, 2785, 2787, 2789, 2791, 2793, 2795, 2797, 2799, 2801, 2803, 2805, 2807, 2809, 2811, 2813, 2815, 2817, 2819, 2821, 2823, 2825, 2827, 2829, 2831, 2833, 2835, 2837, 2839, 2841, 2843, 2845, 2847, 2849, 2851, 2853, 2855, 2857, 2859, 2861, 2863, 2865, 2867, 2869, 2871, 2873, 2875, 2877, 2879, 2881, 2883, 2885, 2887, 2889, 2891, 2893, 2895, 2897, 2899, 2901, 2903, 2905, 2907, 2909, 2911, 2913, 2915, 2917, 2919, 2921, 2923, 2925, 2927, 2929, 2931, 2933, 2935, 2937, 2939, 2941, 2943, 2945, 2947, 2949, 2951, 2953, 2955, 2957, 2959, 2961, 2963, 2965, 2967, 2969, 2971, 2973, 2975, 2977, 2979, 2981, 2983, 2985, 2987, 2989, 2991, 2993, 2995, 2997, 2999, 3001, 3003, 3005, 3007, 3009, 3011, 3013, 3015, 3017, 3019, 3021, 3023, 3025, 3027, 3029, 3031, 3033, 3035, 3037, 3039, 3041, 3043, 3045, 3047, 3049, 3051, 3053, 3055, 3057, 3059, 3061, 3063, 3065, 3067, 3069, 3071, 3073, 3075, 3077, 3079, 3081, 3083, 3085, 3087, 3089, 3091, 3093, 3095, 3097, 3099, 3101, 3103, 3105, 3107, 3109, 3111, 3113, 3115, 3117, 3119, 3121, 3123, 3125, 3127, 3129, 3131, 3133, 3135, 3137, 3139, 3141, 3143,

3145, 3147, 3149, 3151, 3153, 3155, 3157, 3159, 3161, 3163, 3165, 3167, 3169, 3171, 3173, 3175, 3177, 3179, 3181, 3183, 3185, 3187, 3189, 3191, 3193, 3195, 3197, 3199, 3201, 3203, 3205, 3207, 3209, 3211, 3213, 3215, 3217, 3219, 3221, 3223, 3225, 3227, 3229, 3231, 3233, 3235, 3237, 3239, 3241, 3243, 3245, 3247, 3249, 3251, 3253, 3255, 3257, 3259, 3261, 3263, 3265, 3267, 3269, 3271, 3273, 3275, 3277, 3279, 3281, 3283, 3285, 3287, 3289, 3291, 3293, 3295, 3297, 3299, 3301, 3303, 3305, 3307, 3309, 3311, 3313, 3315, 3317, 3319, 3321, 3323, 3325, 3327, 3329, 3331, 3333, 3335, 3337, 3339, 3341, 3343, 3345, 3347, 3349, 3351, 3353, 3355, 3357, 3359, 3361, 3363, 3365, 3367, 3369, 3371, 3373, 3375, 3377, 3379, 3381, 3383, 3385, 3387, 3389, 3391, 3393, 3395, 3397, 3399, 3401, 3403, 3405, 3407, 3409, 3411, 3413, 3415, 3417, 3419, 3421, 3423, 3425, 3427, 3429, 3431, 3433, 3435, 3437, 3439, 3441, 3443, 3445, 3447, 3449, 3451, 3453, 3455, 3457, 3459, 3461, 3463, 3465, 3467, 3469, 3471, 3473, 3475, 3477, 3479, 3481, 3483, 3485, 3487, 3489, 3491, 3493, 3495, 3497, 3499, 3501, 3503, 3505, 3507, 3509, 3511, 3513, 3515, 3517, 3519, 3521, 3523, 3525, 3527, 3529, 3531, 3533, 3535, 3537, 3539, 3541, 3543, 3545, 3547, 3549, 3551, 3553, 3555, 3557, 3559, 3561, 3563, 3565, 3567, 3569, 3571, 3573, 3575, 3577, 3579, 3581, 3583, 3585, 3587, 3589, 3591, 3593, 3595, 3597, 3599, 3601, 3603, 3605, 3607, 3609, 3611, 3613, 3615, 3617, 3619, 3621, 3623, 3625, 3627, 3629, 3631, 3633, 3635, 3637, 3639, 3641, 3643, 3645, 3647, 3649, 3651, 3653, 3655, 3657, 3659, 3661, 3663, 3665, 3667, 3669, 3671, 3673, 3675, 3677, 3679, 3681, 3683, 3685, 3687, 3689, 3691, 3693, 3695, 3697, 3699, 3701, 3703, 3705, 3707, 3709, 3711, 3713, 3715, 3717, 3719, 3721, 3723, 3725, 3727, 3729, 3731, 3733, 3735, 3737, 3739, 3741, 3743, 3747, 3749, 3751, 3753, 3755, 3757, 3759, 3761, 3763, 3765, 3767, 3769, 3771, 3773, 3775, 3777, 3779, 3781, 3783, 3785, 3787, 3789, 3791, 3793, 3795, 3797, 3799, 3801, 3803, 3805, 3807, 3809, 3811, 3813, 3815, 3817, 3819, 3821, 3823, 3825, 3827, 3829, 3831, 3833, 3835, 3837, 3839, 3841, 3843, 3845, 3847, 3849, 3851, 3853, 3855, 3857, 3859, 3861, 3863, 3865, 3867, 3869, 3871, 3873, 3875, 3877, 3879, 3881, 3883, 3885, 3887, 3889, 3891, 3893, 3895, 3897, 3899, 3901, 3903, 3905, 3907, 3909, 3911, 3913, 3915, 3917, 3919, 3923, 3925, 3927, 3929, 3931, 3933, 3935, 3937, 3939, 3941, 3943, 3945, 3947, 3949, 3951, 3953, 3955, 3957, 3959, 3961, 3963, 3965, 3967, 3969, 3971, 3973, 3975, 3977, 3979, 3981, 3983, 3985, 3987, 3989, 3991, 3993, 3995, 3997, 3999, 4001, 4003, 4005, 4007, 4009, 4011, 4013, 4015, 4017, 4019, 4021, 4023, 4025, 4027, 4029, 4031, 4033, 4035, 4037, 4039, 4041, 4043, 4045, 4047, 4049, 4051, 4053, 4055, 4057, 4059, 4061, 4063, 4065, 4067, 4069, 4071, 4073, 4075, 4077, 4079, 4081, 4083, 4085, 4087, 4089, 4091, 4093, 4095, 4097, 4099, 4101, 4103, 4105, 4107, 4109, 4111, 4113, 4115, 4117, 4119, 4121, 4123, 4125, 4127, 4129, 4131, 4133, 4135, 4137, 4139, 4141, 4143, 4145, 4147, 4149, 4151, 4153, 4155, 4157, 4159, 4161, 4163, 4165, 4167, 4169, 4171, 4173, 4175, 4177, 4179, 4181, 4183, 4185, 4187, 4189, 4191, 4193, 4195, 4197, 4199, 4201, 4203, 4205, 4207, 4209, 4211, 4213, 4215, 4217, 4219, 4221, 4223, 4225, 4227, 4229, 4231, 4233, 4235, 4237, 4239, 4241, 4243, 4245, 4247, 4249, 4251, 4253, 4255, 4257, 4259, 4261, 4263, 4265, 4267, 4269, 4271, 4273, 4275, 4277, 4279, 4281, 4283, 4285, 4287, 4289, 4291, 4293, 4295, 4297, 4299, 4301, 4303, 4305, 4307, 4309, 4311, 4313, 4315, 4317, 4319, 4321, 4323, 4325, 4327, 4329, 4331, 4333, 4335, 4337, 4339, 4341, 4343, 4345, 4347, 4349, 4351, 4353, 4355, 4357, 4359, 4361, 4363, 4365, 4367, 4369, 4371, 4373, 4375, 4377, 4379, 4381, 4383, 4385, 4387, 4389, 4391, 4393, 4395, 4397, 4399, 4401, 4403, 4405, 4407, 4409, 4411, 4413, 4415, 4417, 4419, 4421, 4423, 4425, 4427, 4429, 4431, 4433, 4435, 4437, 4439, 4441, 4443, 4445, 4447, 4449, 4451, 4453, 4455, 4457, 4459, 4461, 4463, 4465, 4467, 4469, 4471, 4473, 4475, 4477, 4479, 4481, 4483, 4485, 4487, 4489, 4491, 4493, 4495, 4497, 4499, 4501, 4503, 4505, 4507, 4509, 4511, 4513, 4515, 4517, 4519, 4521, 4523, 4525, 4527, 4529, 4531, 4533, 4535, 4537, 4539, 4541, 4543, 4545, 4547, 4549, 4551, 4553, 4555, 4557, 4559, 4561, 4563, 4565, 4567, 4569, 4571, 4573, 4575, 4577, 4579, 4581, 4583, 4585, 4587, 4589, 4591, 4593, 4595, 4597, 4599, 4601, 4603, 4605, 4607, 4609, 4611, 4613, 4615, 4617, 4619, 4621, 4623, 4625, 4627, 4629, 4631, 4633, 4635, 4637, 4639, 4641, 4643, 4645, 4647, 4649, 4651, 4653, 4655, 4657, 4659, 4661, 4663,

4665, 4667, 4669, 4671, 4673, 4675, 4677, 4679, 4681, 4683, 4685, 4687, 4689, 4691, 4693, 4695, 4697, 4699, 4701, 4703, 4705, 4707, 4709, 4711, 4713, 4715, 4717, 4719, 4721, 4723, 4725, 4727, 4729, 4731, 4733, 4735, 4737, 4739, 4741, 4743, 4745, 4747, 4749, 4751, 4753, 4755, 4757, 4759, 4761, 4763, 4765, 4767, 4769, 4771, 4773, 4775, 4777, 4779, 4781, 4783, 4785, 4787, 4789, 4791, 4793, 4795, 4797, 4799, 4801, 4803, 4805, 4807, 4809, 4811, 4813, 4815, 4817, 4819, 4821, 4823, 4825, 4827, 4829, 4831, 4833, 4835, 4837, 4839, 4841, 4843, 4845, 4847, 4849, 4851, 4853, 4855, 4857, 4859, 4861, 4863, 4865, 4867, 4869, 4871, 4873, 4875, 4877, 4879, 4881, 4883, 4885, 4887, 4889, 4891, 4893, 4895, 4897, 4899, 4901, 4903, 4905, 4907, 4909, 4911, 4913, 4915, 4917, 4919, 4921, 4923, 4925, 4927, 4929, 4931, 4933, 4935, 4937, 4939, 4941, 4943, 4945, 4947, 4949, 4951, 4953, 4955, 4957, 4959, 4961, 4963, 4965, 4967, 4969, 4971, 4973, 4975, 4977, 4979, 4981, 4983, 4985, 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5001, 5003, 5005, 5007, 5009, 5011, 5013, 5015, 5017, 5019, 5021, 5023, 5025, 5027, 5029, 5031, 5033, 5035, 5037, 5039, 5041, 5043, 5045, 5047, 5049, 5051, 5053, 5055, 5057, 5059, 5061, 5063, 5065, 5067, 5069, 5071, 5073, 5075, 5077, 5079, 5081, 5083, 5085, 5087, 5089, 5091, 5093, 5095, 5097, 5099, 5101, 5103, 5105, 5107, 5109, 5111, 5113, 5115, 5117, 5119, 5121, 5123, 5125, 5127, 5129, 5131, 5133, 5135, 5137, 5139, 5141, 5143, 5145, 5147, 5149, 5151, 5153, 5155, 5157, 5159, 5161, 5163, 5165, 5167, 5169, 5171, 5173, 5175, 5177, 5179, 5181, 5183, 5185, 5187, 5189, 5191, 5193, 5195, 5197, 5199, 5201, 5203, 5205, 5207, 5209, 5211, 5213, 5215, 5217, 5219, 5221, 5223, 5225, 5227, 5229, 5231, 5233, 5235, 5237, 5239, 5241, 5243, 5245, 5247, 5249, 5251, 5253, 5255, 5257, 5259, 5261, 5263, 5265, 5267, 5269, 5271, 5273, 5275, 5277, 5279, 5281, 5283, 5285, 5287, 5289, 5291, 5293, 5295, 5297, 5299, 5301, 5303, 5305, 5307, 5309, 5311, 5313, 5315, 5317, 5319, 5321, 5323, 5325, 5327, 5329, 5331, 5333, 5335, 5337, 5339, 5341, 5343, 5345, 5347, 5349, 5351, 5353, 5355, 5357, 5359, 5361, 5363, 5365, 5367, 5369, 5371, 5373, 5375, 5377, 5379, 5381, 5383, 5385, 5387, 5389, 5391, 5393, 5395, 5397, 5399, 5401, 5403, 5405, 5407, 5409, 5411, 5413, 5415, 5417, 5419, 5421, 5423, 5425, 5427, 5429, 5431, 5433, 5435, 5437, 5439, 5441, 5443, 5445, 5447, 5449, 5451, 5453, 5455, 5457, 5459, 5461, 5463, 5465, 5467, 5469, 5471, 5473, 5475, 5477, 5479, 5481, 5483, 5485, 5487, 5489, 5491, 5493, 5495, 5497, 5499, 5501, 5503, 5505, 5507, 5509, 5511, 5513, 5515, 5517, 5519, 5521, 5523, 5525, 5527, 5529, 5531, 5533, 5535, 5537, 5539, 5541, 5543, 5545, 5547, 5549, 5551, 5553, 5555, 5557, 5559, 5561, 5563, 5565, 5567, 5569, 5571, 5573, 5575, 5577, 5579, 5581, 5583, 5585, 5587, 5589, 5591, 5593, 5595, 5597, 5599, 5601, 5603, 5605, 5607, 5609, 5611, 5613, 5615, 5617, 5619, 5621, 5623, 5625, 5627, 5629, 5631, 5633, 5635, 5637, 5639, 5641, 5643, 5645, 5647, 5649, 5651, 5653, 5655, 5657, 5659, 5661, 5663, 5665, 5667, 5669, 5671, 5673, 5675, 5677, 5679, 5681, 5683, 5685, 5687, 5689, 5691, 5693, 5695, 5697, 5699, 5701, 5703, 5705, 5707, 5709, 5711, 5713, 5715, 5717, 5719, 5721, 5723, 5725, 5727, 5729, 5731, 5733, 5735, 5737, 5739, 5741, 5743, 5745, 5747, 5749, 5751, 5753, 5755, 5757, 5759, 5761, 5763, 5765, 5767, 5769, 5771, 5773, 5775, 5777, 5779, 5781, 5783, 5785, 5787, 5789, 5791, 5793, 5795, 5797, 5799, 5801, 5803, 5805, 5807, 5809, 5811, 5813, 5815, 5817, 5819, 5821, 5823, 5825, 5827, 5829, 5831, 5833, 5835, 5837, 5839, 5841, 5843, 5845, 5847, 5849, 5851, 5853, 5855, 5857, 5859, 5861, 5863, 5865, 5867, 5869, 5871, 5873, 5875, 5877, 5879, 5881, 5883, 5885, 5887, 5889, 5891, 5893, 5895, 5897, 5899, 5901, 5903, 5905, 5907, 5909, 5911, 5913, 5915, 5917, 5919, 5921, 5923, 5925, 5927, 5929, 5931, 5933, 5935, 5937, 5939, 5941, 5943, 5945, 5947, 5949, 5951, 5953, 5955, 5957, 5959, 5961, 5963, 5965, 5967, 5969, 5971, 5973, 5975, 5977, 5979, 5981, 5983, 5985, 5987, 5989, 5991, 5993, 5995, 5997, 5999, 6001, 6003, 6005, 6007, 6009, 6011, 6013, 6015, 6017, 6019, 6021, 6023, 6025, 6027, 6029, 6031, 6033, 6035, 6037, 6039, 6041, 6043, 6045, 6047, 6049, 6051, 6053, 6055, 6057, 6059, 6061, 6063, 6065, 6067, 6069, 6071, 6073, 6075, 6077, 6079, 6081, 6083, 6085, 6087, 6089, 6091, 6093, 6095, 6097, 6099, 6101, 6103, 6105, 6107, 6109, 6111, 6113, 6115, 6117, 6119, 6121, 6123, 6125, 6127, 6129, 6131, 6133, 6135, 6137, 6139, 6141, 6143, 6145, 6147, 6149, 6151, 6153, 6155, 6157, 6159, 6161, 6163, 6165, 6167, 6169, 6171, 6173, 6175, 6177, 6179, 6181, 6183,

6185, 6187, 6189, 6191, 6193, 6195, 6197, 6199, 6201, 6203, 6205, 6207, 6209, 6211, 6213, 6215, 6217, 6219, 6221, 6223, 6225, 6227, 6229, 6231, 6233, 6235, 6237, 6239, 6241, 6243, 6245, 6247, 6249, 6251, 6253, 6255, 6257, 6259, 6261, 6263, 6265, 6267, 6269, 6271, 6273, 6275, 6277, 6279, 6281, 6283, 6285, 6287, 6289, 6291, 6293, 6295, 6297, 6299, 6301, 6303, 6305, 6307, 6309, 6311, 6313, 6315, 6317, 6319, 6321, 6323, 6325, 6327, 6329, 6331, 6333, 6335, 6337, 6339, 6341, 6343, 6345, 6347, 6349, 6351, 6353, 6355, 6357, 6359, 6361, 6363, 6365, 6367, 6369, 6371, 6373, 6375, 6377, 6379, 6381, 6383, 6385, 6387, 6389, 6391, 6393, 6395, 6397, 6399, 6401, 6403, 6405, 6407, 6409, 6411, 6413, 6415, 6417, 6419, 6421, 6423, 6425, 6427, 6429, 6431, 6433, 6435, 6437, 6439, 6441, 6443, 6445, 6447, 6449, 6451, 6453, 6455, 6457, 6459, 6461, 6463, 6465, 6467, 6469, 6471, 6473, 6475, 6477, 6479, 6481, 6483, 6485, 6487, 6489, 6491, 6493, 6495, 6497, 6499, 6501, 6503, 6505, 6507, 6509, 6511, 6513, 6515, 6517, 6519, 6521, 6523, 6525, 6527, 6529, 6531, 6533, 6535, 6537, 6539, 6541, 6543, 6545, 6547, 6549, 6551, 6553, 6555, 6557, 6559, 6561, 6563, 6565, 6567, 6569, 6571, 6573, 6575, 6577, 6579, 6581, 6583, 6585, 6587, 6589, 6591, 6593, 6595, 6597, 6599, 6601, 6603, 6605, 6607, 6609, 6611, 6613, 6615, 6617, 6619, 6621, 6623, 6625, 6627, 6629, 6631, 6633, 6635, 6637, 6639, 6641, 6643, 6645, 6647, 6649, 6651, 6653, 6655, 6657, 6659, 6661, 6663, 6665, 6667, 6669, 6671, 6673, 6675, 6677, 6679, 6681, 6683, 6685, 6687, 6689, 6691, 6693, 6695, 6697, 6699, 6701, 6703, 6705, 6707, 6709, 6711, 6713, 6715, 6717, 6719, 6721, 6723, 6725, 6727, 6729, 6731, 6733, 6735, 6737, 6739, 6741, 6743, 6745, 6747, 6749, 6751, 6753, 6755, 6757, 6759, 6761, 6763, 6765, 6767, 6769, 6771, 6773, 6775, 6777, 6779, 6781, 6783, 6785, 6787, 6789, 6791, 6793, 6795, 6797, 6799, 6801, 6803, 6805, 6807, 6809, 6811, 6813, 6815, 6817, 6819, 6821, 6823, 6825, 6827, 6829, 6831, 6833, 6835, 6837, 6839, 6841, 6843, 6845, 6847, 6849, 6851, 6853, 6855, 6857, 6859, 6861, 6863, 6865, 6867, 6869, 6871, 6873, 6875, 6877, 6879, 6881, 6883, 6885, 6887, 6889, 6891, 6893, 6895, 6897, 6899, 6901, 6903, 6905, 6907, 6909, 6911, 6913, 6915, 6917, 6919, 6921, 6923, 6925, 6927, 6929, 6931, 6933, 6935, 6937, 6939, 6941, 6943, 6945, 6947, 6949, 6951, 6953, 6955, 6957, 6959, 6961, 6963, 6965, 6967, 6969, 6971, 6973, 6975, 6977, 6979, 6981, 6983, 6985, 6987, 6989, 6991, 6993, 6995, 6997, 6999, 7001, 7003, 7005, 7007, 7009, 7011, 7013, 7015, 7017, 7019, 7021, 7023, 7025, 7027, 7029, 7031, 7033, 7035, 7037, 7039, 7041, 7043, 7045, 7047, 7049, 7051, 7053, 7055, 7057, 7059, 7061, 7063, 7065, 7067, 7069, 7071, 7073, 7075, 7077, 7079, 7081, 7083, 7085, 7087, 7089, 7091, 7093, 7095, 7097, 7099, 7101, 7103, 7105, 7107, 7109, 7111, 7113, 7115, 7117, 7119, 7121, 7123, 7125, 7127, 7129, 7131, 7133, 7135, 7137, 7139, 7141, 7143, 7145, 7147, 7149, 7151, 7153, 7155, 7157, 7159, 7161, 7163, 7165, 7167, 7169, 7171, 7173, 7175, 7177, 7179, 7181, 7183, 7185, 7187, 7189, 7191, 7193, 7195, 7197, 7199, 7201, 7203, 7205, 7207, 7209, 7211, 7213, 7215, 7217, 7219, 7221, 7223, 7225, 7227, 7229, 7231, 7233, 7235, 7237, 7239, 7241, 7243, 7245, 7247, 7249, 7251, 7253, 7255, 7257, 7259, 7261, 7263, 7265, 7267, 7269, 7271, 7273, 7275, 7277, 7279, 7281, 7283, 7285, 7287, 7289, 7291, 7293, 7295, 7297, 7299, 7301, 7303, 7305, 7307, 7309, 7311, 7313, 7315, 7317, 7319, 7321, 7323, 7325, 7327, 7329, 7331, 7333, 7335, 7337, 7339, 7341, 7343, 7345, 7347, 7349, 7351, 7353, 7355, 7357, 7359, 7361, 7363, 7365, 7367, 7369, 7371, 7373, 7375, 7377, 7379, 7381, 7383, 7385, 7387, 7389, 7391, 7393, 7395, 7397, 7399, 7401, 7403, 7405, 7407, 7409, 7411, 7413, 7415, 7417, 7419, 7421, 7423, 7425, 7427, 7429, 7431, 7433, 7435, 7437, 7439, 7441, 7443, 7445, 7447, 7449, 7451, 7453, 7455, 7457, 7459, 7461, 7463, 7467, 7469, 7471, 7473, 7475, 7477, 7479, 7481, 7483, 7485, 7487, 7489, 7491, 7493, 7495, 7497, 7499, 7501, 7503, 7505, 7507, 7509, 7511, 7513, 7515, 7517, 7519, 7521, 7523, 7525, 7527, 7529, 7531, 7533, 7535, 7537, 7539, 7541, 7543, 7545, 7547, 7549, 7551, 7553, 7555, 7557, 7559, 7561, 7563, 7565, 7567, 7569, 7571, 7573, 7575, 7577, 7579, 7581, 7583, 7585, 7587, 7589, 7591, 7593, 7595, 7597, 7599, 7601, 7603, 7605, 7607, 7609, 7611, 7613, 7615, 7617, 7619, 7621, 7623, 7625, 7627, 7629, 7631, 7633, 7635, 7637, 7639, 7641, 7643, 7645, 7647, 7649, 7651, 7653, 7655, 7657, 7659, 7661, 7663, 7665, 7667, 7669, 7671, 7673, 7675, 7677, 7679, 7681, 7683, 7685, 7687, 7689, 7691, 7693, 7695, 7697, 7699, 7701, 7703,

7705, 7707, 7709, 7711, 7713, 7715, 7717, 7719, 7721, 7723, 7725, 7727, 7729, 7731, 7733, 7735, 7737, 7739, 7741, 7743, 7745, 7747, 7749, 7751, 7753, 7755, 7757, 7759, 7761, 7763, 7765, 7767, 7769, 7771, 7773, 7775, 7777, 7779, 7781, 7783, 7785, 7787, 7789, 7791, 7793, 7795, 7797, 7799, 7801, 7803, 7805, 7807, 7809, 7811, 7813, 7815, 7817, 7819, 7821, 7823, 7825, 7827, 7829, 7831, 7833, 7835, 7837, 7839, 7841, 7843, 7845, 7847, 7849, 7851, 7853, 7855, 7857, 7859, 7861, 7863, 7865, 7867, 7869, 7871, 7873, 7875, 7877, 7879, 7881, 7883, 7885, 7887, 7889, 7891, 7893, 7895, 7897, 7899, 7901, 7903, 7905, 7907, 7909, 7911, 7913, 7915, 7917, 7919, 7921, 7923, 7925, 7927, 7929, 7931, 7933, 7935, 7937, 7939, 7941, 7943, 7945, 7947, 7949, 7951, 7953, 7955, 7957, 7959, 7961, 7963, 7965, 7967, 7969, 7971, 7973, 7975, 7977, 7979, 7981, 7983, 7985, 7987, 7989, 7991, 7993, 7995, 7997, 7999, 8001, 8003, 8005, 8007, 8009, 8011, 8013, 8015, 8017, 8019, 8021, 8023, 8025, 8027, 8029, 8031, 8033, 8035, 8037, 8039, 8041, 8043, 8045, 8047, 8049, 8051, 8053, 8055, 8057, 8059, 8061, 8063, 8065, 8067, 8069, 8071, 8073, 8075, 8077, 8079, 8081, 8083, 8085, 8087, 8089, 8091, 8093, 8095, 8097, 8099, 8101, 8103, 8105, 8107, 8109, 8111, 8113, 8115, 8117, 8119, 8121, 8123, 8125, 8127, 8129, 8131, 8133, 8135, 8137, 8139, 8141, 8143, 8145, 8147, 8149, 8151, 8153, 8155, 8157, 8159, 8161, 8163, 8165, 8167, 8169, 8171, 8173, 8175, 8177, 8179, 8181, 8183, 8185, 8187, 8189, 8191, 8193, 8195, 8197, 8199, 8201, 8203, 8205, 8207, 8209, 8211, 8213, 8215, 8217, 8219, 8221, 8223, 8225, 8227, 8229, 8231, 8233, 8235, 8237, 8239, 8241, 8243, 8245, 8247, 8249, 8251, 8253, 8255, 8257, 8259, 8261, 8263, 8265, 8267, 8269, 8271, 8273, 8275, 8277, 8279, 8281, 8283, 8285, 8287, 8289, 8291, 8293, 8295, 8297, 8299, 8301, 8303, 8305, 8307, 8309, 8311, 8313, 8315, 8317, 8319, 8321, 8323, 8325, 8327, 8329, 8331, 8333, 8335, 8337, 8339, 8341, 8343, 8345, 8347, 8349, 8351, 8353, 8355, 8357, 8359, 8361, 8363, 8365, 8367, 8369, 8371, 8373, 8375, 8377, 8379, 8381, 8383, 8385, 8387, 8389, 8391, 8393, 8395, 8397, 8399, 8401, 8403, 8405, 8407, 8409, 8411, 8413, 8415, 8417, 8419, 8421, 8423, 8425, 8427, 8429, 8431, 8433, 8435, 8437, 8439, 8441, 8443, 8445, 8447, 8449, 8451, 8453, 8455, 8457, 8459, 8461, 8463, 8465, 8467, 8469, 8471, 8473, 8475, 8477, 8479, 8481, 8483, 8485, 8487, 8489, 8491, 8493, 8495, 8497, 8499, 8501, 8503, 8505, 8507, 8509, 8511, 8513, 8515, 8517, 8519, 8521, 8523, 8525, 8527, 8529, 8531, 8533, 8535, 8537, 8539, 8541, 8543, 8545, 8547, 8549, 8551, 8553, 8555, 8557, 8559, 8561, 8563, 8565, 8567, 8569, 8571, 8573, 8575, 8577, 8579, 8581, 8583, 8585, 8587, 8589, 8591, 8593, 8595, 8597, 8599, 8601, 8603, 8605, 8607, 8609, 8611, 8613, 8615, 8617, 8619, 8621, 8623, 8625, 8627, 8629, 8631, 8633, 8635, 8637, 8639, 8641, 8643, 8645, 8647, 8649, 8651, 8653, 8655, 8657, 8659, 8661, 8663, 8665, 8667, 8669, 8671, 8673, 8675, 8677, 8679, 8681, 8683, 8685, 8687, 8689, 8691, 8693, 8695, 8697, 8699, 8701, 8703, 8705, 8707, 8709, 8711, 8713, 8715, 8717, 8719, 8721, 8723, 8725, 8727, 8729, 8731, 8733, 8735, 8737, 8739, 8741, 8743, 8745, 8747, 8749, 8751, 8753, 8755, 8757, 8759, 8761, 8763, 8765, 8767, 8769, 8771, 8773, 8775, 8777, 8781, 8783, 8785, 8787, 8789, 8791, 8793, 8795, 8797, 8799, 8801, 8803, 8805, 8807, 8809, 8811, 8813, 8815, 8817, 8819, 8821, 8823, 8825, 8827, 8829, 8831, 8833, 8835, 8837, 8839, 8841, 8843, 8845, 8847, 8849, 8851, 8853, 8855, 8857, 8859, 8861, 8863, 8865, 8867, 8869, 8871, 8873, 8875, 8877, 8879, 8881, 8883, 8885, 8887, 8889, 8891, 8893, 8895, 8897, 8899, 8901, 8903, 8905, 8907, 8909, 8911, 8913, 8915, 8917, 8919, 8921, 8923, 8925, 8927, 8929, 8931, 8933, 8935, 8937, 8939, 8941, 8943, 8945, 8947, 8949, 8951, 8953, 8955, 8957, 8959, 8961, 8963, 8965, 8967, 8969, 8971, 8973, 8975, 8977, 8979, 8981, 8983, 8985, 8987, 8989, 8991, 8993, 8995, 8997, 8999, 9001, 9003, 9005, 9007, 9009, 9011, 9013, 9015, 9017, 9019, 9021, 9023, 9025, 9027, 9029, 9031, 9033, 9035, 9037, 9039, 9041, 9043, 9045, 9047, 9049, 9051, 9053, 9055, 9057, 9059, 9061, 9063, 9065, 9067, 9069, 9071, 9073, 9075, 9077, 9079, 9081, 9083, 9085, 9087, 9089, 9091, 9093, 9095, 9097, 9099, 9101, 9103, 9105, 9107, 9109, 9111, 9113, 9115, 9117, 9119, 9121, 9123, 9125, 9127, 9129, 9131, 9133, 9135, 9137, 9139, 9141, 9143, 9145, 9147, 9149, 9151, 9153, 9155, 9157, 9159, 9161, 9163, 9165, 9167, 9169, 9171, 9173, 9175, 9177, 9179, 9181, 9183, 9185, 9189, 9191, 9193, 9195, 9197, 9199, 9201, 9203, 9205, 9207, 9209, 9211, 9213, 9215, 9217, 9219, 9221, 9223,

9225, 9227, 9229, 9231, 9233, 9235, 9237, 9239, 9241, 9243, 9245, 9247, 9249, 9251, 9253, 9255, 9257, 9259, 9261, 9263, 9265, 9267, 9269, 9271, 9273, 9275, 9277, 9279, 9281, 9283, 9285, 9287, 9289, 9291, 9293, 9295, 9297, 9299, 9301, 9303, 9305, 9307, 9309, 9311, 9313, 9315, 9317, 9319, 9321, 9323, 9325, 9327, 9329, 9331, 9333, 9335, 9337, 9339, 9341, 9343, 9345, 9347, 9349, 9351, 9353, 9355, 9357, 9359, 9361, 9363, 9365, 9367, 9369, 9371, 9373, 9375, 9377, 9379, 9381, 9383, 9385, 9387, 9389, 9391, 9393, 9395, 9397, 9399, 9401, 9403, 9405, 9407, 9409, 9411, 9413, 9415, 9417, 9419, 9421, 9423, 9425, 9427, 9429, 9431, 9433, 9435, 9437, 9439, 9441, 9443, 9445, 9447, 9449, 9451, 9453, 9455, 9457, 9459, 9461, 9463, 9465, 9467, 9469, 9471, 9473, 9475, 9477, 9479, 9481, 9483, 9485, 9487, 9489, 9491, 9493, 9495, 9497, 9499, 9501, 9503, 9505, 9507, 9509, 9511, 9513, 9515, 9517, 9519, 9521, 9523, 9525, 9527, 9529, 9531, 9533, 9535, 9537, 9539, 9541, 9543, 9545, 9547, 9549, 9551, 9553, 9555, 9557, 9559, 9561, 9563, 9565, 9567, 9569, 9571, 9573, 9575, 9577, 9579, 9581, 9583, 9585, 9587, 9589, 9591, 9593, 9595, 9597, 9599, 9601, 9603, 9605, 9607, 9609, 9611, 9613, 9615, 9617, 9619, 9621, 9623, 9625, 9627, 9629, 9631, 9633, 9635, 9637, 9639, 9641, 9643, 9645, 9647, 9649, 9651, 9653, 9655, 9657, 9659, 9661, 9663, 9665, 9667, 9669, 9671, 9673, 9675, 9677, 9679, 9681, 9683, 9685, 9687, 9689, 9691, 9693, 9695, 9697, 9699, 9701, 9703, 9705, 9707, 9709, 9711, 9713, 9715, 9717, 9719, 9721, 9723, 9725, 9727, 9729, 9731, 9733, 9735, 9737, 9739, 9741, 9743, 9745, 9747, 9749, 9751, 9753, 9755, 9757, 9759, 9761, 9763, 9765, 9767, 9769, 9771, 9773, 9775, 9777, 9779, 9781, 9783, 9785, 9787, 9789, 9791, 9793, 9795, 9797, 9799, 9801, 9803, 9805, 9807, 9809, 9811, 9813, 9815, 9817, 9819, 9821, 9823, 9825, 9827, 9829, 9831, 9833, 9835, 9837, 9839, 9841, 9843, 9845, 9847, 9849, 9851, 9853, 9855, 9857, 9859, 9861, 9863, 9865, 9867, 9869, 9871, 9873, 9875, 9877, 9879, 9881, 9883, 9885, 9887, 9889, 9891, 9893, 9895, 9897, 9899, 9901, 9903, 9905, 9907, 9909, 9911, 9913, 9915, 9917, 9919, 9921, 9923, 9925, 9927, 9929, 9931, 9933, 9935, 9937, 9939, 9941, 9943, 9945, 9947, 9949, 9951, 9953, 9955, 9957, 9959, 9961, 9963, 9965, 9967, 9969, 9971, 9973, 9975, 9977, 9979, 9981, 9983, 9985, 9987, 9989, 9991, 9993, 9995, 9997, 9999, 10001, 10003, 10005, 10007, 10009, 10011, 10013, 10015, 10017, 10019, 10021, 10023, 10025, 10027, 10029, 10031, 10033, 10035, 10037, 10039, 10041, 10043, 10045, 10047, 10049, 10051, 10053, 10055, 10057, 10059, 10061, 10063, 10065, 10067, 10069, 10071, 10073, 10075, 10077, 10079, 10081, 10083, 10085, 10087, 10089, 10091, 10093, 10095, 10097, 10099, 10101, 10103, 10105, 10107, 10109, 10111, 10113, 10115, 10117, 10119, 10121, 10123, 10125, 10127, 10129, 10131, 10133, 10135, 10137, 10139, 10141, 10143, 10145, 10147, 10149, 10151, 10153, 10155, 10157, 10159, 10161, 10163, 10165, 10167, 10169, 10171, 10173, 10175, 10177, 10179, 10181, 10183, 10185, 10187, 10189, 10191, 10193, 10195, 10197, 10199, 10201, 10203, 10205, 10207, 10209, 10211, 10213, 10215, 10217, 10219, 10221, 10223, 10225, 10227, 10229, 10231, 10233, 10235, 10237, 10239, 10241, 10243, 10245, 10247, 10249, 10251, 10253, 10255, 10257, 10259, 10261, 10263, 10265, 10267, 10269, 10271, 10273, 10275, 10277, 10279, 10281, 10283, 10285, 10287, 10289, 10291, 10293, 10295, 10297, 10299, 10301, 10303, 10305, 10307, 10309, 10311, 10313, 10315, 10317, 10319, 10321, 10323, 10325, 10327, 10329, 10331, 10333, 10335, 10337, 10339, 10341, 10343, 10345, 10347, 10349, 10351, 10353, 10355, 10357, 10359, 10361, 10363, 10365, 10367, 10369, 10371, 10373, 10375, 10377, 10379, 10381, 10383, 10385, 10387, 10389, 10391, 10393, 10395, 10397, 10399, 10401, 10403, 10405, 10407, 10409, 10411, 10413, 10415, 10417, 10419, 10421, 10423, 10425, 10427, 10429, 10431, 10433, 10435, 10437, 10439, 10441, 10443, 10445, 10447, 10449, 10451, 10453, 10455, 10457, 10459, 10461, 10463, 10465, 10467, 10469, 10471, 10473, 10475, 10477, 10479, 10481, 10483, 10485, 10487, 10489, 10491, 10493, 10495, 10497, 10499, 10501, 10503, 10505, 10507, 10509, 10511, 10513, 10515, 10517, 10519, 10521, 10523, 10525, 10527, 10529, 10531, 10533, 10535, 10537, 10539, 10541, 10543, 10545, 10547, 10549, 10551, 10553, 10555, 10557, 10559, 10561, 10563, 10565, 10567, 10569, 10571, 10573, 10575, 10577, 10579, 10581, 10583, 10585, 10587, 10589, 10591, 10593, 10595, 10597, 10599, 10601, 10603, 10605, 10607, 10609, 10611, 10613, 10615, 10617, 10619, 10621, 10623, 10625,

10627, 10629, 10631, 10633, 10635, 10637, 10639, 10641, 10643, 10645, 10647, 10649, 10651, 10653, 10655, 10657, 10659, 10661, 10663, 10665, 10667, 10669, 10671, 10673, 10675, 10677, 10679, 10681, 10683, 10685, 10687, 10689, 10691, 10693, 10695, 10697, 10699, 10701, 10703, 10705, 10707, 10709, 10711, 10713, 10715, 10717, 10719, 10721, 10723, 10725, 10727, 10729, 10731, 10733, 10735, 10737, 10739, 10741, 10743, 10745, 10747, 10749, 10751, 10753, 10755, 10757, 10759, 10761, 10763, 10765, 10767, 10769, 10771, 10773, 10775, 10777, 10779, 10781, 10783, 10785, 10787, 10789, 10791, 10793, 10795, 10797, 10799, 10801, 10803, 10805, 10807, 10809, 10811, 10813, 10815, 10817, 10819, 10821, 10823, 10825, 10827, 10829, 10831, 10833, 10835, 10837, 10839, 10841, 10843, 10845, 10847, 10849, 10851, 10853, 10855, 10857, 10859, 10861, 10863, 10865, 10867, 10869, 10871, 10873, 10875, 10877, 10879, 10881, 10883, 10885, 10887, 10889, 10891, 10893, 10895, 10897, 10899, 10901, 10903, 10905, 10907, 10909, 10911, 10913, 10915, 10917, 10919, 10921, 10923, 10925, 10927, 10929, 10931, 10933, 10935, 10937, 10939, 10941, 10943, 10945, 10947, 10949, 10951, 10953, 10955, 10957, 10959, 10961, 10963, 10965, 10967, 10968, 10969, 10970, 10971, 10972, 10973, 10974, 10975, 10976, 10977, 10978, 10979, 10980, 10981, 10982, 10983, 10984, 10985, 10986, 10987, 10988, 10989, 10990, 10991, 10992, 10993, 10994, 10995, 10996, 10997, 10998, 10999, 11000, 11001, 11002, 11003, 11004, 11005, 11006, 11007, 11008, 11009, 11010, 11011, 11012, 11013, 11014, 11015, 11016, 11017, 11018, 11019, 11020, 11021, 11022, 11023, 11024, 11025, 11026, 11027, 11028, 11029, 11030, 11031, 11032, 11033, 11034, 11035, 11036, 11037, 11038, 11039, 11040, 11041, 11042, 11043, 11044, 11045, 11046, 11047, 11048, 11049, 11050, 11051, 11052, 11053, 11054, 11055, 11056, 11057, 11058, 11059, 11060, 11061, 11062, 11063, 11064, 11065, 11066, 11067, 11068, 11069, 11070, 11071, 11072, 11073, 11074, 11075, 11076, 11077, 11078, 11079, 11080, 11081, 11082, 11083, 11084, 11085, 11086, 11087, 11088, 11089, 11090, 11091, 11092, 11093, 11094, 11095, 11096, 11097, 11098, 11099, 11100, 11101, 11102, 11103, 11104, 11105, 11106, 11107, 11108, 11109, 11110, 11111, 11112, 11113, 11114, 11115, 11116, 11117, 11118, 11119, 11120, 11121, 11122, 11123, 11124, 11125, 11126, 11127, 11128, 11129, 11130, 11131, 11132, 11133, 11134, 11135, 11136, 11137, 11138, 11139, 11140, 11141, 11142, 11143, 11144, 11145, 11146, 11147, 11148, 11149, 11150, 11151, 11152, 11153, 11154, 11155, 11156, 11157, 11158, 11159, 11160, 11161, 11162, 11163, 11164, 11165, 11166, 11167, 11168, 11169, 11170, 11171, 11172, 11173, 11174, 11175, 11176, 11177, 11178, 11179, 11180, 11181, 11182, 11183, 11184, 11185, 11186, 11187, 11188, 11189, 11190, 11191, 11192, 11193, 11194, 11195, 11196, 11197, 11198, 11199, 11200, 11201, 11202, 11203, 11204, 11205, 11206, 11207, 11208, 11209, 11210, 11211, 11212, 11213, 11214, 11215, 11216, 11217, 11218, 11219, 11220, 11221, 11222, 11223, 11224, 11225, 11226, 11227, 11228, 11229, 11230, 11231, 11232, 11233, 11234, 11235, 11236, 11237, 11238, 11239, 11240, 11241, 11242, 11243, 11244, 11245, 11246, 11247, 11248, 11249, 11250, 11251, 11252, 11253, 11254, 11255, 11256, 11257, 11258, 11259, 11260, 11261, 11262, 11263, 11264, 11265, 11266, 11267, 11268, 11269, 11270, 11271, 11272, 11273, 11274, 11275, 11276, 11277, 11278, 11279, 11280, 11281, 11282, 11283, 11284, 11285, 11286, 11287, 11288, 11289, 11290, 11291, 11292, 11293, 11294, 11295, 11296, 11297, 11298, 11299, 11300, 11301, 11302, 11303, 11304, 11305, 11306, 11307, 11308, 11309, 11310, 11311, 11312, 11313, 11314, 11315, 11316, 11317, 11318, 11319, 11320, 11321, 11322, 11323, 11324, 11325, 11326, 11327, 11328, 11329, 11330, 11331, 11332, 11333, 11334, 11335, 11336, 11337, 11338, 11339, 11340, 11341, 11342, 11343, 11344, 11345, 11346, 11347, 11348, 11349, 11350, 11351, 11352, 11353, 11354, 11355, 11356, 11357, 11358, 11359, 11360, 11361, 11362, 11363, 11364, 11365, 11366, 11367, 11368, 11369, 11370, 11371, 11372, 11373, 11374, 11375, 11376, 11377, 11378, 11379, 11380, 11381, 11382, 11383, 11384, 11385, 11386, 11387, 11388, 11389, 11390, 11391, 11392, 11393, 11394, 11395, 11396, 11397, 11398, 11399, 11400, 11401, 11402, 11403, 11404, 11405, 11406, 11407, 11408, 11409, 11410, 11411, 11412, 11413, 11414, 11415, 11416, 11417, 11418, 11419, 11420, 11421, 11422, 11423, 11424, 11425, 11426, 11427, 11428, 11429, 11430, 11431, 11432, 11433, 11434, 11435, 11436,

11437, 11438, 11439, 11440, 11441, 11442, 11443, 11444, 11445, 11446, 11447, 11448, 11449, 11450, 11451, 11452, 11453, 11454, 11455, 11456, 11457, 11458, 11459, 11460, 11461, 11462, 11463, 11464, 11465, 11466, 11467, 11468, 11469, 11470, 11471, 11472, 11473, 11474, 11475, 11476, 11477, 11478, 11479, 11480, 11481, 11482, 11483, 11484, 11485, 11486, 11487, 11488, 11489, 11490, 11491, 11492, 11493, 11494, 11495, 11496, 11497, 11498, 11499, 11500, 11501, 11502, 11503, 11504, 11505, 11506, 11507, 11508, 11509, 11510, 11511, 11512, 11513, 11514, 11515, 11516, 11517, 11518, 11519, 11520, 11521, 11522, 11523, 11524, 11525, 11526, 11527, 11528, 11529, 11530, 11531, 11532, 11533, 11534, 11535, 11536, 11537, 11538, 11539, 11540, 11541, 11542, 11543, 11544, 11545, 11546, 11547, 11548, 11549, 11550, 11551, 11552, 11553, 11554, 11555, 11556, 11557, 11558, 11559, 11560, 11561, 11562, 11563, 11564, 11565, 11566, 11567, 11568, 11569, 11570, 11571, 11572, 11573, 11574, 11575, 11576, 11577, 11578, 11579, 11580, 11581, 11582, 11583, 11584, 11585, 11586, 11587, 11588, 11589, 11590, 11591, 11592, 11593, 11594, 11595, 11596, 11597, 11598, 11599, 11600, 11601, 11602, 11603, 11604, 11605, 11606, 11607, 11608, 11609, 11610, 11611, 11612, 11613, 11614, 11615, 11616, 11617, 11618, 11619, 11620, 11621, 11622, 11623, 11624, 11625, 11626, 11627, 11628, 11629, 11630, 11631, 11632, 11633, 11634, 11635, 11636, 11637, 11638, 11639, 11640, 11641, 11642, 11643, 11644, 11645, 11646, 11647, 11648, 11649, 11650, 11651, 11652, 11653, 11654, 11655, 11656, 11657, 11658, 11659, 11660, 11661, 11662, 11663, 11664, 11665, 11666, 11667, 11668, 11669, 11670, 11671, 11672, 11673, 11674, 11675, 11676, 11677, 11678, 11679, 11680, 11681, 11682, 11683, 11684, 11685, 11686, 11687, 11688, 11689, 11690, 11691, 11692, 11693, 11694, 11695, 11696, 11697, 11698, 11699, 11700, 11701, 11702, 11703, 11704, 11705, 11706, 11707, 11708, 11709, 11710, 11711, 11712, 11713, 11714, 11715, 11716, 11717, 11718, 11719, 11720, 11721, 11722, 11723, 11724, 11725, 11726, 11727, 11728, 11729, 11730, 11731, 11732, 11733, 11734, 11735, 11736, 11737, 11738, 11739, 11740, 11741, 11742, 11743, 11744, 11745, 11746, 11747, 11748, 11749, 11750, 11751, 11752, 11753, 11754, 11755, 11756, 11757, 11758, 11759, 11760, 11761, 11762, 11763, 11764, 11765, 11766, 11767, 11768, 11769, 11770, 11771, 11772, 11773, 11774, 11775, 11776, 11777, 11778, 11779, 11780, 11781, 11782, 11783, 11784, 11785, 11786, 11787, 11788, 11789, 11790, 11791, 11792, 11793, 11794, 11795, 11796, 11797, 11798, 11799, 11800, 11801, 11802, 11803, 11804, 11805, 11806, 11807, 11808, 11809, 11810, 11811, 11812, 11813, 11814, 11815, 11816, 11817, 11818, 11819, 11820, 11821, 11822, 11823, 11824, 11825, 11826, 11827, 11828, 11829, 11830, 11831, 11832, 11833, 11834, 11835, 11836, 11837, 11838, 11839, 11840, 11841, 11842, 11843, 11844, 11845, 11846, 11847, 11848, 11849, 11850, 11851, 11852, 11853, 11854, 11855, 11856, 11857, 11858, 11859, 11860, 11861, 11862, 11863, 11864, 11865, 11866, 11867, 11868, 11869, 11870, 11871, 11872, 11873, 11874, 11875, 11876, 11877, 11878, 11879, 11880, 11881, 11882, 11883, 11884, 11885, 11886, 11887, 11888, 11889, 11890, 11891, 11892, 11893, 11894, 11895, 11896, 11897, 11898, 11899, 11900, 11901, 11902, 11903, 11904, 11905, 11906, 11907, 11908, 11909, 11910, 11911, 11912, 11913, 11914, 11915, 11916, 11917, 11918, 11919, 11920, 11921, 11922, 11923, 11924, 11925, 11926, 11927, 11928, 11929, 11930, 11931, 11932, 11933, 11934, 11935, 11936, 11937, 11938, 11939, 11940, 11941, 11942, 11943, 11944, 11945, 11946, 11947, 11948, 11949, 11950, 11951, 11952, 11953, 11954, 11955, 11956, 11957, 11958, 11959, 11960, 11961, 11962, 11963, 11964, 11965, 11966, 11967, 11968, 11969, 11970, 11971, 11972, 11973, 11974, 11975, 11976, 11977, 11978, 11979, 11980, 11981, 11982, 11983, 11984, 11985, 11986, 11987, 11988, 11989, 11990, 11991, 11992, 11993, 11994, 11995, 11996, 11997, 11998, 11999, 12000, 12001, 12002, 12003, 12004, 12005, 12006, 12007, 12008, 12009, 12010, 12011, 12012, 12013, 12014, 12015, 12016, 12017, 12018, 12019, 12020, 12021, 12022, 12023 and 12024.

10. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 6 to 9.

11. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 6 to 10.

12. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any one of claims 6 to 11 under high stringency conditions.

13. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.

14. A composition according to claim 13, being an immunogenic composition, a vaccine composition or a diagnostic composition.

15. A composition according to claim 13 or claim 14 for use as a pharmaceutical.

16. The use of a composition according to claim 13 in the manufacture of a medicament for the treatment or prevention of infection or disease caused by streptococcus bacteria, particularly *S.agalactiae* and *S.pyogenes*.

17. A hybrid protein represented by the formula NH_2-A-[-X-L-]$_n$-B-COOH, wherein X is an amino acid sequence as defined in claim 1, L is an optional linker amino acid sequence, A is an optional N-terminal amino acid sequence, B is an optional C-terminal amino acid sequence, and n is an integer greater than 1.

18. A kit comprising primers for amplifying a template sequence contained within a *Streptococcus* nucleic acid sequence, the kit comprising a first primer and a second primer, wherein the first primer is substantially complementary to said template sequence and the second primer is substantially complementary to a complement of said template sequence, wherein the parts of said primers which have substantial complementarity define the termini of the template sequence to be amplified.

19. A kit comprising first and second single-stranded oligonucleotides which allow amplification of a Streptococcus template nucleic acid sequence contained in a single- or double-stranded nucleic acid (or mixture thereof), wherein: (a) the first oligonucleotide comprises a primer sequence which is substantially complementary to said template nucleic acid sequence; (b) the second oligonucleotide comprises a primer sequence which is substantially complementary to the complement of said template nucleic acid sequence; (c) the first oligonucleotide and/or the second oligonucleotide comprise(s) sequence which is not compementary to said template nucleic acid; and (d) said primer sequences define the termini of the template sequence to be amplified.

20. The kit of claim 19, wherein the non-complementary sequence(s) of (c) comprise a restriction site and/or a promoter sequence.

21. A computer-readable medium containing one or more of SEQ IDs 1 to 12024, excluding SEQ IDs 3745, 3746, 3921, 3922, 7465, 7466, 8779, 8780, 9187 and 9188.

22. A process for detecting Streptococcus in a biological sample, comprising the step of contacting nucleic acid according to any of claims 6 to 12 with the biological sample under hybridising conditions.

23. The process of claim 22, wherein the process involves nucleic acid amplification.

24. A process for determining whether a compound binds to a protein according to claim 1, claim 2 or claim 3, comprising the step of contacting a test compound with a protein according to claim 1, claim 2 or claim 3 and determining whether the test compound binds to said protein.

25. A compound identified by the process of claim 24.

26. A composition comprising a protein according to claim 1, claim 2 or claim 3 and one or more of the following antigens:

- a protein antigen from *Helicobacter pylori*;
- a protein antigen from *N.meningitidis* serogroup B;
- an outer-membrane vesicle (OMV) preparation from *N.meningitidis* serogroup B;
- a saccharide antigen from N.meningitidis serogroup A, C, W135 and/or Y;
- a saccharide antigen from *Streptococcus pneumoniae*;
- an antigen from hepatitis A virus;
- an antigen from hepatitis B virus;
- an antigen from hepatitis C virus;
- an antigen from *Bordetella pertussis*;
- a diphtheria antigen;
- a tetanus antigen;
- a saccharide antigen from *Haemophilus influenzae* B.
- an antigen from N.gonorrhoeae;
- an antigen from *Chlamydia pneumoniae*;
- an antigen from *Chlamydia trachomatis*;
- an antigen from *Porphyromonas gingivalis*;
- polio antigen(s);
- rabies antigen(s);
- measles, mumps and/or rubella antigens;
- influenza antigen(s);
- an antigen from *Moraxella catarrhalis*; and/or
- an antigen from *Staphylococcus aureus*.

27. A composition comprising two or more proteins, wherein each protein is a protein according to claim 1, claim 2 or claim 3.