```
Sbjct: 171 PTYLSSLIKKELGLPFQDYLVRERVKQA-KLLLLTTDLKIYEIAEKVGFEDMNYFTQRFK 229

Query: 238 KYLGQTVKAFKE 249
           +  G T + FK+
Sbjct: 230 QIAGVTPRQFKK 241
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1621> which encodes the amino acid sequence <SEQ ID 1622>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4239(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 193/257 (75%), Positives = 226/257 (87%)

Query: 1   MYRLLIVEDEHLIRKWLRYAIDYQSLNILVVGEAKDGKEGAQLIQEEQPDIVLSDINMPI 60
           MY+L+I+EDEH+IRKWLRYAIDY++L+ILV+GEAKDGKEGA LI+E QPDIVL+DINMPI
Sbjct: 1   MYKLVIIEDEHIIRKWLRYAIDYKALDILVIGEAKDGKEGAVLIKESQPDIVLTDINMPI 60

Query: 61  MTAFDMFEATKGQSYAKIILSGYADFPNAQSAIHYGVLEFLTKPLEKQALIDCLKTIMAR 120
           MTAFDMFE TK Q+YAKIILSGYADFPNA+SAIHYGVLEFLTKP+EK AL +CL+TI+A+
Sbjct: 61  MTAFDMFEVTKDQTYAKIILSGYADFPNARSAIHYGVLEFLTKPIEKAALWECLQTIIAK 120

Query: 121 IEEHKEKHLQEHTELYLPLPQANDQVPEVIKDMLAWIHSHFHGKIVISQLAHDLGYSESY 180
           IE+ K  + +     +Y+PLPQ  DQ+PEV+KD+L W+H+HF  KI  S+LAHDLGYSESY
Sbjct: 121 IEKQKGSNQKTDACVYIPLPQMTDQIPEVVKDILEWVHAHFQDKISTSRLAHDLGYSESY 180

Query: 181 LYTVTKKHLHITLSDYINQYRINQAIQLMFREPDLMVYQIAEAVGIYDYRYFDRVFKKYL 240
           +Y   KKHL + LSDYINQYRINQAIQLM +EPDLMVY+IA+AVGIYDYRYFDRVFKKYL
Sbjct: 181 IYQNIKKHLQMPLSDYINQYRINQAIQLMQQEPDLMVYEIAQAVGIYDYRYFDRVFKKYL 240

Query: 241 GQTVKAFKEEHIFKQMD 257
           GQTVKAFKEEH  K  D
Sbjct: 241 GQTVKAFKEEHFMKDTD 257
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 507

A DNA sequence (GBSx0545) was identified in *S.agalactiae* <SEQ ID 1623> which encodes the amino acid sequence <SEQ ID 1624>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2964(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 508

A DNA sequence (GBSx0546) was identified in *S.agalactiae* <SEQ ID 1625> which encodes the amino acid sequence <SEQ ID 1626>. This protein is predicted to be two-component sensor histidine kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.80   Transmembrane  266 - 282 ( 257 - 285)
   INTEGRAL    Likelihood =-12.90   Transmembrane   29 -  45 (  24 -  51)

----- Final Results -----
               bacterial membrane --- Certainty=0.6519(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10197> which encodes amino acid sequence <SEQ ID 10198> was also identified.

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

```
>GP:BAB05628 GB:AP001513 two-component sensor histidine kinase
          [Bacillus halodurans]
 Identities = 84/258 (32%), Positives = 138/258 (52%), Gaps = 23/258 (8%)

Query: 298 SSAINQMVLDMDAISRQEKSSIELDSQDEFQYLSVQINQMVSRLKDLHEKTLDLETQKLL 357
           S  INQ+        S    K+ I +D +DE    LSVQ NQMV+ L+ L  +  +    QK L
Sbjct: 327 SERINQVA------SGDLKTKIVVDGKDEIGQLSVQFNQMVANLRSLIHQVHETNRQKRL 380

Query: 358 FEK-------RMLEAQFNPHFLYNTLETILITSHYDSQL-TERIVIQLTKLLRYSLSGST 409
            EK        +ML +Q NPHFL+NTLE+I + SH   +     ++V QL KL+R SL  +
Sbjct: 381 LEKSQNEIKLKMLASQINPHFLFNTLESIRMKSHMKGETEIAKVVKQLGKLMRKSLEVTG 440

Query: 410 EAAVLKDDLAIIESYLLINQVRF-EELTYTISVSPELEHMRVPKLFLLPLIENAIKYGLK 468
               L+++L ++   YL I    R+ + L Y + + P+ E + +  L + PL+ENA+ +GL+
Sbjct: 441 HHIPLRNELDMVRCYLEIQTFRYGDRLHYELYIDPQSEMVEILPLIIQPLVENAVIHGLE 500

Query: 469 ERHD-VAINIDIWQDSDGIWFTVSNNGSGISLARQQAIRTMLRSTH----SHHGLINSYR 523
             D   + I     + + +   V+++G G+    + +AI+ ML          +  GL+N ++
Sbjct: 501 RTEDGGTVTISTIVNGNDLTVIVNDDGCGMDEEKLEAIQNMLHHPQEVDGNKIGLLNVHK 560

Query: 524 RLQYQF---STVLLEFTK 538
           RLQ  +   S +++E  K
Sbjct: 561 RLQLTYGKTSGLIIESAK 578
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1627> which encodes the amino acid sequence <SEQ ID 1628>. Analysis of this protein sequence reveals the following:

```
     Possible site: 43

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.88   Transmembrane   27 -  43 (  22 -  49)
   INTEGRAL    Likelihood = -9.08   Transmembrane  263 - 279 ( 258 - 282)

----- Final Results -----
               bacterial membrane --- Certainty=0.5352(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:

```
>GP:BAB05628 GB:AP001513 two-component sensor histidine kinase
          [Bacillus halodurans]
 Identities = 85/270 (31%), Positives = 139/270 (51%), Gaps = 20/270 (7%)
```

```
Query: 276 IFVILQRKSSGLANRIAAKNSRAINQMVRDMSAISRQEKRRIDLESQDEFQYLSDQINQM 335
           + V+L    S L ++  +  S  INQ+       S   K +I ++ +DE   LS Q NQM
Sbjct: 307 VAVLLIVHFSWLISKRLSHLSERINQVA------SGDLKTKIVVDGKDEIGQLSVQFNQM 360

Query: 336 VERLQQLHDKTLDLETQKLLFEK-------RMLEAQFNPHFLYNTLETILITSHYDSAL- 387
           V  L+ L  +  +   QK L EK        +ML +Q NPHFL+NTLE+I + SH
Sbjct: 361 VANLRSLIHQVHETNRQKRLLEKSQNEIKLKMLASQINPHFLFNTLESIRMKSHMKGETE 420

Query: 388 TEKIVIQLTKLLRYSLTDSSKPVLLKDDLSVIESYLVINQVRF-EELQYSINLSPDLDSL 446
             K+V QL KL+R SL  +   + L+++L ++  YL I   R+ + L Y + + P  + +
Sbjct: 421 IAKVVKQLGKLMRKSLEVTGHHIPLRNELDMVRCYLEIQTFRYGDRLHYELYIDPQSEMV 480

Query: 447 EVPKLFLLPLIENAIKYGLKERHD-VKINIACYYQDDHIIFSVRDNGSGIDAHHQKVIRE 505
           E+  L + PL+ENA +GL+   D   + I+     + +   V D+G G+D     + I+
Sbjct: 481 EILPLIIQPLVENAVIHGLERTEDGGTVTISTIVNGNDLTVIVNDDGCGMDEEKLEAIQN 540

Query: 506 QL----EAGESHHGLINSYRRLKYHFSEVS 531
            L     E   + GL+N ++RL+  + + S
Sbjct: 541 MLHHPQEVDGNKIGLLNVHKRLQLTYGKTS 570
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 369/549 (67%), Positives = 449/549 (81%)

Query: 3   MRGYRMEERFKKRLQDDISKHFSRQSLILSLLLIALFVLFSLAPQQIGLYKDVNSVSYSY 62
           MRG ++EE FKK+LQDDIS+HFS QSL+LSLLLI LF++FSLAPQQ+GLY+D+N+ +  Y
Sbjct: 1   MRGEQVEEHFKKQLQDDISRHFSYQSLMLSLLLIGLFIIFSLAPQQLGLYRDINATATRY 60

Query: 63  KQLIQKHDTLLDDLGKNSLKPFVSGHLGSADLSKQYYHLRNHLQSQTELLVFSPNQELLF 122
            +LI K + LDDDLGKNSL PF++ +L +ADLSK Y+HLR+  Q+  ELL+FSP+Q+LLF
Sbjct: 61  HRLISKQEALLDDDLGKNSLLPFLNKNLSTADLSKHYFHLRHSSQTSPELLLFSPSQDLLF 120

Query: 123 ASNSHLGNFFSKSIYISEVLDKAKINQRLLKIIVDSEGGHYLALIKPIIVNKKVSGYAFL 182
           ASN HLGN FSKS+YI EVL      + L K  +DSE GHYL +I P+I    ++ GYAFL
Sbjct: 121 ASNPHLGNVFSKSVYIQEVLRATHSPKTLFKDAMDSEDGHYLMIIMPMIDQNQLKGYAFL 180

Query: 183 LMNGKDFLLPTKAINSDLIIADQLNNSFTFTNRDFISSSLDKVDSQFLTRYFSFHDHRAF 242
           +M+GKDFL PTK + S+L+IAD+L+N+FTF+NR+FI+SSLDK++SQ+L  YF F D+RAF
Sbjct: 181 VMSGKDFLHPTKTLTSELVIADKLDNTFTFSNREFIASSLDKINSQYLHHYFVFQDNRAF 240

Query: 243 VVRKVALQDNILLYMYRPLIPVTLVVLFSLVSSVIIFVILRQKSRVLADRIAVKNSSAIN 302
           + RKVALQ  + LYMYRPLIP+  V+LFSL+SS +IFVIL++KS  LA+RIA KNS AIN
Sbjct: 241 ITRKVALQGGLWLYMYRPLIPMVSVMLFSLISSAVIFVILQRKSSGLANRIAAKNSRAIN 300

Query: 303 QMVLDMDAISRQEKSSIELDSQDEFQYLSVQINQMVSRLKDLHEKTLDLETQKLLFEKRM 362
           QMV DM AISRQEK  I+L+SQDEFQYLS QINQMV RL+ LH+KTLDLETQKLLFEKRM
Sbjct: 301 QMVRDMSAISRQEKRRIDLESQDEFQYLSDQINQMVERLQQLHDKTLDLETQKLLFEKRM 360

Query: 363 LEAQFNPHFLYNTLETILITSHYDSQLTERIVIQLTKLLRYSLSGSTEAAVLKDDLAIIE 422
           LEAQFNPHFLYNTLETILITSHYDS LTE+IVIQLTKLLRYSL+ S++  +LKDDL++IE
Sbjct: 361 LEAQFNPHFLYNTLETILITSHYDSALTEKIVIQLTKLLRYSLTDSSKPVLLKDDLSVIE 420

Query: 423 SYLLINQVRFEELTYTISVSPELEHMRVPKLFLLPLIENAIKYGLKERHDVAINIDIWQD 482
           SYL+INQVRFEEL Y+I++SP+L+ + VPKLFLLPLIENAIKYGLKERHDV INI  +
Sbjct: 421 SYLVINQVRFEELQYSINLSPDLDSLEVPKLFLLPLIENAIKYGLKERHDVKINIACYYQ 480

Query: 483 SDGIWFTVSNNGSGISLARQQAIRTMLRSTHSHHGLINSYRRLQYQFSTVLLEFTKTDDA 542
            D I F+V +NGSGI     Q+ IR  L +  SHHGLINSYRRL+Y FS V L F + D
Sbjct: 481 DDHIIFSVRDNGSGIDAHHQKVIREQLEAGESHHGLINSYRRLKYHFSEVSLVFDQGDKQ 540

Query: 543 FRVSYIVKE 551
           F VSY VKE
Sbjct: 541 FNVSYHVKE 549
```

A related GBS gene <SEQ ID 8587> and protein <SEQ ID 8588> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:       6.23
GvH: Signal Score (-7.5): -0.0500002
     Possible site: 38
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 1 value: -13.80 threshold:  0.0
   INTEGRAL    Likelihood =-13.80   Transmembrane  259 - 275 ( 250 - 278)
   PERIPHERAL  Likelihood =  2.70     404
 modified ALOM score:   3.26

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6519(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:

```
33.2/53.9% over 181aa
                              Streptococcus pneumoniae
  GP|5830535| histidine kinase Insert characterized

ORF00032(1309 - 1848 of 2253)
GP|5830535|emb|CAB54576.1||AJ006396(1  -  182  of  231)  histidine  kinase  (Streptococcus
pneumoniae)
%Match = 5.9
%Identity = 33.2  %Similarity = 53.8
Matches = 61  Mismatches = 78  Conservative Sub.s = 38

1053       1083       1113       1143       1173       1203       1233       1263
FVVRKVALQDNILLYMYRPLIPVTLVVLFSLVSSVIIFVILRQKSRVLADRIAVKNSSAINQMVLDMDAISRQEKSSIEL

1293       1323       1350       1380       1410       1440                  1494
DSQDEFQYLSVQINQMVSRL-KDLHEKTLDLETQKLLFEKRMLEAQFNPHFLYNTLETILITSHYDSQ--LTERIVIQLT
                 |: || |::|:    ||  :     | |:|| ||||:||||| : : :   ||  | : |: ::: .
                 MLDRLEKNIHD-IYQLELSQKDANMRALQAQINPHFMYNTLEFLRMYAVMQSQDELAD-IIYEFS
                          10         20         30         40         50         60

1524       1554       1584       1611       1641       1671       1701       1728
KLLRYSLSGSTEAAVLKDDLAIIESYLLINQVRF-EELTYTISVSPELEHMRVPKLFLLPLIENAIKYGLKERH-DVAIN
 ||| ::|      :|| :|     |  :  ||: : : |   : ||||:|::||: | ||:||   :|:  |  |  |:
SLLRNNIS-DERETLLKQELEFCRKYSYLCMVRYPKSIAYGFKIDPELENMKIPKFTLQPLVENYFAHGVDHRRTDNVIS
                 80         90        100        110        120        130        140

1758       1788       1818       1848       1878       1908       1938       1968
IDIWQDSDGIWFTVSNNGSGISLARQQAIRTMLRSTHSHHGLINSYRRLQYQFSTVLLEFTKTDDAFRVSYIVKE*VMYR
|   :    :   | :|| |:|  :   ||  |   :  |    | :|       |   |           |
IKALKQDGFVEILVVDNGRGMSAEKLANIREKLSQRYFEHQASYSDQRQSIGIVNVHERFVLYFGDRYAITIESAEQAGV
                160        170        180        190        200        210        220
```

SEQ ID 8588 (GBS47) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 2; MW 84kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 4; MW 59.3kDa).

GBS47-His was purified as shown in Figure 221, lane 4-5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 509

A DNA sequence (GBSx0547) was identified in *S.agalactiae* <SEQ ID 1629> which encodes the amino acid sequence <SEQ ID 1630>. This protein is predicted to be phosphotransferase enzyme II, D component. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.46   Transmembrane  258 - 274 ( 252 - 274)
   INTEGRAL    Likelihood = -9.13   Transmembrane  232 - 248 ( 227 - 251)
   INTEGRAL    Likelihood = -5.31   Transmembrane  142 - 158 ( 140 - 161)
   INTEGRAL    Likelihood = -2.50   Transmembrane  119 - 135 ( 118 - 139)

----- Final Results -----
              bacterial membrane --- Certainty=0.5182(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 GENPEPT database:

```
>GP:AAC74889 GB:AE000276 PTS enzyme IID, mannose-specific
          [Escherichia coli K12]
 Identities = 94/280 (33%), Positives = 156/280 (55%), Gaps = 13/280 (4%)

Query: 3   SQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAVGVIYTLLPVINRFYKTDKD-KAEA 61
           ++  LT+ D   +R VF RS     S     +  A+G  ++++P I R Y  + + + +A
Sbjct: 12  TEKKLTQSD---IRGVFLRSNLFQGS-WNFERMQALGFCFSMVPAIRRLYPENNEARKQA 67

Query: 62  LVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFFW 121
           + RH  +FN    +   I+G+  ++E++ +   + D  AI  +K  LMGP++GVGD  FW
Sbjct: 68  IRRHLEFFNTQPFVAAPILGVTLALEEQRANGAEIDDGAINGIKVGLMGPLAGVGDPIFW 127

Query: 122 GILRVIAAGIGISLASTGSAMGAVVFLLLYNIPAFLIHYYSLYGGYSVGAGFIKKLYESG 181
           G +R + A +G  +A +GS +G ++F +L+N+      YY +  GYS G   +K   G
Sbjct: 128 GTVRPVFAALGAGIAMSGSLLGPLLFFILFNLVRLATRYYGVAYGYSKGIDIVKDM-GGG 186

Query: 182 GIKIVTKTSSMLGLMMVGSM----TASNVKFKTILTVAAKGAKEAASIQSYLDQLFVGVV 237
            ++ +T+ +S+LGL ++G++     T  N+            G +   ++Q+ LDQL  G+V
Sbjct: 187 FLQKLTEGASILGLFVMGALVNKWTHVNIPLVVSRITDQTGKEHVTTVQTILDQLMPGLV 246

Query: 238 PLLVTILAFWLLRKKVNINWIMFGIMVLGI---VLGLLGI 274
           PLL+T    WLLRKKVN  WI+ G  V+GI     GLLG+
Sbjct: 247 PLLLTFACMWLLRKKVNPLWIIVGFFVIGIAGYACGLLGL 286
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1631> which encodes the amino acid sequence <SEQ ID 1632>. Analysis of this protein sequence reveals the following:

```
     Possible site: 32
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.98   Transmembrane  255 - 271 ( 251 - 274)
   INTEGRAL    Likelihood = -7.01   Transmembrane  232 - 248 ( 228 - 250)
   INTEGRAL    Likelihood = -5.68   Transmembrane  142 - 158 ( 140 - 161)
   INTEGRAL    Likelihood = -2.50   Transmembrane  119 - 135 ( 118 - 139)

----- 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>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC74889 GB:AE000276 PTS enzyme IID, mannose-specific
          [Escherichia coli]
 Identities = 94/281 (33%), Positives = 157/281 (55%), Gaps = 13/281 (4%)

Query: 2   TSQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAVGVIYTLLPVINRFYKTDKD-KAE 60
```

```
           T++  LT+ D   +R VF RS     S     +  A+G  ++++P I R Y  + + + +
Sbjct: 11  TTEKKLTQSD---IRGVFLRSNLFQGS-WNFERMQALGFCFSMVPAIRRLYPENNEARKQ 66

Query: 61  ALVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFF 120
           A+ RH  +FN    +   I+G+  ++E++ +   + D  AI  +K  LMGP++GVGD  F
Sbjct: 67  AIRRHLEFFNTQPFVAAPILGVTLALEEQRANGAEIDDGAINGIKVGLMGPLAGVGDPIF 126

Query: 121 WGILRVIAAGIGISLASAGSAMGAVVFLLLYNIPAFIIHYYSLYGGYSVGAGFIKKLYES 180
           WG +R + A +G   +A +GS +G ++F +L+N+      YY +  GYS G   +K +
Sbjct: 127 WGTVRPVFAALGAGIAMSGSLLGPLLFFILFNLVRLATRYYGVAYGYSKGIDIVKDM-GG 185

Query: 181 GGIKIVTKTSSMLGLMMVGSM----TASNVKFKTILTVAAKGAKEAASIQDYLDQLFIGI 236
           G ++ +T+ +S+LGL ++G++    T  N+          G +   ++Q  LDQL  G+
Sbjct: 186 GFLQKLTEGASILGLFVMGALVNKWTHVNIPLVVSRITDQTGKEHVITVQTILDQLMPGL 245

Query: 237 VPLMVTLAAFWLLRKKVNIIWIMFGIMFLGI---ILGLLGI 274
           VPL++T A  WLLRKKVN +WI+ G   +GI     GLLG+
Sbjct: 246 VPLLLTFACMWLLRKKVNPLWIIVGFFVIGIAGYACGLLGL 286
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 263/275 (95%), Positives = 269/275 (97%)

Query: 1   MKSQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAVGVIYTLLPVINRFYKTDKDKAE 60
           M SQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAVGVIYTLLPVINRFYKTDKDKAE
Sbjct: 1   MTSQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAVGVIYTLLPVINRFYKTDKDKAE 60

Query: 61  ALVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFF 120
           ALVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFF
Sbjct: 61  ALVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFF 120

Query: 121 WGILRVIAAGIGISLASTGSAMGAVVFLLLYNIPAFLIHYYSLYGGYSVGAGFIKKLYES 180
           WGILRVIAAGIGISLAS GSAMGAVVFLLLYNIPAF+IHYYSLYGGYSVGAGFIKKLYES
Sbjct: 121 WGILRVIAAGIGISLASAGSAMGAVVFLLLYNIPAFIIHYYSLYGGYSVGAGFIKKLYES 180

Query: 181 GGIKIVTKTSSMLGLMMVGSMTASNVKFKTILTVAAKGAKEAASIQSYLDQLFVGVVPLL 240
           GGIKIVTKTSSMLGLMMVGSMTASNVKFKTILTVAAKGAKEAASIQ YLDQLF+G+VPL+
Sbjct: 181 GGIKIVTKTSSMLGLMMVGSMTASNVKFKTILTVAAKGAKEAASIQDYLDQLFIGIVPLM 240

Query: 241 VTILAFWLLRKKVNINWIMFGIMVLGIVLGLLGIC 275
           VT+ AFWLLRKKVNI WIMFGIM LGI+LGLLGIC
Sbjct: 241 VTLAAFWLLRKKVNIIWIMFGIMFLGIILGLLGIC 275
```

There is also homology to SEQ ID 5236.

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9077> which encodes the amino acid sequence <SEQ ID 9078>. An alignment of the GAS and GBS sequences follows:

```
 Score =  178 bits (448), Expect = 3e-47
 Identities = 83/136 (61%), Positives = 108/136 (79%)

Query: 2   IMEEITIYHNPNCGTSRNVLAMIRHAGIEPTIIEYLQTPPNRETLIELLQSMGISARELL 61
           +ME+I IYHNPNCGTSRNVLA+IRH GIEP II YL+TPP+R  L+ELL  M +SARELL
Sbjct: 1   MMEKIRIYHNPNCGTSRNVLAIIRHCGIEPEIIYYLKTPPSRMELVELLLEMKLSARELL 60

Query: 62  RTNVPEFEAYGLANQAVAEKDIINAMLADPILINRPIVVTRKGVKLCRPSETLLDILPVP 121
           RT+VP +E + L + +V ++++I+AM+ DPILINRPIVVT KG KLCRP E +L ILPV
Sbjct: 61  RTDVPAYEKFNLESSSVTDEEMIDAMIQDPILINRPIVVTSKGAKLCRPCEAILTILPVK 120

Query: 122 LPSPYIKEDGESVNPI 137
           +   ++KEDG+ +  +
Sbjct: 121 MEKDFVKEDGQIIQSL 136
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 510

A DNA sequence (GBSx0548) was identified in *S.agalactiae* <SEQ ID 1633> which encodes the amino acid sequence <SEQ ID 1634>. This protein is predicted to be PTS permease for mannose subunit IIPMan. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.70   Transmembrane  144 - 160 ( 140 - 165)
   INTEGRAL    Likelihood = -8.07   Transmembrane  220 - 236 ( 215 - 239)
   INTEGRAL    Likelihood = -7.27   Transmembrane   95 - 111 (  91 - 116)
   INTEGRAL    Likelihood = -3.77   Transmembrane    2 -  18 (   1 -  18)
   INTEGRAL    Likelihood = -1.44   Transmembrane  180 - 196 ( 179 - 196)
   INTEGRAL    Likelihood = -1.33   Transmembrane   32 -  48 (  30 -  49)
   INTEGRAL    Likelihood = -0.53   Transmembrane  198 - 214 ( 198 - 214)

----- Final Results -----
              bacterial membrane --- Certainty=0.4482(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 GENPEPT database:

```
>GP:AAC44680 GB:U65015 PTS permease for mannose subunit IIPMan
          [Vibrio furnissii]
 Identities = 70/251 (27%), Positives = 132/251 (51%), Gaps = 6/251 (2%)

Query: 2   IMPATMAALAVLICFGGNYLTGQSMMERPLVVGLVTGMLLGDIKVGILMGASLEALFLGN 61
           +  A M  L   +  G +   G +   RP+V+G + G++LGD+  GIL+G +LE +++G
Sbjct: 5   LFQALMLGLLAFLA-GLDLFNGLTHFHRPVVLGPLVGLILGDLHTGILVGGTLELIWMGL 63

Query: 62  VNIGGVIAAEPVTATAMATTFTIISNIDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFAP 121
             + G      +  T + TTF I +N++   A+ +AVP  +     +  L +    + +
Sbjct: 64  APLAGAQPPNVIIGTIVGTTFAITTNVEPNVAVGVAVPFAVAVQMGITLLFSAMSAVMSK 123

Query: 122 MVDKAAAANHQGKLVMLHYGTWII--YYLIIASISFIGILVGSGPVNSFVHHIPQNLMNG 179
             + A  A+ +G   + ++   ++ +Y + A   F+ I +G+    + V  +P+ L++G
Sbjct: 124 CDEYAKNADTRGIERVNYFALAVLGSFYFLCA---FLPIYLGADHAGAMVAALPKALIDG 180

Query: 180 LSAAGGLLPAVGFAMLMKLLWTNKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQR 239
           L  AGG++PA+GFA+LMK++  N    +++LGFV  A+L+LP +A+      + +I   R
Sbjct: 181 LGVAGGIMPAIGFAVLMKIMMKNAYIPYFILGFVAAAWLQLPILAIRCAATAMAIIDFMR 240

Query: 240 DIELDAITRGA 250
             E   +  A
Sbjct: 241 KSEPTPVNASA 251
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1635> which encodes the amino acid sequence <SEQ ID 1636>. Analysis of this protein sequence reveals the following:

```
     Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.70   Transmembrane  144 - 160 ( 140 - 165)
   INTEGRAL    Likelihood = -8.07   Transmembrane  220 - 236 ( 215 - 239)
   INTEGRAL    Likelihood = -7.27   Transmembrane   95 - 111 (  91 - 116)
   INTEGRAL    Likelihood = -4.62   Transmembrane    2 -  18 (   1 -  19)
   INTEGRAL    Likelihood = -1.44   Transmembrane  180 - 196 ( 179 - 196)
   INTEGRAL    Likelihood = -0.96   Transmembrane   32 -  48 (  31 -  49)
   INTEGRAL    Likelihood = -0.53   Transmembrane  198 - 214 ( 198 - 214)

----- Final Results -----
              bacterial membrane --- Certainty=0.4482(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:

```
>GP:AAC44680 GB:U65015 PTS permease for mannose subunit IIPMan
          [Vibrio furnissii]
 Identities = 72/251 (28%), Positives = 132/251 (51%), Gaps = 6/251 (2%)

Query: 2   LVPATMAALAVLICFGGNYLTGQSMMERPLVVGLVTGLLLGDMKVGILMGASLEALFLGN 61
           L  A M  L   +  G +   G +   RP+V+G + GL+LGD+  GIL+G +LE +++G
Sbjct: 5   LFQALMLGLLAFLA-GLDLFNGLTHFHRPVVLGPLVGLILGDLHTGILVGGTLELIWMGL 63

Query: 62  VNIGGVIAAEPVTATAMATTFTIISHIDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFAP 121
             + G      +  T + TTF I ++++   A+ +AVP  +     +  L +    + +
Sbjct: 64  APLAGAQPPNVIIGTIVGTTFAITTNVEPNVAVGVAVPFAVAVQMGITLLFSAMSAVMSK 123

Query: 122 MVDKAAAANHQGKLVMLHYGTWII--YYLIIASISFIGILVGSGPVNAFVEHIPQNLMNG 179
             + A  A+ +G   + ++   ++ +Y + A   F+ I +G+    A V  +P+ L++G
Sbjct: 124 CDEYAKNADTRGIERVNYFALAVLGSFYFLCA---FLPIYLGADHAGAMVAALPKALIDG 180

Query: 180 LSAAGGLLPAVGFAMLMKLLWTNKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQR 239
           L  AGG++PA+GFA+LMK++  N   +++LGFV  A+L+LP +A+      + +I   R
Sbjct: 181 LGVAGGIMPAIGFAVLMKIMMKNAYIPYFILGFVAAAWLQLPILAIRCAATAMAIIDFMR 240

Query: 240 DLELDAITRGA 250
             E  +  A
Sbjct: 241 KSEPTPVNASA 251
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 261/269 (97%), Positives = 268/269 (99%)

Query: 1   MIMPATMAALAVLICFGGNYLTGQSMMERPLVVGLVTGMLLGDIKVGILMGASLEALFLG 60
           M++PATMAALAVLICFGGNYLTGQSMMERPLVVGLVTG+LLGD+KVGILMGASLEALFLG
Sbjct: 1   MLVPATMAALAVLICFGGNYLTGQSMMERPLVVGLVTGLLLGDMKVGILMGASLEALFLG 60

Query: 61  NVNIGGVIAAEPVTATAMATTFTIISNIDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFA 120
           NVNIGGVIAAEPVTATAMATTFTIIS+IDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFA
Sbjct: 61  NVNIGGVIAAEPVTATAMATTFTIISHIDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFA 120

Query: 121 PMVDKAAAANHQGKLVMLHYGTWIIYYLIIASISFIGILVGSGPVNSFVHHIPQNLMNGL 180
           PMVDKAAAANHQGKLVMLHYGTWIIYYLIIASISFIGILVGSGPVN+FV HIPQNLMNGL
Sbjct: 121 PMVDKAAAANHQGKLVMLHYGTWIIYYLIIASISFIGILVGSGPVNAFVEHIPQNLMNGL 180

Query: 181 SAAGGLLPAVGFAMLMKLLWTNKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQRD 240
           SAAGGLLPAVGFAMLMKLLWTNKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQRD
Sbjct: 181 SAAGGLLPAVGFAMLMKLLWTNKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQRD 240

Query: 241 IELDAITRGAISKQTTFDSKESEEEDFFA 269
           +ELDAITRGAISKQTTFDSKESEEEDFFA
Sbjct: 241 LELDAITRGAISKQTTFDSKESEEEDFFA 269
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 511

A DNA sequence (GBSx0549) was identified in *S.agalactiae* <SEQ ID 1637> which encodes the amino acid sequence <SEQ ID 1638>. This protein is predicted to be pts system, sorbose-specific iib component. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1874(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 GENPEPT database:

```
>GP:CAA46858 GB:X66059 EIII-B Sor PTS [Klebsiella pneumoniae]
 Identities = 49/158 (31%), Positives = 94/158 (59%), Gaps = 8/158 (5%)

Query: 2   ITQIRVDDRLIHGQVAVVWTKELNAPLLVVANDEAAKNEITQMTLKMAVPNGMKLLIRSV 61
            IT  R+DDRLIHGQV  VW+K  NA  +++ ND+   +E+ +  L+ A P GMK+ + S+
Sbjct: 3   ITLARIDDRLIHGQVTTVWSKVANAQRIIICNDDVFNDEVRRTLLRQAAPPGMKVNVVSL 62

Query: 62  EESIALFKDPRATDKRIFVIVNSVKDACTIAKNITDLEAVNVANVGRFDKSDPATKVKLT 121
            E+++A++ +P+  D+ +F +  +  D  T+ +    +  +N+  +     +    K +LT
Sbjct: 63  EKAVAVYHNPQYQDETVFYLFTNPHDVLTMVRQGVQIATLNIGGM-----AWRPGKKQLT 117

Query: 122 SSLLLNTEELEAAKELASL-PDLDVFNQVLPSNTKVNL 158
            ++ L+ ++++A +EL  L   LD+  +V+ S+  VN+
Sbjct: 118 KAVSLDPQDIQAFRELDKLGVKLDL--RVVASDPSVNI 153
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1639> which encodes the amino acid sequence <SEQ ID 1640>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1874(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 145/162 (89%), Positives = 152/162 (93%)

Query: 1   MITQIRVDDRLIHGQVAVVWTKELNAPLLVVANDEAAKNEITQMTLKMAVPNGMKLLIRS 60
            MITQIRVDDRLIHGQVAVVWTKELNAPLLVVANDEAAKNEITQMTLKMAVPNGMKLLIRS
Sbjct: 1   MITQIRVDDRLIHGQVAVVWTKELNAPLLVVANDEAAKNEITQMTLKMAVPNGMKLLIRS 60

Query: 61  VEESIALFKDPRATDKRIFVIVNSVKDACTIAKNITDLEAVNVANVGRFDKSDPATKVKL 120
            VE+SI LF DPRA DKRIFVIVNSVKDAC IAK + DLEAVNVANVGRFDKSDPA+KVK+
Sbjct: 61  VEDSIKLFNDPRAKDKRIFVIVNSVKDACAIAKEVPDLEAVNVANVGRFDKSDPASKVKV 120

Query: 121 TSSLLLNTEELEAAKELASLPDLDVFNQVLPSNTKVNLSQLV 162
            T SLLLN EE+ AAKEL SLP+LDVFNQVLPSNTKV+LSQLV
Sbjct: 121 TPSLLLNPEEMAAAKELVSLPELDVFNQVLPSNTKVHLSQLV 162
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 512

A DNA sequence (GBSx0550) was identified in *S.agalactiae* <SEQ ID 1641> which encodes the amino acid sequence <SEQ ID 1642>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.22   Transmembrane   87 - 103 (  87 - 104)

----- 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>
```

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

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1643> which encodes the amino acid sequence <SEQ ID 1644>. Analysis of this protein sequence reveals the following:

```
      Possible site: 33

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.44   Transmembrane   87 - 103 (  87 - 104)

----- 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>
```

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

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 115/141 (81%), Positives = 125/141 (88%)

Query: 1   MKRKFLIGSHGKLASGLQSSIDILTGKGQEIQTIDAYIDDSDYTKSIVEFIDEIAPDEQG 60
           MKRKFLIGSHG+LASGLQSSIDIL G GQ ++TIDAY+DDSDYT  I +FI  +A DEQG
Sbjct: 1   MKRKFLIGSHGRLASGLQSSIDILAGMGQALETIDAYVDDSDYTSQIDDFIAGVAADEQG 60

Query: 61  LIFTDLLGGSVNQKMATAVMNSGKNNIFLITNSNLATLLSLLFLKPEEELTKEEIVTVIN 120
           LIFTDLLGGSVNQKM TAVMNSGK+NIFLITNSNLATLLSL+FLKP E LTK+EIVTVIN
Sbjct: 61  LIFTDLLGGSVNQKMVTAVMNSGKDNIFLITNSNLATLLSLVFLKPGEALTKDEIVTVIN 120

Query: 121 ESQVQLVDLSFKAGSEDDFFD 141
           ESQVQLVDL  +  SEDDFFD
Sbjct: 121 ESQVQLVDLVPETNSEDDFFD 141
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 513

A DNA sequence (GBSx0551) was identified in *S.agalactiae* <SEQ ID 1645> which encodes the amino acid sequence <SEQ ID 1646>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2469(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 514

A DNA sequence (GBSx0552) was identified in *S.agalactiae* <SEQ ID 1647> which encodes the amino acid sequence <SEQ ID 1648>. This protein is predicted to be racemase. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
```

```
    INTEGRAL    Likelihood = -8.65   Transmembrane  319 - 335 ( 316 - 339)
    INTEGRAL    Likelihood = -6.10   Transmembrane   18 -  34 (  17 -  37)
    INTEGRAL    Likelihood = -5.68   Transmembrane  230 - 246 ( 227 - 248)
    INTEGRAL    Likelihood = -3.98   Transmembrane  254 - 270 ( 254 - 271)
    INTEGRAL    Likelihood = -3.56   Transmembrane  110 - 126 ( 110 - 129)
    INTEGRAL    Likelihood = -3.19   Transmembrane  161 - 177 ( 156 - 177)
    INTEGRAL    Likelihood = -1.97   Transmembrane  132 - 148 ( 132 - 153)
    INTEGRAL    Likelihood = -1.33   Transmembrane  286 - 302 ( 286 - 302)
    INTEGRAL    Likelihood = -0.59   Transmembrane   53 -  69 (  52 -  69)

----- 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>
```

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

```
>GP:AAF71283 GB:AF253562 racemase [Enterococcus faecalis]
 Identities = 78/262 (29%), Positives = 129/262 (48%), Gaps = 29/262 (11%)

Query: 13  KQHNTSMISLLQYLFSILVILVHSGRLFS-QDVIHFTFKSFLGRMAVPYFLICTAFFLRG 71
           K  + S I    +++ ++L++ +H+  LFS  +  +F F   +  +AVP+F + + FFL
Sbjct: 3   KNESYSGIDYFRFIAALLIVAIHTSPLFSFSETGNFIFTRIVAPVAVPFFFMTSGFFL-- 60

Query: 72  RIQQGLCNHSYFRKLIKK----YSMWTIIYLPY----GYFFFESLNIAKIYLLPGFIVAF 123
            I +  CN        IKK    Y +  ++Y+P     GYF  ++L        LP  I
Sbjct: 61  -ISRYTCNAEKLGAFIKKTTLIYGVAILLYIPINVYNGYFKMDNL-------LPNIIKDI 112

Query: 124 LYLGMSHTLWYIPAVILGWVIIQGLLKYVGTRGTFITVVVLYCIGAV-ETYSVFIQSTKF 182
           ++ G  + LWY+PA I+G  I   L+K V  R  F+    +LY IG    ++Y   ++S
Sbjct: 113 VFDGTLYHLWYLPASIIGAAIAWYLVKKVHYRKAFLIASILYIIGLFGDSYYGIVKSVSC 172

Query: 183 YPLMSTYMSIFQT---TRNGLFYTPVYLLAGYLLYDYFNTDLFTKSRGLK-YILFLLLLA 238
             L   Y  IFQ    TRNG+F+ P++ + G  + D  + + + K   ++ Y LF L+
Sbjct: 173 --LNVFYNLIFQLTDYTRNGIFFAPIFFVLGGYISD--SPNRYRKKNYIRIYSLFCLMFG 228

Query: 239 LENVLIYFN-QGLDKNFFLLAP 259
              L +F+ Q  D  + LL P
Sbjct: 229 KTLTLQHFDIQKHDSMYVLLLP 250
```

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

A related GBS gene <SEQ ID 8589> and protein <SEQ ID 8590> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:      0.23
GvH: Signal Score (-7.5): -5.77
     Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 3 value:  -5.68 threshold:  0.0
   INTEGRAL    Likelihood = -5.68   Transmembrane   41 -  57 (  38 -  59)
   INTEGRAL    Likelihood = -3.98   Transmembrane   65 -  81 (  65 -  82)
   INTEGRAL    Likelihood = -1.33   Transmembrane   97 - 113 (  97 - 113)
   PERIPHERAL  Likelihood =  5.78      10
 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>
```

A related GBS gene <SEQ ID 8591> and protein <SEQ ID 8592> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:     11.50
GvH: Signal Score (-7.5): -2.69
     Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 9 value:  -8.65 threshold:  0.0
   INTEGRAL    Likelihood = -8.65   Transmembrane  310 - 326 ( 307 - 330)
   INTEGRAL    Likelihood = -6.10   Transmembrane    9 -  25 (   8 -  28)
   INTEGRAL    Likelihood = -5.68   Transmembrane  221 - 237 ( 218 - 239)
   INTEGRAL    Likelihood = -3.98   Transmembrane  245 - 261 ( 245 - 262)
   INTEGRAL    Likelihood = -3.56   Transmembrane  101 - 117 ( 101 - 120)
   INTEGRAL    Likelihood = -3.19   Transmembrane  152 - 168 ( 147 - 168)
   INTEGRAL    Likelihood = -1.97   Transmembrane  123 - 139 ( 123 - 144)
   INTEGRAL    Likelihood = -1.33   Transmembrane  277 - 293 ( 277 - 293)
   INTEGRAL    Likelihood = -0.59   Transmembrane   44 -  60 (  43 -  60)
   PERIPHERAL  Likelihood =  5.78     190
 modified ALOM score:   2.23

*** Reasoning Step: 3

----- 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>
```

The protein has homology with the following sequences in the databases:

```
ORF00153(307 - 1140 of 1632)
GP|7960293|gb|AAF71283.1|AF253562_7|AF253562(2   -   284   of   711)   racemase   (Enterococcus
faecalis)
%Match = 8.5
%Identity = 32.7  %Similarity = 54.0
Matches = 91  Mismatches = 113  Conservative Sub.s = 59

150       180       210       240       270       300       330       360
CEISFFIS*YG**GINNNYQIPFKAFQ*LFGIIEIFF*RDWYHSNDNL*KVMLRMKRSQCVDNKQHNTSMISLLQYLFSI
                                                                 | : | |  :::: ::
                                                               MTKNESYSGIDYFRFIAAL
                                                                       10

390       417       447       477       507       537       555
LVILVHSGRLFS-QDVIHFTFKSFLGRMAVPYFLICTAFFLRGRIQQGLCNHSYFRKLIKK----YSMWTIIYLP---Y-
|:: :|:  |||  :  :| |   :  :|||:|:: : |||   | :  ||   :  :|||    | :  ::|:|   |
LIVAIHTSPLFSFSETGNFIFTRIVAPVAVPFFFMTSGFFL---ISRYTCNAEKLGAFIKKTTLIYGVAILLYIPINVYN
         30        40        50        60          70        80        90

603       633       663       693       723       753       783       810
GYFFFESLNIAKIYLLPGFIVAFLYLGMSHTLWYIPAVILGWVIIQGLLKYVGTRGTFITVVVLYCIGAV-ETYSVFIQS
|||  ::|       ||  |   ::  |  : |||:|| |:|  |   |:| |  |  |:   :|| ||   ::|   ::|
GYFKMDNL--------LPNIIKDIVFDGTLYHLWYLPASIIGAAIAWYLVKKVHYRKAFLIASILYIIGLFGDSYYGIVKS
                  110       120       130       140       150       160

840                891       921       951       978      1008      1035
TKFYPLMSTYMSIFQTT---RNGLFYTPVYLLAGYLLYDYFNTDLFTKSRGLK-YILFLLLLALENVLIYFN-QGLDKNF
     |  |  ||| |   |||:|: |:::: |  : |  |   : |   :: | || |::     | :|: |  |  :
--VSCLNVFYNLIFQLTDYTRNGIFFAPIFFVLGGYISDSPNR--YRKKNYIRIYSLFCLMFGKTLTLQHFDIQKHDSMY
          180       190       200       210       220       230       240

1053      1080      1110      1140      1170      1200      1230      1260
FLLAP----LCAVFL-FNWSIRTSLFKEYRLSPLKQLSVYYFFLPPLFIGIVSYCLKSTSLVAHHQGKVIFVVTLALTHA
 || |    |  ::| |    || | :   |  | : |||
VLLLPSVWCLFNLLLHFRGKRRTGL-RTISLDQLYHSSVYDCCNTIVCAELLHLQSLLVENSLVHYIAVCFASVVLAVVI
             260       270       280       290       300       310       320
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 515

A DNA sequence (GBSx0553) was identified in *S.agalactiae* <SEQ ID 1649> which encodes the amino acid sequence <SEQ ID 1650>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3088(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 516

A DNA sequence (GBSx0554) was identified in *S.agalactiae* <SEQ ID 1651> which encodes the amino acid sequence <SEQ ID 1652>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1446(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 517

A DNA sequence (GBSx0555) was identified in *S.agalactiae* <SEQ ID 1653> which encodes the amino acid sequence <SEQ ID 1654>. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      8.28
GvH: Signal Score (-7.5): -2.11
     Possible site: 20
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 6 value:  -8.33 threshold:  0.0
   INTEGRAL    Likelihood = -8.33   Transmembrane  358 - 374 ( 354 - 376)
   INTEGRAL    Likelihood = -8.23   Transmembrane  264 - 280 ( 257 - 290)
   INTEGRAL    Likelihood = -6.37   Transmembrane  210 - 226 ( 206 - 232)
   INTEGRAL    Likelihood = -5.95   Transmembrane  163 - 179 ( 160 - 180)
   INTEGRAL    Likelihood = -5.10   Transmembrane   23 -  39 (  21 -  40)
   INTEGRAL    Likelihood = -1.70   Transmembrane  297 - 313 ( 296 - 314)
   PERIPHERAL  Likelihood =  1.75     322
 modified ALOM score:   2.17

*** Reasoning Step: 3

----- Final Results -----
```

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

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

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

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

Example 518

A DNA sequence (GBSx0556) was identified in *S.agalactiae* <SEQ ID 1655> which encodes the amino acid sequence <SEQ ID 1656>. This protein is predicted to be ABC transporter (ATP-bindingprot). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1510(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10199> which encodes amino acid sequence <SEQ ID 10200> was also identified.

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

```
>GP:CAB88481 GB:AL353816 putative ABC transport system ATP-binding
          protein [Streptomyces coelicolor A3(2)]
 Identities = 104/284 (36%), Positives = 159/284 (55%), Gaps = 18/284 (6%)

Query: 6   TMLLQLDNITKSYGKKIVLNQISYQFTPGLYGLLGANGTGKTTLLNLMSHFTLADSGNIY 65
           T  +    ++  YG+   L+ +S + TPG+ GLLG NG GKTTLL +++    AD G
Sbjct: 2   TPTVSASGLSLHYGRTRALDDVSLRLTPGVTGLLGPNGAGKTTLLRVLATAVPADRGAFT 61

Query: 66  WNGQEQS-----EEFYRHIGFLPQHFRYYDQFTGIAFLNYIATLKGV-DKKKAKQEIPRL 119
             G +        +E  R +G+LPQ    ++  FT   F++Y+A LK + D+++  +E+ R+
Sbjct: 62  VLGHDPGSSRGRQEVRRRLGYLPQTPGFHPDFTAFEFVDYVAILKELADRRERHREVRRV 121

Query: 120 LELVGLGDVGKKKISSYSGGMKQRLGIAQALINDPEILILDEPTVGLDPKERVKFRHILS 179
           LE V LG+V  ++I   SGGM+QR+ +A AL+ DP  L+LDEPTVGLDP++R++FR +++
Sbjct: 122 LEEVDLGEVRGRRIKKLSGGMRQRVALAAALVGDPGFLVLDEPTVGLDPEQRMRFRELIA 181

Query: 180 QLSTNKIIILSTHIVSDVEAVAKEIIVLKNGKFIEHGNTAQLLKTIEGKVWEIT-TEPGL 238
                + ++LSTH    DV  +   +IV+  G     G  A+L     G+VW  T  +PG
Sbjct: 182 GAGEGRTVLLSTHQTEDVAMLCHRVIVMAAGAVRFDGTPAELTARAAGRVWSSTEKDPG- 240

Query: 239 SQIPNIAIVNEKVFSDSRVFRVVSDICPSDSAQLVVPTLEDFYI 282
                A    +  + S  FR V D  P   A+   PTLED Y+
Sbjct: 241 ------AKAGWRTGTGS--FRNVGD--PPPGAEPAEPTLEDGYL 274
```

There is also homology to SEQ ID 686.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 519

A DNA sequence (GBSx0557) was identified in *S.agalactiae* <SEQ ID 1657> which encodes the amino acid sequence <SEQ ID 1658>. This protein is predicted to be response regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3781(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 GENPEPT database:

```
>GP:CAC10170 GB:AJ278301 response regulator [Streptococcus pneumoniae]
 Identities = 136/242 (56%), Positives = 183/242 (75%)

Query: 1   MNIFILEDDFVQQAHFEKIIKEIRVQYNLHFKTVETFAKPVQLLESIYEIGLHNLFFLDI 60
           M IF+LEDDF QQ   E  I+++  ++++   + E F KP QLL  ++E G H LFFLDI
Sbjct: 1   MRIFVLEDDFSQQTRIETTIEKLLKEHHITLSSFEVFGKPDQLLAEVHEKGAHQLFFLDI 60

Query: 61  EIKNDEQMGLEVAKQIRQVDPYAQIVFVTTHSELMPLTFRYQVSALDYIDKGLSQEEFSQ 120
           EI+N+E  GLEVA++IR+ DPYA IVFVTTHSE MPL+FRYQVSALDYIDK LS EEF
Sbjct: 61  EIRNEEMKGLEVARKIREQDPYALIVFVTTHSEFMPLSFRYQVSALDYIDKALSAEEFES 120

Query: 121 RIEEVLLYVDGICNKPLVENSFYFKSRYSQVQLPFNDLLYIETSSRSHRVVLYTEKDRME 180
           RIE  LLY +   +K L E+ FYFKS+++Q Q PF ++ Y+ETS R HRV+LYT+ DR+E
Sbjct: 121 RIETALLYANSQDSKSLAEDCFYFKSKFAQFQYPFKEVYYLETSPRPHRVILYTKTDRLE 180

Query: 181 FTATLGDILKQEPRLFQCHRSFLVNPLNIFKVDRIDRLVYFQNGTTCLVSRNKVRDIVSI 240
           FTA+L ++ KQEPRL QCHRSFL+NP N+  +D+ ++L++F NG +CL++R KVR++
Sbjct: 181 FTASLEEVFKQEPRLLQCHRSFLINPANVVHLDKKEKLLFFPNGGSCLIARYKVREVSEA 240

Query: 241 VD 242
           ++
Sbjct: 241 IN 242
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1659> which encodes the amino acid sequence <SEQ ID 1660>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2098(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 106/235 (45%), Positives = 159/235 (67%)

Query: 1   MNIFILEDDFVQQAHFEKIIKEIRVQYNLHFKTVETFAKPVQLLESIYEIGLHNLFFLDI 60
           MNIFILEDDF+QQ   E I+  I  +  +    +E F+ P +L ESI E G H L+FLDI
Sbjct: 2   MNIFILEDDFIQQTRIESIVVGILKETRIPCNQLEVFSTPQKLFESIQERGDHQLYFLDI 61

Query: 61  EIKNDEQMGLEVAKQIRQVDPYAQIVFVTTHSELMPLTFRYQVSALDYIDKGLSQEEFSQ 120
           EI    + GLE+A  IRQ DP A IVFVTTHSE  P++F+Y+VSALD+IDK   Q++F +
Sbjct: 62  EIGEYTRCGLELAAAIRQKDPNAVIVFVTTHSEFAPISFKYKVSALDFIDKAGGQKQFKE 121

Query: 121 RIEEVLLYVDGICNKPLVENSFYFKSRYSQVQLPFNDLLYIETSSRSHRVVLYTEKDRME 180
           +IEE + Y   + +    ++ F F++  ++++LP+ D+LY  T++  H+V L+T+ +R+E
Sbjct: 122 QIEECIRYTYDMMSSRESKDMFLFETPQTRLKLPYKDILYFATATTPHKVCLWTQTERLE 181
```

```
Query: 181 FTATLGDILKQEPRLFQCHRSFLVNPLNIFKVDRIDRLVYFQNGTTCLVSRNKVR 235
           F   L +I     P+LF CHRS+LVN   + ++D+  +L+YF+NG +C+VSR K++
Sbjct: 182 FYGNLSEIQAVAPKLFLCHRSYLVNLDKVVRIDKSKQLLYFENGDSCMVSRLKMK 236
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 520

A DNA sequence (GBSx0558) was identified in *S.agalactiae* <SEQ ID 1661> which encodes the amino acid sequence <SEQ ID 1662>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2651(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1663> which encodes the amino acid sequence <SEQ ID 1664>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0535(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 177/269 (65%), Positives = 219/269 (80%)

Query: 6   MAKCLTLNTHSWMEVNALKKLFDLAEHIFREKYDIICLQEVNQSISSPLAKSSPNYHPIE 65
           M K LTLNTHSWM+ N LKKL  LAEHI  EKYDIICLQE+NQ I S LA   P Y  +
Sbjct: 1   MTKVLTLNTHSWMQANTLKKLVALAEHILAEKYDIICLQEINQLIESELATDLPRYQALS 60

Query: 66  GTPALHQDNFALQLVHYLNLQGLHYHWTWAYNHIGYSKYHEGVAILSLKPLKPEDILVSA 125
           GTP++H+D+FAL L+HYL  +G HY+W+WAYNHIGY  Y EGVAILS +P+   DILVSA
Sbjct: 61  GTPSIHKDHFALLLIHYLQKRGQHYYWSWAYNHIGYDIYQEGVAILSKQPIHVSDILVSA 120

Query: 126 VDDETDYHTRRALVAETTLNDKVVTVVSLHFSWFEKGFAEEWKRLETTLLEVETPLLLMG 185
           +DDETDYHTRR+L+A+TTL+ K V VV++H SWF+KGF  EW++LE  LL +  PLLLMG
Sbjct: 121 MDDETDYHTRRSLIAKTTLDGKEVAVVNVHLSWFDKGFLGEWEKLEKELLTLNCPLLLMG 180

Query: 186 DFNNPTGNQGYELVLNSPLALKDSHQIANHVFGDHTIMADIDGWEGNKKALKVDHIFTSE 245
           DFNNPT   GY++++ SPL L+DSH+ A+HVFGDH+I+ADIDGW+GNK+ALKVDH+FTS+
Sbjct: 181 DFNNPTDQDGYQVMMGSPLDLQDSHKGADHVFGDHSIVADIDGWQGNKEALKVDHVFTSK 240

Query: 246 DLSISSSQVVFEGGEAPVVSDHYGLEITM 274
           D  I SS++ FEGG+APVVSDHYGLE+T+
Sbjct: 241 DFIIRSSKITFEGGDAPVVSDHYGLEVTL 269
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 521

A DNA sequence (GBSx0559) was identified in *S.agalactiae* <SEQ ID 1665> which encodes the amino acid sequence <SEQ ID 1666>. This protein is predicted to be PTS system, glucose-specific enzyme II, A component (ptsG). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.07   Transmembrane  193 - 209 ( 189 - 217)
   INTEGRAL    Likelihood = -7.86   Transmembrane   28 -  44 (  24 -  48)
   INTEGRAL    Likelihood = -6.48   Transmembrane  431 - 447 ( 421 - 449)
   INTEGRAL    Likelihood = -2.92   Transmembrane  153 - 169 ( 153 - 170)
   INTEGRAL    Likelihood = -2.81   Transmembrane   93 - 109 (  93 - 111)
   INTEGRAL    Likelihood = -2.39   Transmembrane  370 - 386 ( 370 - 388)
   INTEGRAL    Likelihood = -2.28   Transmembrane   68 -  84 (  68 -  84)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 10201> which encodes amino acid sequence <SEQ ID 10202> was also identified.

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

```
>GP:AAD00281 GB:U78600 putative ptsG protein [Streptococcus mutans]
 Identities = 294/409 (71%), Positives = 342/409 (82%), Gaps = 7/409 (1%)

Query: 293 DLINLKGS-NSSQYHHLLTSVTPARFKVGQMIGASGILMGLSYAMYRNVDKDKKLKYKSM 351
           DLI+LKG+ + SQYHHLLTSVTPARFKVGQMIG+SGILMGL+ AMYRNVD DKK KYK M
Sbjct: 3   DLIHLKGAGHMSQYHHLLTSVTPARFKVGQMIGSSGILMGLTLAMYRNVDPDKKEKYKGM 62

Query: 352 FISAAAATFLTGVTEPIEYMFMFAAMPLYLVYAVVQGCAFAMADIVNLRVHSFGNIEFLT 411
           F+SAA A FLTGVTEP+EYMFMFAA+PLYLVYAVVQG AFA AD+++LRVHSFGNIEFLT
Sbjct: 63  FLSAAVAVFLTGVTEPLEYMFMFAALPLYLVYAVVQGLAFASADLIHLRVHSFGNIEFLT 122

Query: 412 RVPMGIKAGLGGDIFNFVWVTLLFAVLMYFIANFMIKKFNLATAGRNGNYDNEEVDNAPS 471
           + PM IKAGL  DI NF+ V+++F V MYFI NFMIKKFNLAT+GRNGNYD  + D +
Sbjct: 123 KTPMAIKAGLAMDIVNFIVVSVVFGVAMYFITNFMIKKFNLATSGRNGNYDTGD-DASDE 181

Query: 472 TAS----GSADANSQVVQVINLLGGRDNIEDVDACMTRLRVTVKDGNSVGSEAAWKKAGA 527
           TAS    G+A+ANSQ+V++INLLGG++NI DVDACMTRLR+TV D   VG EAAWKKAGA
Sbjct: 182 TASNSNAGTANANSQIVKIINLLGGKENISDVDACMTRLRITVTDVAKVGDEAAWKKAGA 241

Query: 528 MGLVLKGNGVQAIYGPKADVLKSDIQDLLDSGTVIPIVDLETGQPVAAAPVTTYKGITEE 587
           MGL++KGNGVQA+YGPKADVLKSDIQDLLDSG  IP  D+   +    A V ++KG+TEE
Sbjct: 242 MGLIVKGNGVQAVYGPKADVLKSDIQDLLDSGVDIPKTDVTAPEEDKTADV-SFKGVTEE 300

Query: 588 IVSVANGQVEALDVVKDPVFSQKMMGDGFAVEPTDGNIYVPVSGTVTSVFPTKHAFGLLT 647
           + +VA+GQV  +  V DPVFSQKMMGDGFAVEP +GNIY PV+G VTSVFPTKHA GLLT
Sbjct: 301 VATVADGQVLPITQVHDPVFSQKMMGDGFAVEPENGNIYSPVAGLVTSVFPTKHALGLLT 360

Query: 648 ESGLEVLVHIGLDTVALDGQPFEVKISSGQKVVAGDLAVVADLEAIKAA 696
           + GLEVLVH+GLDTVAL+G PF  K+  GQ+V  GDL +VADLEAIK+A
Sbjct: 361 DDGLEVLVHVGLDTVALNGAPFSAKVKDGQRVALGDLLLVADLEAIKSA 409
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1667> which encodes the amino acid sequence <SEQ ID 1668>. Analysis of this protein sequence reveals the following:

```
     Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-13.43   Transmembrane  186 - 202 ( 181 - 213)
   INTEGRAL    Likelihood = -6.79   Transmembrane  419 - 435 ( 412 - 442)
   INTEGRAL    Likelihood = -5.52   Transmembrane   61 -  77 (  57 -  82)
```

```
   INTEGRAL     Likelihood = -3.56   Transmembrane  363 - 379 ( 363 - 381)
   INTEGRAL     Likelihood = -1.97   Transmembrane  143 - 159 ( 142 - 160)
   INTEGRAL     Likelihood = -0.16   Transmembrane  343 - 359 ( 343 - 359)

----- Final Results -----
               bacterial membrane --- Certainty=0.6371(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:

```
>GP:AAD00281 GB:U78600 putative ptsG protein [Streptococcus mutans]
 Identities = 288/407 (70%), Positives = 331/407 (80%), Gaps = 2/407 (0%)

Query: 286 DLVHLKGSD-ASAYSHLMDSVTPARFKVGQMIGATGTLMGVALAMYRNVDADKKHTYKMM 344
           DL+HLKG+   S Y HL+ SVTPARFKVGQMIG++G LMG+ LAMYRNVD DKK  YK M
Sbjct: 3   DLIHLKGAGHMSQYHHLLTSVTPARFKVGQMIGSSGILMGLTLAMYRNVDPDKKEKYKGM 62

Query: 345 FISAAAAVFLTGVTEPLEYLFMFAAMPLYIVYALVQGASFAMADLVNLRVHSFGNIELLT 404
           F+SAA AVFLTGVTEPLEY+FMFAA+PLY+VYA+VQG +FA ADL++LRVHSFGNIE LT
Sbjct: 63  FLSAAVAVFLTGVTEPLEYMFMFAALPLYLVYAVVQGLAFASADLIHLRVHSFGNIEPLT 122

Query: 405 RTPMALKAGLGMDVINFVWVSVLFAVIMYFIADMMIKKMHLATAGRLGNYDA-DILGDRN 463
           +TPMA+KAGL MD++NF+ VSV+F V MYFI + MIKK +LAT+GR GNYD  D   D
Sbjct: 123 KTPMAIKAGLAMDIVNFIVVSVVFGVAMYFITNFMIKKFNLATSGRNGNYDTGDDASDET 182

Query: 464 TQTRPTQVADSNSQVVQIVNLLGGAGNIDDVDACMTRLRVTVKDPAKVGAEDDWKKAGAI 523
                A++NSQ+V+I+NLLGG  NI DVDACMTRLR+TV D AKVG E  WKKAGA+
Sbjct: 183 ASNSNAGTANANSQIVKIINLLGGKENISDVDACMTRLRITVTDVAKVGDEAAWKKAGAM 242

Query: 524 GLIQKGNGVQAVYGPKADILKSDIQDLLDSGALIPEVNMSQLTSKPTPAKDFKHVTEDVL 583
           GLI KGNGVQAVYGPKAD+LKSDIQDLLDSG  IP+ +++       T    FK VTE+V
Sbjct: 243 GLIVKGNGVQAVYGPKADVLKSDIQDLLDSGVDIPKTDVTAPEEDKTADVSFKGVTEEVA 302

Query: 584 SVADGMVLPITGVKDQVFAAKMMGDGFAVEPTHGNIYAPVAGLVTSVFPTKHAFGLLTDN 643
           +VADG VLPIT V D VF+ KMMGDGFAVEP +GNIY+PVAGLVTSVFPTKHA GLLTD+
Sbjct: 303 TVADGQVLPITQVHDPVFSQKMMGDGFAVEPENGNIYSPVAGLVTSVFPTKHALGLLTDD 362

Query: 644 GLEVLVHVGLDTVALNGVPFSVKVSEGQRVHAGDLLVVADLAAIKSA 690
           GLEVLVHVGLDTVALNG PFS KV +GQRV  GDLL+VADL AIKSA
Sbjct: 363 GLEVLVHVGLDTVALNGAPFSAKVKDGQRVALGDLLLVADLEAIKSA 409
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 517/731 (70%), Positives = 606/731 (82%), Gaps = 7/731 (0%)

Query: 8   MKNNVKQLFSFEFWQKFGKALMVVIAVMPAAGLMVSIGNSISLLDPSNVLLGRIANVIAQ 67
           MK + KQLF FEFWQKFGK LMVVIAVMPAAGLM+SIGNSI +++  +  L  + N+IAQ
Sbjct: 1   MKTSFKQLFRFEFWQKFGKCLMVVIAVMPAAGLMISIGNSIPMINHDSAFLASLGNIIAQ 60

Query: 68  IGWGVIGNLHILFALAIGGSWAKERAGGAFAAGLSFILINLITGNFFGVKTDMLADSKAT 127
           IGW VI NLH+LFALAIGGSWAKERAGGAFA+GL+F+LIN ITG F+GV + MLAD +A
Sbjct: 61  IGWAVIVNLHLLFALAIGGSWAKERAGGAFASGLAFVLINRITGAFYGVSSTMLADPEAK 120

Query: 128 VQTVFGATIRVSDYFVNVLGQPALNMGVFVGIISGFVGATAFNKYYNYRKLPDALTFFNG 187
           + ++ G  + V DYF +VL  PALN GVFVGII+GFVGATA+NKYYNYRKLP+ LTFFNG
Sbjct: 121 ITSLLGTQMIVKDYFTSVLESPALNTGVFVGIIAGFVGATAYNKYYNYRKLPEVLTFFNG 180

Query: 188 KRFVPFVVIYRSVIVALILSVFWPVVQSGINGFGKWIASSQDSAPILAPFVYGTLERLLL 247
           KRFVPFVVI RS+ VALIL V WPV+QSGIN FG WIASSQDSAPILAPF+YGTLERLLL
Sbjct: 181 KRFVPFVVILRSIFVALILVVVWPVIQSGINSFGMWIASSQDSAPILAPFLYGTLERLLL 240

Query: 248 PFGLHHMLTIPMNYTQLGGTYTVLTGATKGAQVLGQDPLWLAWVGDLINLKGSNSSQYHH 307
           PFGLHHMLTIPMNYT LGGTY V+TGA  G +V GQDPLWLAWV DL++LKGS++S Y H
Sbjct: 241 PFGLHHMLTIPMNYTALGGTYEVMTGAAAGTKVFGQDPLWLAWVTDLVHLKGSDASAYSH 300

Query: 308 LLTSVTPARFKVGQMIGASGILMGLSYAMYRNVDKDKKLKYKSMFISAAAATFLTGVTEP 367
           L+ SVTPARFKVGQMIGA+G LMG++ AMYRNVD DKK  YK MFISAAAA FLTGVTEP
```

```
Sbjct: 301 LMDSVTPARFKVGQMIGATGTLMGVALAMYRNVDADKKHTYKMMFISAAAAVFLTGVTEP 360

Query: 368 IEYMFMFAAMPLYLVYAVVQGCAFAMADIVNLRVHSFGNIEFLTRVPMGIKAGLGGDIFN 427
           +EY+FMFAAMPLY+VYA+VQG +FAMAD+VNLRVHSFGNIE LTR PM +KAGLG D+ N
Sbjct: 361 LEYLFMFAAMPLYIVYALVQGASFAMADLVNLRVHSFGNIELLTRTPMALKAGLGMDVIN 420

Query: 428 FVWVTLLFAVLMYFIANFMIKKFNLATAGRNGNYDNEEVD--NAPSTASGSADANSQVVQ 485
           FVWV++LFAV+MYFIA+ MIKK +LATAGR GNYD + +   N  +  + AD+NSQVVQ
Sbjct: 421 FVWVSVLFAVIMYFIADMMIKKMHLATAGRLGNYDADILGDRNTQTRPTQVADSNSQVVQ 480

Query: 486 VINLLGGRDNIEDVDACMTRLRVTVKDGNSVGSEAAWKKAGAMGLVLKGNGVQAIYGPKA 545
           ++NLLGG NI+DVDACMTRLRVTVKD  VG+E WKKAGA+GL+ KGNGVQA+YGPKA
Sbjct: 481 IVNLLGGAGNIDDVDACMTRLRVTVKDPAKVGAEDDWKKAGAIGLIQKGNGVQAVYGPKA 540

Query: 546 DVLKSDIQDLLDSGTVIPIVDLE--TGQPVAAAPVTTYKGITEEIVSVANGQVEALDVVK 603
           D+LKSDIQDLLDSG +IP V++   T +P    P   +K +TE+++SVA+G V  +  VK
Sbjct: 541 DILKSDIQDLLDSGALIPEVNMSQLTSKP---TPAKDFKHVTEDVLSVADGMVLPITGVK 597

Query: 604 DPVFSQKMMGDGFAVEPTDGNIYVPVSGTVTSVFPTKHAFGLLTESGLEVLVHIGLDTVA 663
           D VF+ KMMGDGFAVEPT GNIY PV+G VTSVFPTKHAFGLLT++GLEVLVH+GLDTVA
Sbjct: 598 DQVFAAKMMGDGFAVEPTHGNIYAPVAGLVTSVFPTKHAFGLLTDNGLEVLVHVGLDTVA 657

Query: 664 LDGQPFEVKISSGQKVVAGDLAVVADLEAIKAAGKETSVIIVFTNVSDIKTVKLEKSGPQ 723
           L+G PF VK+S GQ+V AGDL VVADL AIK+A +ET +++ FTN ++I+ V L   G Q
Sbjct: 658 LNGVPFSVKVSEGQRVHAGDLLVVADLAAIKSAERETIIVVAFTNTTEIQDVTLTSLGAQ 717

Query: 724 IAKTVVAKVEL 734
            AKT VA VEL
Sbjct: 718 PAKTKVATVEL 728
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 522

A DNA sequence (GBSx0560) was identified in *S.agalactiae* <SEQ ID 1669> which encodes the amino acid sequence <SEQ ID 1670>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2266(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 523

A DNA sequence (GBSx0561) was identified in *S.agalactiae* <SEQ ID 1671> which encodes the amino acid sequence <SEQ ID 1672>. This protein is predicted to be alkaline phosphatase synthesis sensor protein phor (hpkA). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.96   Transmembrane  160 - 176 ( 148 - 183)
   INTEGRAL    Likelihood = -8.65   Transmembrane   20 -  36 (  13 -  41)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.6583(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8595> which encodes amino acid sequence <SEQ ID 8596> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
SRCFLG: 0
McG: Length of UR:   26
     Peak Value of UR:   3.27
     Net Charge of CR: 3
McG: Discrim Score:     14.63
GvH: Signal Score (-7.5): -5.64
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 2 value: -13.96 threshold:  0.0
   INTEGRAL    Likelihood =-13.96   Transmembrane  152 - 168 ( 140 - 175)
   INTEGRAL    Likelihood = -8.65   Transmembrane   12 -  28 (   5 -  33)
   PERIPHERAL  Likelihood =  1.59     135
 modified ALOM score:   3.29
icm1 HYPID: 7  CFP: 0.658

*** Reasoning Step:. 3

----- Final Results -----
              bacterial membrane --- Certainty=0.6583(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS gene <SEQ ID 8593> and protein <SEQ ID 8594> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:     14.63
GvH: Signal Score (-7.5): -5.64
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 2 value: -13.96 threshold:  0.0
   INTEGRAL    Likelihood =-13.96   Transmembrane  152 - 168 ( 140 - 175)
   INTEGRAL    Likelihood = -8.65   Transmembrane   12 -  28 (   5 -  33)
   PERIPHERAL  Likelihood =  1.59     135
 modified ALOM score:   3.29

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.6583(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:

```
34.9/61.1% over 363aa
                                                  Thermotoga maritima
  EGAD|131465| sensor histidine kinase HpkA Insert characterized
  GP|1575578|gb|AAC44437.1||U67196 histidine protein kinase Insert characterized
  GP|4982228|gb|AAD36721.1|AE001807_12|AE001807   sensor   histidine   kinase   HpkA   Insert
characterized
  PIR|C72228|C72228 sensor histidine kinase HpkA - (strain MSB8) Insert characterized

ORF00680(919 - 1977 of 2277)
```

```
EGAD|131465|TM1654(48 - 411 of 412) sensor histidine kinase HpkA {Thermotoga maritima}
GP|1575578|gb|AAC44437.1||U67196 histidine protein kinase {Thermotoga maritima}
GP|4982228|gb|AAD36721.1|AE001807_12|AE001807 sensor histidine kinase HpkA {Thermotoga
maritima} PIR|C72228|C72228 sensor histidine kinase HpkA - Thermotoga maritima (strain
MSB8)
%Match = 13.6
%Identity = 34.8  %Similarity = 61.0
Matches = 125  Mismatches = 134  Conservative Sub.s = 94

720       750       780       810       840       870       900       930
AAQRLNNGTIVRLSVAQQTIFYLLLGMISPLAIIILLAIILSVLIARYIAKKVSEPLNNIDLDHPLSNDSYEEITPLLRR
                                  : ::  :| ||:     |::||    |:    :  :       |: |
                                MSVFLFVIVAVLFVLLFLVFKKRLSEYKILIEKLSDMLGEKGVPPLYLFER
                                        10        20        30        40        50

960       990       1020      1050      1080      1110      1140      1170
LDSHQAKIQHQKLLLQKRQKEFDTIISKIKEGMILLDDQARIVSINAEALKLFQINDDWHGRFMMEVSRDLTLKDLIDQG
|  :   ::     ::  :  | ||:: : | : :|| : :|  :|  | :| |   :  ||   |:  |  : :::::
LKKYVDNLKETISRVEVSRDNFLTILNSLSEPIFILDREGKITFLNEIARELVQGRINPEGRPYYEIFEDYYINEMVEET
                  70        80        90        100       110       120       130

1197      1215      1245      1275      1305      1335      1365      1395
LKGKK-KEAN----IGIENNHYRVLVRPTTDNNRVTGLVVLLFDVTDQLQMEQLQREFTANVSHELKTPLHVISGYSELL
:| :: :|      :| |  :: | | |    :    :|:|: ||| : ::::::||| | |||||:|||  | ||:| |
IKSEEPQEGTLVTYVGNEKKYFHVKVIPVELKSGDKIFVILFHDVTKERKLDEMRREFIATVSHELRTPLTSIHGYAETL
                  150       160       170       180       190       200       210

1452      1482      1512      1539      1569      1599      1629
ANQMVPNEE-VPQFAAKIHKESERLVKLVEDIINLSHLDEQE-KLPQETVNLYDLTQKVLEGLQAKADKKHIQINFNGEE
    : |:| | :|   | :|| |: :|: |:::|  ::| |  :  : |:| :: : |   :|  |::  : :    |:
LEDDLENKELVKRFLKIIEEESARMTRLINDLLDLEKIEESEANFEMKDVDLCEVIEYVYRIIQPIAEENEVDLIVECED
                  230       240       250       260       270       280       290

1659      1689      1713                1767      1797      1827      1857
AILRGNPVLLNSLVYNLCDNAITYNH--EKGQVNVTLK--NSPDTITLEVSDTGLGIAEKDKKRIFERFYRVDKSRSKIV
 ::|||.  |  :: || |||: |    |||:  | ::  ::|| : :|| ||| || :: : ||||:||||||:||: :
VVVRGNKERLIQMLLNLVDNAVKYTSLKEKGEKKVWVRAYDTPDWVVVEVEDTGPGIPKEAQSRIFEKFYRVDKARSRKM
                  310       320       330       340       350       360       370

1887      1917      1947      1977      2007      2037      2067      2097
GGTGLGLSIVKSALDFHNGSIKVDSHLGQGTTMTVLLHKQ*KLTNKSLDDII*TFLVIQKKYISKGLQKTNKCYNKTXX*
|||||||:|||: :| | | |:|:| : ||| | ||| |:
GGTGLGLTIVKTIVDKHGGKIEVESEINQGTLMRVLLPKRR
                  390       400       410
```

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

```
>GP:BAB06875 GB:AP001517 two-component sensor histidine kinase
          involved in phosphate regulation [Bacillus halodurans]
 Identities = 176/589 (29%), Positives = 315/589 (52%), Gaps = 47/589 (7%)

Query: 9   MTKKIFRTTLSASLGIVLVTILMIMG------------FLYNYFNHIQREQLRTQTALAS 56
           MTK  +R  L+     ++ VT+L++ G             +L N  + +++E      + + +
Sbjct: 1   MTKFRYRLVLA----VLTVTLLVMAGLGLVIGQIFKNVYLENLTDRLKKETYLAASMVEN 56

Query: 57  QGISF-EGKDYFENLKTS-NVRITWVDNKGQVLYDTQSDAKHMKNHANRQEIKEAIKSGY 114
           + + F E +    E +     + R+T +   G V+ ++ +D   M+NHA+R E  E ++ G
Sbjct: 57  EAVLFNEVQTLTEEISQKLDARVTIILADGTVVGESAADPAEMENHADRPEFTE-LEEGI 115

Query: 115 GESTRWSATL-TEKSIYAAQRLN--NGTI--VRLSVAQQTIFYLLLGMISPLAIIILLAI 169
               R+S T+ TE    YA    N  N TI  VRL +  + +  +   + + L +   +A
Sbjct: 116 ---VRYSTTVETELLFYAVPIQNEANETIGYVRLGLPIEAVNSVNRTLWAILIVSFTIAF 172

Query: 170 ILSVLIARYIAKKVSEPLNNI----------DLDHPLSNDSYEEITPLLRRLDSHQAKIQ 219
           ++ V +   IA ++  P+ +           D     S +S +E+  L R ++      ++
Sbjct: 173 LVIVSVTYRIANQMIRPIESATVVANKLAEGDYQARTSEESRDEVGQLNRSINVLAYNLE 232

Query: 220 HQKLLLQKRQKEFDTIISKIKEGMILLDDQARIVSINAEALKLFQINDD-WHGRFMMEVS 278
```

```
                Q +++  +T+I  +  G+IL++ +  I  IN     +FQ + D W  +   +V
Sbjct: 233 QLTKRHQVQKERLETLIENMGSGLILINTRGDISLINKTCHDIFQEDTDLWLHQLYHDVI 292

Query: 279 RDLTLKDLIDQGLKGKKKEAN-----IGIENNHYRVLVRPTT-DNNRVTGLVVLLFDVTD 332
           +   +  ++      +K++        I +E  H+ V   P   +N ++ G+ ++  D+T+
Sbjct: 293 KHKEIIKIVQDIFLTEKRQRRQVKLPIHLEYRHFDVHGAPIVRENGKLKGIALVFHDITE 352

Query: 333 QLQMEQLQREFTANVSHELKTPLHVISGYSELLANQMVPNEEV-PQFAAKIHKESERLVK 391
             ++EQ++++F ANVSHELKTP+  I G++E L +  + +E++  QF   I KESERL
Sbjct: 353 LKKLEQVRKDFVANVSHELKTPVTSIKGFTETLLDGAMHDEQLRDQFLHIIWKESERLQS 412

Query: 392 LVEDIINLSHLDEQE-KLPQETVNLYDLTQKVLEGLQAKADKKHIQINFNGEEAI-LRGN 449
           L+ D++ LS +++   +L  +  NL+ +  +V+  L+ KA++K I I+ + E +  L G+
Sbjct: 413 LIHDLLELSKIEQNYFQLNWQQTNLFAVVSEVMTLLKGKAEEKGIDISLSAEGSFDLEGD 472

Query: 450 PVLLNSLVYNLCDNAITYNHEKGQVNVTLKNSPDTITLEVSDTGLGIAEKDKKRIFERFY 509
           P  L  +  NL +NAITY    G++++ LK+  D +  EV+DTG+GI E +  RIFERFY
Sbjct: 473 PERLKQIAINLVNNAITYTSNGGRIDLALKDHGDVVEFEVNDTGIGIRESEIPRIFERFY 532

Query: 510 RVDKSRSKIVGGTGLGLSIVKSALDFHNGSIKVDSHLGQGTTMTVLLHK 558
           RVD++RS+  GGTGLGL+IVK  ++ H G I V+S  G+GTT T+  H+
Sbjct: 533 RVDRARSRNSGGTGLGLAIVKHLVEAHQGKILVESEFGKGTTFTIQFHR 581
```

There is also homology to SEQ ID 1178.

SEQ ID 8594 (GBS340) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 10; MW 86kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 7; MW 61.5kDa) and in Figure 77 (lane 10; MW 62kDa).

Purified GBS340-GST is shown in Figure 223, lane 2; purified GBS340-His is shown in Fig. 191, lane 9.

The purified GBS340-GST fusion product was used to immunise mice. The resulting antiserum was used for Western blot (Figure 254A), FACS (Figure 254B), 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 was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 524

A DNA sequence (GBSx0562) was identified in *S.agalactiae* <SEQ ID 1673> which encodes the amino acid sequence <SEQ ID 1674>. This protein is predicted to be phosphate regulon transcriptional regulatory protein phob (phoB). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2617(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10203> which encodes amino acid sequence <SEQ ID 10204> was also identified.

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

```
>GP:AAC73502 GB:AE000146 positive response regulator for pho
```

```
          regulon, sensor is PhoR (or CreC) [Escherichia coli K12]
 Identities = 98/224 (43%), Positives = 138/224 (60%), Gaps = 2/224 (0%)

Query: 2   IYCVEDDADIREMMLYTLQMAGFKAQGFSSSELFWEAIQEKVPDLILLDIMLPGDDGLTI 61
           I  VED+A IREM+ + L+  GF+       +     + E  PDLILLD MLPG  G+
Sbjct: 5   ILVVEDEAPIREMVCFVLEQNGFQPVEAEDYDSAVNQLNEPWPDLILLDWMLPGGSGIQF 64

Query: 62  LERLRRKHQTEMIPVIMTTAKGSEYDKVKGLDLGADDYLVKPFGMMEMISRIKAVLRRSR 121
           ++ L+R+  T  IPV+M TA+G E D+V+GL+ GADDY+ KPF   E+++RIKAV+RR
Sbjct: 65  IKHLKRESMTRDIPVVMLTARGEEEDRVRGLETGADDYITKPFSPKELVARIKAVMRRIS 124

Query: 122 QVDSKAHIIIGNLEIDPTNYWVKRGTEKIHLTLKEFELLVLFFRNPNRVFTRQELLDKVW 181
            +  +  I +  L +DPT++ V  G E + +   EF+LL  F  +P RV++R++LL+ VW
Sbjct: 125 PMAVEEVIEMQGLSLDPTSHRVMAGEEPLEMGPTEFKLLHFFMTHPERVYSREQLLNHVW 184

Query: 182 GEQFLGETRTVDVHIGTLRTKLGEDGY--LIATVRGVGYRLEER 223
           G     E RTVDVHI  LR  L   G+  ++ TVRG GYR   R
Sbjct: 185 GTNVYVEDRTVDVHIRRLRKALEPGGHDRMVQTVRGTGYRFSTR 228
```

There is also homology to SEQ ID 1182.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 525

A DNA sequence (GBSx0563) was identified in *S.agalactiae* <SEQ ID 1675> which encodes the amino acid sequence <SEQ ID 1676>. This protein is predicted to be phosphate transport system regulatory protein (phoU). Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1188(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 GENPEPT database:

```
>GP:AAG08750 GB:AE004948 phosphate uptake regulatory protein PhoU
          [Pseudomonas aeruginosa]
 Identities = 66/213 (30%), Positives = 119/213 (54%), Gaps = 4/213 (1%)

Query: 2   IRSRFASQLNDLNKEIIFMGALCEDIIGKSLGALTNSNDVYLDDISETYHKIEQMERDIE 61
           I  +F ++L D+   ++ MG L E  +  ++ AL +++      + E   +I QMER+I+
Sbjct: 11  ISQQFNAELEDVRSHLLAMGGLVEKQVNDAVNALIDADSGLAQQVREIDDQINQMERNID 70

Query: 62  ERCLKLLLRQQPVAKDLRRISSALKMVYDMKRIGAQAYEIAEIVSLGHIIQGSGSERD-- 119
           E C+++L R+QP A DLR I S  K V D++RIG +A ++A       + +   S R
Sbjct: 71  EECVRILARRQPAASDLRLIISISKSVIDLERIGDEASKVARRAI--QLCEEGESPRGYV 128

Query: 120 QLNSMSNNVISMLTKSIDAFIYDNEEQAHQVIEQDRTVNQEFDTIKKQLVLYFSVQDVDG 179
           ++  + + V  M+ +++DAF   + + A  V + D+TV++E+ T  ++LV Y
Sbjct: 129 EVRHIGSQVQKMVQEALDAFARFDADLALSVAQYDKTVDREYKTALRELVTYMMEDPRAI 188

Query: 180 EYPIDVLMIAKYLERIGDHTVNIAKWVLFSITG 212
              ++++   + LERIGDH  NIA+ V++ + G
Sbjct: 189 SRVLNIIWALRSLERIGDHARNIAELVIYLVRG 221
```

There is also homology to SEQ ID 1678.

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

Example 526

A DNA sequence (GBSx0564) was identified in *S.agalactiae* <SEQ ID 1679> which encodes the amino acid sequence <SEQ ID 1680>. This protein is predicted to be ATP-binding cassette protein PstB (pstB-2). Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2432(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10205> which encodes amino acid sequence <SEQ ID 10206> was also identified.

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

```
>GP:AAD22041 GB:AF118229 ATP-binding cassette protein PstB
          [Streptococcus pneumoniae]
 Identities = 166/245 (67%), Positives = 211/245 (85%), Gaps = 1/245 (0%)

Query: 10  INNLDLYYGEFHALKDVNLDIEEKEITAFIGPSGCGKSTLLKSINRMNDLVKNCKITGDI 69
           + +LDL+YG+F ALK++++ + E++ITA IGPSGCGKST LK++NRMNDLV +C I G +
Sbjct: 6   VRHLDLFYGDFQALKNISIQLPERQITALIGPSGCGKSTFLKTLNRMNDLVPSCHIEGQV 65

Query: 70  TLEGEDVYR-QLDINQLRKKVGMVFQKPNPFPMSIYDNVAFGPRTHGIHSKAELDDIVER 128
            L+ +D+Y  + ++NQLRK+VGMVFQ+PNPF MSIYDNVA+GPRTHGI  K +LD +VE+
Sbjct: 66  LLDEQDIYSSKFNLNQLRKRVGMVFQQPNPFAMSIYDNVAYGPRTHGIRDKKQLDALVEK 125

Query: 129 SLKQAALWDEVKDRLHKSALGMSGGQQQRLCIARALAIEPDVLLMDEPTSALDPISTAKI 188
           SLK AA+W+EVKD L KSA+ +SGGQQQRLCIARALA+EPD+LLMDEPTSALDPIST KI
Sbjct: 126 SLKGAAIWEEVKDDLKKSAMSLSGGQQQRLCIARALAVEPDILLMDEPTSALDPISTLKI 185

Query: 189 EELVIQLKKNYTIVIVTHNMQQAVRISDKTAFFLMGEVVEYNKTSQLFSLPQDERTENYI 248
           E+L+ QLKK+YTI+IVTHNMQQA RISDKTAFFL GE+ E+  T  +F+ P+D+RTE+YI
Sbjct: 186 EDLIQQLKKDYTIIIVTHNMQQASRISDKTAFFLTGEICEFGDTVDVFTNPKDQRTEDYI 245

Query: 249 TGRFG 253
           +GRFG
Sbjct: 246 SGRFG 250
```

There is also homology to SEQ ID 1682.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 527

A DNA sequence (GBSx0565) was identified in *S.agalactiae* <SEQ ID 1683> which encodes the amino acid sequence <SEQ ID 1684>. This protein is predicted to be transmembrane protein PstA (pstA-2). Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood =-13.11   Transmembrane  265 - 281 ( 255 - 286)
    INTEGRAL    Likelihood = -8.81   Transmembrane   79 -  95 (  68 - 100)
    INTEGRAL    Likelihood = -4.78   Transmembrane  195 - 211 ( 192 - 213)
    INTEGRAL    Likelihood = -4.67   Transmembrane  147 - 163 ( 143 - 164)
    INTEGRAL    Likelihood = -2.92   Transmembrane  122 - 138 ( 120 - 138)
    INTEGRAL    Likelihood = -0.90   Transmembrane   40 -  56 (  39 -  56)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.6243(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 GENPEPT database:

```
>GP:AAD22040 GB:AF118229 transmembrane protein PstA [Streptococcus pneumoniae]
 Identities = 135/263 (51%), Positives = 203/263 (76%)

Query: 23  FFLFAIVYLGAILSFATIAFVVIYILVKGLPHVNTGLFAWTYNTQNVSLLPAFINTIFII 82
           + L  +VY  + L+F ++  ++ +IL+KGLPH++  LF+WTY ++N+SL+PA I+T+ ++
Sbjct: 4   YLLKLLVYCFSALTFGSLFLIIGFILIKGLPHLSLSLFSWTYTSENISLMPAIISTVILV 63

Query: 83  ALTLLFAVPLGIGGSIYLTEYARRDNPYLKIIRVATETLAGIPSIIYGLFGALFFVKYTH 142
              LL A+P+GI    YL EY ++D+  +KI+R+A++TL+GIPSI++GLFG LFFV +
Sbjct: 64  FGALLLALPIGIFAGFYLVEYTKKDSLCVKIMRLASDTLSGIPSIVFGLFGMLFFVVFLG 123

Query: 143 LGLSLISGSLTLSIMILPLIMRTTEEALLSVPDSYREGAFALGAGKLRTIFKIVLPSAMS 202
              SL+SG LT  IM+LP+I+R+TEEALLSV DS R+ ++ LGAGKLRT+F+IVLP AM
Sbjct: 124 FQYSLLSGILTSVIMVLPVIIRSTEEALLSVSDSMRQASYGLGAGKLRTVFRIVLPVAMP 183

Query: 203 GIFAGIILAVGRIIGESAALIFTAGTVAKVAHSVFSSSRTLAVHMYAISGEGLYVDQTYA 262
           GI AG+ILA+GRI+GE+AAL++T GT       S+ SS R+LA+HMY +S EGL+V++ YA
Sbjct: 184 GILAGVILAIGRIVGETAALMYTLGTSTNTPSSLMSSGRSLALHMYMLSSEGLHVNEAYA 243

Query: 263 TAVILLLLVIIVNFVSGLVAKRL 285
           T VIL++ V+++N +S L++++L
Sbjct: 244 TGVILIITVLMINTLSSLLSRKL 266
```

There is also homology to SEQ ID 1686.

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

Example 528

A DNA sequence (GBSx0566) was identified in *S.agalactiae* <SEQ ID 1687> which encodes the amino acid sequence <SEQ ID 1688>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2687(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 529

A DNA sequence (GBSx0567) was identified in *S.agalactiae* <SEQ ID 1689> which encodes the amino acid sequence <SEQ ID 1690>. This protein is predicted to be transmembrane protein PstC (pstC-2). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
```

```
    INTEGRAL    Likelihood =-10.67   Transmembrane  256 - 272 ( 251 - 279)
    INTEGRAL    Likelihood = -8.86   Transmembrane  141 - 157 ( 133 - 162)
    INTEGRAL    Likelihood = -4.99   Transmembrane  111 - 127 ( 109 - 132)
    INTEGRAL    Likelihood = -4.30   Transmembrane   76 -  92 (  72 -  95)
    INTEGRAL    Likelihood = -1.86   Transmembrane   25 -  41 (  24 -  42)
    INTEGRAL    Likelihood = -1.33   Transmembrane   59 -  75 (  59 -  75)
    INTEGRAL    Likelihood = -0.27   Transmembrane  203 - 219 ( 202 - 219)

----- Final Results -----
                bacterial membrane --- Certainty=0.5267(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 GENPEPT database:

```
>GP:AAD22039 GB:AF118229 transmembrane protein PstC [Streptococcus pneumoniae]
 Identities = 162/266 (60%), Positives = 212/266 (78%), Gaps = 3/266 (1%)

Query: 15  ITACVSVISAILICLFLFSSGLPAITKIGWGNFIFGKVWHPSN--NIFGIFPMIVGSLYV 72
           ++A V+V++ +LIC F+FS+GLP I   G+  F+ G  W P+N    +GI PMIVGSL +
Sbjct: 1   MSATVAVVAILLICFFIFSNGLPFIANYGFARFLLGSDWSPTNIPASYGILPMIVGSLLI 60

Query: 73  TAGALLLGGPIGILTAVFMAYFCPENIYKPLKSAINLMAGIPSVVYGFFGLVVIVPMIRQ 132
           T GA+++G P GILT+VFM Y+CP+ +Y  LKSAINLMA IPS+VYGFFGL ++VP IR
Sbjct: 61  TLGAIVIGVPTGILTSVFMVYYCPKPVYGFLKSAINLMAAIPSIVYGFFGLQLLVPWIRS 120

Query: 133 YIGGFGMGVLAASILLGIMILPTIVSISESSLRAVPESYYEGGIALGASHERSVFFAVLP 192
           ++G  GM VL AS+LLGIMILPTI+S+SES++R VP++YY G +ALGASHERS+F  +LP
Sbjct: 121 FLGN-GMSVLTASLLLGIMILPTIISLSESAIRTVPKTYYSGSLALGASHERSIFSVILP 179

Query: 193 AAKRGILASVVLGIGRAIGETMAVIMVAGNQAVLPQSLTSGVRTLTTNIVMEMGYSSGLH 252
           AA+ GIL++V+LGIGRA+GETMAVI+VAGNQ ++P  L SG RTLTTNIV+EM Y+SG H
Sbjct: 180 AARSGILSAVILGIGRAVGETMAVILVAGNQPIIPSGLFSGTRTLTTNIVLEMAYASGQH 239

Query: 253 RQALIGTAVVLFIFILMINISFSALQ 278
           R+ALI T+ VLF  IL+IN  F+ L+
Sbjct: 240 REALIATSAVLFFLILLINAYFAYLK 265
```

There is also homology to SEQ ID 1692.

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

Example 530

A DNA sequence (GBSx0568) was identified in *S.agalactiae* <SEQ ID 1693> which encodes the amino acid sequence <SEQ ID 1694>. This protein is predicted to be probable hemolysin precursor (pstS). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> 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>
```

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

```
>GP:AAD22038 GB:AF118229 phosphate binding protein PstS
           [Streptococcus pneumoniae]
 Identities = 134/295 (45%), Positives = 185/295 (62%), Gaps = 9/295 (3%)

Query: 1   MKKHKMLSLLAVSGLMGIGILAGCSNDSSSSSK---GTINIVSREEGSGTRGAFIELFGI 57
           MK  KML+L A+ GL G G++A C N S++S +    GTI ++SRE GSGTRGAF E+ GI
```

```
Sbjct: 1   MKFKKMLTLAAI-GLSGFGLVA-CGNQSAASKQSASGTIEVISRENGSGTRGAFTEITGI 58

Query: 58  ESKNKKGEKVDHTSDAATVTNSTSVMLTTVSKDPSAIGYSSLGSLNSSVKVLKIDGKNAT 117
             K+   +K+D+T+  A + NST  +L+ V  + +AIGY SLGSL  SVK L+IDG  A+
Sbjct: 59  LKKDGD-KKIDNTAKTAVIQNSTEGVLSAVQGNANAIGYISLGSLTKSVKALEIDGVKAS 117

Query: 118 VKDIKSGSYKISRPFNIVTKEGKEKEATKDFIDYILSKDGQAVVEKNGYIPL-DNAKAYQ 176
               + G Y + RPFNIV      K   +DFI +I SK GQ VV  N +I        Y
Sbjct: 118 RDTVLDGEYPLQRPFNIVWSSNLSK-LGQDFISFIHSKQGQQVVTDNKFIEAKTETTEYT 176

Query: 177 AKVSSGKVVIAGSSSVTPVMEKIKEAYHKVNAKVDVEIQQSDSSTGITSAIDGSADIGMA 236
           ++  SGK+ + GS+SV+ +MEK+ EAY K N +V ++I  + SS GIT+  + +ADIGM
Sbjct: 177 SQHLSGKLSVVGSTSVSSLMEKLAEAYKKENPEVTIDITSNGSSAGITAVKEKTADIGMV 236

Query: 237 SRELDKTESSKGVKATVIATDGIAVVVNKKNKVNDLSTKQVKDIFTGKTTSWSDL 291
           SREL   E  K +     IA DGIAVVVN  NK + +S  ++ D+F+GK T+W  +
Sbjct: 237 SREL-TPEEGKSLTHDAIALDGIAVVVNNDNKASQVSMAELADVFSGKLTTWDKI 290
```

There is also homology to SEQ ID 1696.

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

A related GBS gene <SEQ ID 8597> and protein <SEQ ID 8598> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: 23   Crend: 4
McG: Discrim Score:      7.91
GvH: Signal Score (-7.5): -3.72
     Possible site: 34
>>> May be a lipoprotein
ALOM program   count: 0 value:   2.44 threshold:  0.0
   PERIPHERAL  Likelihood =  2.44     248
 modified ALOM score:  -0.99

*** Reasoning Step: 3

----- 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>
```

SEQ ID 1694 (GBS24) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 9; MW 33kDa).

GBS24-His was purified as shown in Figure 194, lane 10.

Example 531

A DNA sequence (GBSx0569) was identified in *S.agalactiae* <SEQ ID 1697> which encodes the amino acid sequence <SEQ ID 1698>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1725(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 532

A DNA sequence (GBSx0570) was identified in *S.agalactiae* <SEQ ID 1699> which encodes the amino acid sequence <SEQ ID 1700>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2741(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 GENPEPT database:

```
>GP:BAB05069 GB:AP001511 unknown conserved protein [Bacillus halodurans]
 Identities = 119/250 (47%), Positives = 149/250 (59%), Gaps = 9/250 (3%)

Query: 1   MQQYFVNGE--AGAYVTIEDKDTIKHMFNVMRLTEDDQVVLVFDDAIKRLAKVVDSSAHR 58
           MQ+YFV  E     YVTI   D +KH+  VMR+T  D+  L+  D   R  +   A+
Sbjct: 1   MQRYFVPKEQMTDTYVTITGDD-VKHIIKVMRMTIGDE--LICSDGHGRTVRCEIEKAND 57

Query: 59  FQIL----EELDNNVEMPVQVTIASGFPKGDKLDFVTQKATELGAAAIWGFPADWSVVKW 114
            ++L    E L  N E+P++VTIA   PKGDKLD++ QK TELGA A W F A  S+VKW
Sbjct: 58  SEVLARVIEPLIPNTELPIRVTIAQALPKGDKLDYIVQKGTELGAQAFWPFSASRSIVKW 117

Query: 115 DGKKLAKKEDKLAKIALGAAEQSKRNRLPQVRLFEKKADFQAELAGFDKIFIAYEESAKE 174
           D KK  KK ++L KIA  AAEQS R R+P +      +    E++GF K  +AYEE AKE
Sbjct: 118 DEKKGRKKTERLMKIAKEAAEQSYRERIPSIETPLAFSKLLQEISGFTKTIVAYEEEAKE 177

Query: 175 GELSALAQNLQTVKAGDKLLFIFGPEGGISPKEIAAFEEVGAIKVGLGPRIMRTETAPLY 234
           G L   A  L  +  GD LL I GPEGG + +EI A +  G    GLGPRI+RTETA LY
Sbjct: 178 GRLMTFAACLNELHHGDSLLVIIGPEGGFTTEEIDAIQRAGGAPAGLGPRILRTETASLY 237

Query: 235 ALSVISYSAE 244
           AL+ ISY  E
Sbjct: 238 ALAAISYHFE 247
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1701> which encodes the amino acid sequence <SEQ ID 1702>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2274(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 173/245 (70%), Positives = 202/245 (81%)

Query: 1   MQQYFVNGEAGAYVTIEDKDTIKHMFNVMRLTEDDQVVLVFDDAIKRLAKVVDSSAHRFQ 60
           MQQYF+ G+A   VTI DKDTIKHMF VMRL ++ +VVLVFDD +K LAKV +S AH  +
Sbjct: 1   MQQYFIKGKAEKKVTITDKDTIKHMFQVMRLADEAEVVLVFDDGVKYLAKVTNSMAHELE 60

Query: 61  ILEELDNNVEMPVQVTIASGFPKGDKLDFVTQKATELGAAAIWGFPADWSVVKWDGKKLA 120
           I+E L + VE+PV+VTIASGFPKGDKLD + QK TELGA+A+WG+PADWSVVKWDGKKLA
Sbjct: 61  IIEALPDQVELPVKVTIASGFPKGDKLDTIAQKVTELGASALWGYPADWSVVKWDGKKLA 120

Query: 121 KKEDKLAKIALGAAEQSKRNRLPQVRLFEKKADFQAELAGFDKIFIAYEESAKEGELSAL 180
```

```
            KKEDKLAKI LGAAEQSKRNR+P+V LFE KA+F   L+ FD IFIAYEE+AK G+L+ L
Sbjct: 121 KKEDKLAKIVLGAAEQSKRNRVPEVHLFEHKAEFLKSLSSFDHIFIAYEETAKAGQLATL 180

Query: 181 AQNLQTVKAGDKLLFIFGPEGGISPKEIAAFEEVGAIKVGLGPRIMRTETAPLYALSVIS 240
           A+ ++ VK G K+LFIFGPEGGISP EI  FE   AIKVGLGPRIMR ETAPLYALS +S
Sbjct: 181 AREVKEVKPGAKILFIFGPEGGISPTEITQFEAASAIKVGLGPRIMRAETAPLYALSALS 240

Query: 241 YSAEL 245
           Y+ EL
Sbjct: 241 YALEL 245
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 533

A DNA sequence (GBSx0571) was identified in *S.agalactiae* <SEQ ID 1703> which encodes the amino acid sequence <SEQ ID 1704>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.28   Transmembrane  238 - 254 ( 237 - 254)

----- 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>
```

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

```
>GP:BAA82791 GB:AB023064 orf35 [Listeria monocytogenes]
 Identities = 138/309 (44%), Positives = 193/309 (61%), Gaps = 5/309 (1%)

Query: 4   WNELTVHVNREAEEAVSNLLIETGSQGVAISDSADYLGQ-EDRFGELYP---EVEQSDMI 59
           W+E+ VH   EA E V+N+L E G+ GV+I D AD+L + ED+FGE+Y    E    D +
Sbjct: 3   WSEVEVHTTNEAVEPVANVLTEFGAAGVSIEDVADFLRERERDKFGEIYALRREDYPEDGV 62

Query: 60  AITAYYPDTLDIEAVKADLADRLANFEGFGLATGSVNLDSQELVEEDWADNWKKYYEPAR 119
            I AY+  T +      ++   L N   F +  G       ++ +E+WA  WKKYY P +
Sbjct: 63  IIKAYFLKTTEFVEQIPEIEQTLKNLSTFDIPLGKFQFVVNDVDDEEWATAWKKYYHPVQ 122

Query: 120 ITHDLTIVPSWTDYEAKAGEKIIKMDPGMAFGTGTHPTTKMSLFALEQVLRGGETVIDVG 179
           IT  +TIVPSW  Y   A E II++DPGMAFGTGTHPTT++ + AL   L+ G+ VIDVG
Sbjct: 123 ITDRITIVPSWESYTPSANEIIIELDPGMAFGTGTHPTTQLCIRALSNYLQPGDEVIDVG 182

Query: 180 TGSGVLSIASSLLGAKDIYAYDLDDVAVRVAQENIDMNPGTENIHVAAGDLLKGVQQ-EV 238
           TGSGVLSIAS+ LGAK I A DLD++A R A+ENI +N      I V   +LL+ + +  V
Sbjct: 183 TGSGVLSIASAKLGAKSILATDLDEIATRAAEENITLNKTEHIITVKQNNLLQDINKTNV 242

Query: 239 DVIVANILADILIHLTDDAYRLVKDEGYLIMSGIISEKWDMVRESAEKAGFFLETHMVQG 298
           D++VANILA++++   +D Y+ +K  G  I SGII +K  +V E+ + AG  +E    QG
Sbjct: 243 DIVVANILAEVILLFPEDVYKALKPGGVFIASGIIEDKAKVVEEALKNAGLIIEKMEQQG 302

Query: 299 EWNACVFKK 307
           +W A + K+
Sbjct: 303 DWVAIISKR 311
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1705> which encodes the amino acid sequence <SEQ ID 1706>. Analysis of this protein sequence reveals the following:

```
     Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.57   Transmembrane  238 - 254 ( 237 - 257)

----- 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>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAA82791 GB:AB023064 orf35 [Listeria monocytogenes]
 Identities = 139/309 (44%), Positives = 203/309 (64%), Gaps = 5/309 (1%)

Query: 4   WQEVTVHVHRDAQEAVSHVLIETGSQGVAIADSADYIGQK-DRFGELYP---DVEQSDMI 59
           W EV VH   +A E V++VL E G+ GV+I D AD++ ++ D+FGE+Y    +    D +
Sbjct: 3   WSEVEVHTTNEAVEPVANVLTEFGAAGVSIEDVADFLREREDKFGEIYALRREDYPEDGV 62

Query: 60  AITAYYPSSTNLADIIATINEQLAELASFGLQVGQVTVDSQELAEEDWADNWKKYYEPAR 119
            I AY+  +T   + I  I + L  L++F + +G+      ++ +E+WA  WKKYY P +
Sbjct: 63  IIKAYFLKTTEFVEQIPEIEQTLKNLSTFDIPLGKFQFVVNDVDDEEWATAWKKYYHPVQ 122

Query: 120 ITHDLTIVPSWTDYDASAGEKVIKLDPGMAFGTGTHPTTKMSLFALEQILRGGETVIDVG 179
           IT  +TIVPSW  Y  SA E +I+LDPGMAFGTGTHPTT++ + AL   L+ G+ VIDVG
Sbjct: 123 ITDRITIVPSWESYTPSANEIIIELDPGMAFGTGTHPTTQLCIRALSNYLQPGDEVIDVG 182

Query: 180 TGSGVLSIASSLLGAKTIYAYDLDDVAVRVAQDNIDLNQGTDNIHVAAGDLLKGVSQ-EA 238
           TGSGVLSIAS+ LGAK+I A DLD++A R A++NI LN+    I V   +LL+ +++
Sbjct: 183 TGSGVLSIASAKLGAKSILATDLDEIATRAAEENITLNKTEHIITVKQNNLLQDINKTNV 242

Query: 239 DVIVANILADILVLLTDDAYRLVKKEGYLILSGIISEKLDMVLEAAFSAGFFLETHMVQG 298
           D++VANILA++++L  +D Y+ +K  G  I SGII +K  +V EA  +AG  +E    QG
Sbjct: 243 DIVVANILAEVILLFPEDVYKALKPGGVFIASGIIEDKAKVVEEALKNAGLIIEKMEQQG 302

Query: 299 EWNALVFKK 307
           +W A++ K+
Sbjct: 303 DWVAIISKR 311
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 259/317 (81%), Positives = 287/317 (89%)

Query: 1   MNTWNELTVHVNREAEEAVSNLLIETGSQGVAISDSADYLGQEDRFGELYPEVEQSDMIA 60
           M TW E+TVHV+R+A+EAVS++LIETGSQGVAI+DSADY+GQ+DRFGELYP+VEQSDMIA
Sbjct: 1   METWQEVTVHVHRDAQEAVSHVLIETGSQGVAIADSADYIGQKDRFGELYPDVEQSDMIA 60

Query: 61  ITAYYPDTLDIEAVKADLADRLANFEGFGLATGSVNLDSQELVEEDWADNWKKYYEPARI 120
           ITAYYP + ++  + A + ++LA    FGL  G V +DSQEL EEDWADNWKKYYEPARI
Sbjct: 61  ITAYYPSSTNLADIIATINEQLAELASFGLQVGQVTVDSQELAEEDWADNWKKYYEPARI 120

Query: 121 THDLTIVPSWTDYEAKAGEKIIKMDPGMAFGTGTHPTTKMSLFALEQVLRGGETVIDVGT 180
           THDLTIVPSWTDY+A AGEK+IK+DPGMAFGTGTHPTTKMSLFALEQ+LRGGETVIDVGT
Sbjct: 121 THDLTIVPSWTDYDASAGEKVIKLDPGMAFGTGTHPTTKMSLFALEQILRGGETVIDVGT 180

Query: 181 GSGVLSIASSLLGAKDIYAYDLDDVAVRVAQENIDMNPGTENIHVAAGDLLKGVQQEVDV 240
           GSGVLSIASSLLGAK IYAYDLDDVAVRVAQ+NID+N GT+NIHVAAGDLLKGV QE DV
Sbjct: 181 GSGVLSIASSLLGAKTIYAYDLDDVAVRVAQDNIDLNQGTDNIHVAAGDLLKGVSQEADV 240

Query: 241 IVANILADILIHLTDDAYRLVKDEGYLIMSGIISEKWDMVRESAEKAGFFLETHMVQGEW 300
           IVANILADIL+ LTDDAYRLVK EGYLI+SGIISEK DMV E+A  AGFFLETHMVQGEW
Sbjct: 241 IVANILADILVLLTDDAYRLVKKEGYLILSGIISEKLDMVLEAAFSAGFFLETHMVQGEW 300

Query: 301 NACVFKKTDDISGVIGG 317
           NA VFKKTDDISGVIGG
Sbjct: 301 NALVFKKTDDISGVIGG 317
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 534

A DNA sequence (GBSx0572) was identified in *S.agalactiae* <SEQ ID 1707> which encodes the amino acid sequence <SEQ ID 1708>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4198(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 535

A DNA sequence (GBSx0573) was identified in *S.agalactiae* <SEQ ID 1709> which encodes the amino acid sequence <SEQ ID 1710>. This protein is predicted to be transcriptional activator tipa. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0683(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 GENPEPT database:

```
>GP:CAB15677 GB:Z99122 transcriptional regulator [Bacillus subtilis]
 Identities = 87/246 (35%), Positives = 139/246 (56%), Gaps = 13/246 (5%)

Query: 4   VKEVSILSGVSVRTLHHYDKIGLFPPTALSEAGYRLYDDEALIRLQEILLFRELEFPLKD 63
           VK+V+ +SGVS+RTLHHYD I L  P+AL++AGYRLY D  L RLQ+IL F+E+ F L +
Sbjct: 5   VKQVAEISGVSIRTLHHYDNIELLNPSALTDAGYRLYSDADLERLQQILFFKEIGFRLDE 64

Query: 64  IKYLLEQAKEERQDLLAQQIKLLEWKRSHLEQVITHAKR--LQEKGDDYMN----FDVYN 117
           IK +L+    +R+  L  Q ++L  K+  ++++I    R  L   G + MN    F   +
Sbjct: 65  IKEMLDHPNFDRKAALQSQKEILMKKKQRMDEMIQTIDRTLLSVDGGETMNKRDLFAGLS 124

Query: 118 KTELEQLQA----EAKEKWGQTAA--YKEFAQKHASDDFAQISQEMAKIMVQFGQLKTQN 171
             ++E+ Q      E ++ +G+  A    ++     +++DD+  I  E   I  +
Sbjct: 125 MKDIEEHQQTYADEVRKLYGKEIAEETEKRTSAYSADDWRTIMAEFDSIYRRIAARMKHG 184

Query: 172 VSDESVQMCVKRLQDYISQNFYTCTNEILAGLGQMYQSDDRFSQSIDKAGGAGTSEFVSQ 231
             D  +Q  V   +D+I Q  Y CT +I  GLG++Y +D+RF+ SI++  G G + F+ +
Sbjct: 185 PDDAEIQAAVGAFRDHICQYHYDCTLDIFRGLGEVYITDERFTDSINQY-GEGLAAFLRE 243

Query: 232 AIAYYC 237
           AI  YC
Sbjct: 244 AIIIYC 249
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1711> which encodes the amino acid sequence <SEQ ID 1712>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.28   Transmembrane  146 - 162 ( 143 - 167)
   INTEGRAL    Likelihood = -2.92   Transmembrane  172 - 188 ( 171 - 190)

----- Final Results -----
               bacterial membrane --- Certainty=0.4312(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:

```
>GP:CAB15677 GB:Z99122 transcriptional regulator [Bacillus subtilis]
 Identities = 40/107 (37%), Positives = 69/107 (64%), Gaps = 6/107 (5%)

Query: 7   YSTGELANLAGVSIRTVQYYDQRGILIPTALTAGGRRLYTDSDLEQLRMICFLRDLGFSI 66
           Y   ++A ++GVSIRT+ +YD    +L P+ALT  G RLY+D+DLE+L+ I F +++GF +
Sbjct: 3   YQVKQVAEISGVSIRTLHHYDNIELLNPSALTDAGYRLYSDADLERLQQILFFKEIGFRL 62

Query: 67  EQIRKVLAEENAAQVLELLLVDHIATAKEDLAAKEQQVDIAVKILDR 113
           ++I+++L   N  +   L        + KE L  K+Q++D  ++ +DR
Sbjct: 63  DEIKEMLDHPNFDRKAAL------QSQKEILMKKKQRMDEMIQTIDR 103
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 40/133 (30%), Positives = 71/133 (53%), Gaps = 6/133 (4%)

Query: 6   EVSILSGVSVRTLHHYDKIGLFPPTALSEAGYRLYDDEALIRLQEILLFRELEFPLKDIK 65
           E++ L+GVS+RT+ +YD+ G+  PTAL+  G RLY D  L +L+ I   R+L F ++ I+
Sbjct: 11  ELANLAGVSIRTVQYYDQRGILIPTALTAGGRRLYTDSDLEQLRMICFLRDLGFSIEQIR 70

Query: 66  YLL--EQAKEERQDLLAQQIKL----LEWKRSHLEQVITHAKRLQEKGDDYMNFDVYNKT 119
           +L  E A +  + LL   I      L  K   ++  +     RL+++     ++F +
Sbjct: 71  KVLAEENAAQVLELLLVDHIATAKEDLAAKEQQVDIAVKILDRLRKQDPQSLDFLMDISL 130

Query: 120 ELEQLQAEAKEKW 132
           ++  +A  K +W
Sbjct: 131 SMKNQKAWKKLQW 143
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 536

A DNA sequence (GBSx0575) was identified in *S.agalactiae* <SEQ ID 1713> which encodes the amino acid sequence <SEQ ID 1714>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.06   Transmembrane   57 -  73 (  57 -  73)

----- 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>
```

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

```
>GP:CAB14586 GB:Z99117 yrkN [Bacillus subtilis]
 Identities = 38/136 (27%), Positives = 60/136 (43%), Gaps = 3/136 (2%)

Query: 2   ITLQKAEASDLEKIIA-IQRASFKAVYEKYHDQYDPYVEEVEQIRWKLVERPDCFYHFVL 60
           + L+ A+ SDL +    +Q A   AV E + D  D  +     ++ +    P    + +L
Sbjct: 9   VILELAKESDLPEFQKKLQEAFAIAVIETFGDCEDGPIPSDNDVQ-ESFNAPGAVVYHIL 67

Query: 61  VDETIVGFLRLVIKDEEKRAWLGTAAILPQYQGQGYGSAAMALLEKTYPKLTKWDLCTIA 120
```

```
            D   VG   + I  +     L    + P+Y  QG G +A   +E  YP    W+  T
Sbjct: 68   QDGKNVGGAVVRINSQTNHNSLDLFYVSPEYHSQGIGLSAWKAIEAQYPDTVLWETVTPY 127

Query: 121 QEKLMVSFY-EKCGYH 135
            EK  ++FY  KCG+H
Sbjct: 128 FEKRNINFYVNKCGFH 143
```

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

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

Example 537

A DNA sequence (GBSx0576) was identified in *S.agalactiae* <SEQ ID 1715> which encodes the amino acid sequence <SEQ ID 1716>. This protein is predicted to be Bacterial mutT protein. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2417(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 GENPEPT database:

```
>GP:AAG06568 GB:AE004742 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 57/131 (43%), Positives = 82/131 (62%)

Query: 10  FSGAKIALFCEGKILTSLRDDFPDLPYAGFWDLPGGGREDNETPLECLFREVDEELSLTL 69
           FSGAK+ALF     ++   RD+ P +P+ G+WD PGGGRE  ETP EC  RE++EE S+ L
Sbjct: 7   FSGAKLALFYGDHLVVYKRDEKPGIPFPGYWDFPGGGREGLETPAECALRELEEEFSIRL 66

Query: 70  TRNHIDWVKTYRGMLKPDKLSVFMVGHISQKEYDSIVLGDEGQDYKLMSIDEFLSHKKVI 129
               I+W + Y          + F+V  +  +E+++I  GDEGQ ++LM +D +L+H   +
Sbjct: 67  EEPRIEWQRQYPSTSGSAPFAYFLVARLEDREFEAIRFGDEGQYWRLMEVDAYLAHAMAV 126

Query: 130 PQLQERLRDYL 140
           P LQ RL DYL
Sbjct: 127 PYLQSRLGDYL 137
```

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

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

Example 538

A DNA sequence (GBSx0577) was identified in *S.agalactiae* <SEQ ID 1717> which encodes the amino acid sequence <SEQ ID 1718>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3299(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1719> which encodes the amino acid sequence <SEQ ID 1720>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5527(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 111/156 (71%), Positives = 128/156 (81%)

Query: 1   MAKFGFLSVLEEELDKHLQYDFAMDWDKKNHTVEVTFILEAQNSSAIETVDDQGETSSED 60
           MA +GFLSVLEEE+DKH QYD+AMDWDKKNH VEVTF+LEAQN  AI+T+DD GE + +D
Sbjct: 1   MATYGFLSVLEEEMDKHFQYDYAMDWDKKNHAVEVTFVLEAQNKEAIKTIDDSGEVTQDD 60

Query: 61  IVFEDYVLFYNPVKSRFDAEDYLVTIPYEPKKGLSREFLAYFAETLNEVATEGLSDLMDF 120
           IVFEDYVLFYNP KS+FDA DYLVTIP++ KKG SREFLAYFA+ LN+VA EG SDLMDF
Sbjct: 61  IVFEDYVLFYNPAKSQFDAADYLVTIPFDAKKGFSREFLAYFAQFLNDVAIEGHSDLMDF 120

Query: 121 LTDDSIEEFGLSWDTDAFENGRAELKETEFYPYPRY 156
           L DDS  +F L W+  AFE G+  L+E   YPYPRY
Sbjct: 121 LADDSKADFFLEWNAQAFEEGQQGLEEAASYPYPRY 156
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 539

A DNA sequence (GBSx0578) was identified in *S.agalactiae* <SEQ ID 1721> which encodes the amino acid sequence <SEQ ID 1722>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2846(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 GENPEPT database:

```
>GP:CAB51273 GB:AL096872 putative acetyltransferase [Streptomyces
          coelicolor A3(2)]
 Identities = 35/109 (32%), Positives = 62/109 (56%), Gaps = 1/109 (0%)

Query: 51  VAEVDDKIAGVLDFGPYYPFPAGKHVATF-GILIAEPYQGQGLGKALLKALLTEAKAQGY 109
           VAE+D  + G +  G   P  +  HV    G+ +A   +G G+G+AL++++ + EA+ +G+
Sbjct: 56  VAELDGAVVGYVRLGFPTPLASNTHVRQIRGLAVAGAARGHGVGRALVRAAVEEARHEGF 115

Query: 110 IKIAMHVMGNNSRAISLYQKYGFTEEARITKAFFIENHYVDALIFAKDL 158
            +I + V+G+N+ A  LY+  GF  E    + F ++  YVD ++  + L
Sbjct: 116 RRITLRVLGHNTAARGLYESEGFVVEGVQPEEFHLDGRYVDDVLMGQML 164
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1723> which encodes the amino acid sequence <SEQ ID 1724>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
bacterial cytoplasm --- Certainty=0.0229(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 34/108 (31%), Positives = 59/108 (54%), Gaps = 7/108 (6%)

Query: 35  TESDLEKNLANGMSFFV-----AEVDDKIAGVLDFGPYYPFPAGKHVATFGILIAEPYQG 89
           T  +L   L+   + F+     A +D+K+ G+L+         G+  A   +L+A+ Y+G
Sbjct: 43  TPQELSDFLSRSQTSFIDFCLLARLDEKVVGLLNLSGEV-LSQGQAEADVFMLVAKTYRG 101

Query: 90  QGLGKALLKALLTEAKAQGYIK-IAMHVMGNNSRAISLYQKYGFTEEA 136
            G+G+ LL+  L  A+   YI+ + + V   N++AI LY+KYGF  E+
Sbjct: 102 YGIGQLLLEIALDWAEENPYIESLKLDVQVRNTKAIYLYKKYGFRIES 149
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 540

A DNA sequence (GBSx0579) was identified in *S.agalactiae* <SEQ ID 1725> which encodes the amino acid sequence <SEQ ID 1726>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2056(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 GENPEPT database:

```
>GP:CAB14712 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 248/417 (59%), Positives = 314/417 (74%), Gaps = 4/417 (0%)

Query: 5   LALRMRPRNINEVIGQQHLVGNGKIIDRMVAANMLSSMILYGPPGIGKTSIASAIAGTTK 64
           LA RMRP  I ++IGQQHLV   KII RMV A  LSSMILYGPPGIGKTSIA+AIAG+T
Sbjct: 4   LAYRMRPTKIEDIIGQQHLVAEDKIIGRMVQAKHLSSMILYGPPGIGKTSIATAIAGSTS 63

Query: 65  YAFRTFNATVDSKKRLQEIAEEAKFSGGLVLLLDEIHRLDKTKQDFLLPLLENGNIIMIG 124
            AFR  NA +++KK ++ +A+EAK SG ++L+LDE+HRLDK KQDFLLP LENG II+IG
Sbjct: 64  IAFRKLNAVINNKKDMEIVAQEAKMSGQVILILDEVHRLDKGKQDFLLPYLENGMIILIG 123

Query: 125 ATTENPFFSVTPAIRSRVQIFELEPLSNEDIKKAIQLAISDKERGF-PFLVTIDDEALDF 183
           ATT NP+ ++ PAIRSR QIFELEPL+ E IK+A++ A+ D+ RG    + V+IDD+A++
Sbjct: 124 ATTANPYHAINPAIRSRTQIFELEPLTPELIKQALERALHDEHRGLGTYSVSIDDQAMEH 183

Query: 184 IVTATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMENSLQCSYITMDKNGDGHYDIL 243
                GD+RSA N+L+LAV+ST  + DG  HI+LET E  LQ     + DK+GD HYD+L
Sbjct: 184 FAHGCGGDVRSALNALELAVLSTKESADGEIHITLETAEECLQKKSFSHDKDGDAHYDVL 243

Query: 244 SALQKSIRGSDVNASLHYAARLVEAGDLPSLARRLTIIAYEDIGLANPEAQIHTVTALEA 303
           SA QKSIRGSD NA+LHY ARL+EAGDL S+ARRL +IAYEDIGLA+P+A    + A++
Sbjct: 244 SAFQKSIRGSDANAALHYLARLIEAGDLESIARRLLVIAYEDIGLASPQAGPRVLNAIQT 303

Query: 304 AQRIGFPEARILIANIVVDLALSPKSNSAYLAMDAALADLRRSGNLPIPRHLRDGHYSGS 363
           A+R+GFPEARI +AN V++L LSPKSNSA LA+D ALAD+R       +P+HL+D HY G+
Sbjct: 304 AERVGFPEARIPLANAVIELCLSPKSNSAILAIDEALADIRAGKIGDVPKHLKDAHYKGA 363

Query: 364 KTLGNARDYKYPHAYPEKWVKQQYLPDKLVGHNYFEANETGKYERALGSNKERIDKL 420
           + LG   DYKYPH Y   WV+QQYLPD L    Y++  +TGK+E AL   K+  DKL
Sbjct: 364 QELGRGIDYKYPHNYDNGWVEQQYLPDPLKNKQYYKPKQTGKFESAL---KQVYDKL 417
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1727> which encodes the amino acid sequence <SEQ ID 1728>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2374(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 394/422 (93%), Positives = 409/422 (96%)

Query: 1   MADNLALRMRPRNINEVIGQQHLVGNGKIIDRMVAANMLSSMILYGPPGIGKTSIASAIA 60
           M D+LALRMRP+ I+EVIGQ+HLVG GKII RMV AN LSSMILYGPPGIGKTSIASAIA
Sbjct: 1   MPDHLALRMRPKTISEVIGQKHLVGEGKIIRRMVEANRLSSMILYGPPGIGKTSIASAIA 60

Query: 61  GTTKYAFRTFNATVDSKKRLQEIAEEAKFSGGLVLLLDEIHRLDKTKQDFLLPLLENGNI 120
           GTT+YAFRTFNAT+DSKKRLQEIAEEAKFSGGLVLLLDEIHRLDKTKQDFLLPLLENG I
Sbjct: 61  GTTRYAFRTFNATIDSKKRLQEIAEEAKFSGGLVLLLDEIHRLDKTKQDFLLPLLENGTI 120

Query: 121 IMIGATTENPFFSVTPAIRSRVQIFELEPLSNEDIKKAIQLAISDKERGFPFLVTIDDEA 180
           IMIGATTENPFFSVTPAIRSRVQIFELEPLSNEDIK AIQLAISDKERGFPFLVTIDDEA
Sbjct: 121 IMIGATTENPFFSVTPAIRSRVQIFELEPLSNEDIKTAIQLAISDKERGFPFLVTIDDEA 180

Query: 181 LDFIVTATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMENSLQCSYITMDKNGDGHY 240
           LDFIVTATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMENSLQ SYITMDKNGDGHY
Sbjct: 181 LDFIVTATNGDLRSAYNSLDLAVMSTSPNEDGSSRHISLETMENSLQRSYITMDKNGDGHY 240

Query: 241 DILSALQKSIRGSDVNASLHYAARLVEAGDLPSLARRLTIIAYEDIGLANPEAQIHTVTA 300
           D+LSALQKSIRGSDVNASLHYAARLVEAGDLPSLARRLTIIAYEDIGLANP+AQ+HTVTA
Sbjct: 241 DVLSALQKSIRGSDVNASLHYAARLVEAGDLPSLARRLTIIAYEDIGLANPDAQVHTVTA 300

Query: 301 LEAAQRIGFPEARILIANIVVDLALSPKSNSAYLAMDAALADLRRSGNLPIPRHLRDGHY 360
           L+AAQRIGFPEARI IAN+V+DLALSPKSNSAYLAMDAALADLR SGNLPIPRHLRDGHY
Sbjct: 301 LDAAQRIGFPEARIPIANVVIDLALSPKSNSAYLAMDAALADLRTSGNLPIPRHLRDGHY 360

Query: 361 SGSKTLGNARDYKYPHAYPEKWVKQQYLPDKLVGHNYFEANETGKYERALGSNKERIDKL 420
           +GSK LGNA+DY YPHAYPEKWVKQQYLPDKLVGH+YFEANETGKYERALGSNKERIDKL
Sbjct: 361 AGSKDLGNAKDYLYPHAYPEKWVKQQYLPDKLVGHHYFEANETGKYERALGSNKERIDKL 420

Query: 421 SD 422
           SD
Sbjct: 421 SD 422
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 541

A DNA sequence (GBSx0580) was identified in *S.agalactiae* <SEQ ID 1729> which encodes the amino acid sequence <SEQ ID 1730>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2991(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10207> which encodes amino acid sequence <SEQ ID 10208> was also identified.

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

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 542

A DNA sequence (GBSx0581) was identified in *S.agalactiae* <SEQ ID 1731> which encodes the amino acid sequence <SEQ ID 1732>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2402(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 543

A DNA sequence (GBSx0582) was identified in *S.agalactiae* <SEQ ID 1733> which encodes the amino acid sequence <SEQ ID 1734>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.40   Transmembrane  231 - 247 ( 225 - 250)
   INTEGRAL    Likelihood = -9.92   Transmembrane  159 - 175 ( 151 - 179)
   INTEGRAL    Likelihood = -9.08   Transmembrane   21 -  37 (  18 -  43)
   INTEGRAL    Likelihood = -9.08   Transmembrane  181 - 197 ( 176 - 201)
   INTEGRAL    Likelihood = -3.35   Transmembrane  111 - 127 ( 110 - 130)
   INTEGRAL    Likelihood = -2.81   Transmembrane   74 -  90 (  74 -  93)

----- 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>
```

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

```
>GP:CAB15891 GB:Z99123 yxlG [Bacillus subtilis]
 Identities = 54/203 (26%), Positives = 100/203 (48%), Gaps = 7/203 (3%)

Query: 1   MTGLIPMLKKEWLENSRSHKALALLLISIIFGILGPLTALLMPEIMA--GILPKKLQEAI 58
            M  ++ +L+KEWLE  +S K + L +  +I G+  PLT   MPEI+A  G LP  ++ +
Sbjct: 1   MKVMMALLQKEWLEGWKSGKLIWLPIAMMIVGLTQPLTIYYMPEIIAHGGNLPDGMKISF 60

Query: 59  PDPTYLDSYSQYFKNINQLGLILLVFLFSGSLTQEFTRGTLINLITKGLSKKAIILAKFI 118
              P+  +          N LG+ L++F   GS+  E  +G     ++++ ++   I++K++
Sbjct: 61  TMPSGSSEVMVSTLSQFNTLGMALVIFSVMGSVANERNQGVTALIMSRPVTAAHYIVSKWL 120
```

```
Query: 119 MMTLIWSISYILGSLTQYAYTLYYFNNHGQHKLIV-YGTSWIFGLLLLSLILFYSVIFRK 177
           + ++I  +S+  G    Y Y    F +    +     G   ++ + +++  L  S IFR
Sbjct: 121 IQSVIGIMSFAAGYGLAYYYVRLLFEDASFSRFAASLGLYALWVIFIVTAGLAGSTIFR- 179

Query: 178 TAGVLIAC---LMTIVAFFISGF 197
           + G   AC   L   V+F +  F
Sbjct: 180 SVGAAAACGIGLTAAVSFAVHYF 202
```

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

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

Example 544

A DNA sequence (GBSx0583) was identified in *S.agalactiae* <SEQ ID 1735> which encodes the amino acid sequence <SEQ ID 1736>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1344(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 GENPEPT database:

```
>GP:CAB15892 GB:Z99123 similar to ABC transporter (ATP-binding
          protein) [Bacillus subtilis]
 Identities = 116/303 (38%), Positives = 175/303 (57%), Gaps = 18/303 (5%)

Query: 4   ISLQNLSKSFGDQIILNQVSLELEENKIYGFVGPNGAGKTTTIKMILGLLKVDSGTISVM 63
           +S+++L KS+     +  VS  + EN+    +GPNGAGKTTT++M+ GLL   SGTI ++
Sbjct: 2   LSIESLCKSYRHHEAVKNVSFHVNENECVALLGPNGAGKTTTLQMLAGLLSPTSGTIKLL 61

Query: 64  GNPVTFGQTKSNQVIGYLPDVPEFYDYMTAQEYLQLC---AGLAQNKTSLPIADLLEQVG 120
           G      +    ++IGYLP  P FY +MTA E+L       +GL++ K    I ++LE VG
Sbjct: 62  GE-----KKLDRRLIGYLPQYPAFYSWMTANEFLTFAGRLSGLSKRKCQEKIGEMLEFVG 116

Query: 121 LADN-QQRISTYSRGMKQRLGLAQALIHNPKILICDEPTSALDPQGRQEILSIISQLRGQ 179
           L +   +RI  YS GMKQRLGLAQAL+H PK LI DEP SALDP GR E+L ++ +L+
Sbjct: 117 LHEAAHKRIGGYSGGMKQRLGLAQALLHKPKFLILDEPVSALDPTGRFEVLDMMRELKKH 176

Query: 180 KTVIFSTHILSDVEKVCDQVLILTKSGIH---NLEDLRDKASASVNQLNLLIKVSDNEAQ 236
             V+FSTH+L D E+VCDQV+I+     I      L++L+ +   +V  L++  K+     +
Sbjct: 177 MAVLFSTHVLHDAEQVCDQVVIMKNGEISWKGELQELKQQQQTNVFTLSVKEKLEGWLEE 236

Query: 237 KLALRFPLNQKDQYYKVHLELSEANNREQALASFYRYLVEQEITPYFIELLEDSLEDFYL 296
           K  +   + +    EL + +     L+       + + +T    E   +SLED YL
Sbjct: 237 KPYVSAIVYKNPS--QAVFELPDIHAGRSLLSD----CIRKGLTVTRFEQKTESLEDVYL 290

Query: 297 EVI 299
           +V+
Sbjct: 291 KVV 293
```

There is also homology to SEQ ID 686.

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

Example 545

A DNA sequence (GBSx0584) was identified in *S.agalactiae* <SEQ ID 1737> which encodes the amino acid sequence <SEQ ID 1738>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4383(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 GENPEPT database:

```
>GP:AAB71491 GB:U53767 ORF6 [Bacillus pumilus]
 Identities = 25/60 (41%), Positives = 41/60 (67%)

Query: 2  IGDTILFERTRLGMTQEKLSDYLHLTKATISKWENNQAKPDIDYLILMAKLFDMTLDELV 61
          +G  I  +R  L ++QE +++ L +++  ISKWE NQ++P +D LI +A+LFD  + ELV
Sbjct: 4  LGSNISNKRKSLKLSQEYVAEQLGVSRQAISKWETNQSEPSMDNLIRLAELFDSDIKELV 63
```

There is also homology to SEQ ID 1740.

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

Example 546

A DNA sequence (GBSx0585) was identified in *S.agalactiae* <SEQ ID 1741> which encodes the amino acid sequence <SEQ ID 1742>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4241(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 GENPEPT database:

```
>GP:CAB15470 GB:Z99121 yvdC [Bacillus subtilis]
 Identities = 59/104 (56%), Positives = 76/104 (72%)

Query: 1   MDITAYQKWVSEFYKKRNWYQYNSFIRSNFLCEEVGELAQAIRKYEIGRDRPDEIEKSNN 60
           M +   +KW+ EFY+KR W +Y  FIR  FL EE GELA+A+R YEIGRDRPDE E S
Sbjct: 1   MQLADAEKWMKEFYEKRGWTEYGPFIRVGFLMEEAGELARAVRAYEIGRDRPDEKESSRA 60

Query: 61  ENLNDIKEELGDVLDNIFILADQYNISLEEIIEAHKNKLEKRFE 104
           E   ++ EE+GDV+ NI ILAD Y +SLE++++AH+ KL KRFE
Sbjct: 61  EQKQELIEEMGDVIGNIAILADMYGVSLEDVMKAHQEKLTKRFE 104
```

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

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

Example 547

A DNA sequence (GBSx0586) was identified in *S.agalactiae* <SEQ ID 1743> which encodes the amino acid sequence <SEQ ID 1744>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0453(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 GENPEPT database:

```
>GP:BAB06803 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 87/187 (46%), Positives = 125/187 (66%)

Query: 1   MKITVFCGASNGNNPIYSQKIVELGEWMIKNNHDLVYGGGKVGLMGVIADTVINNGGQAI 60
           MKI VFCG+SNG + +Y +    +LG+ + +     LVYGG  VG+MG +AD+V+  GG+ I
Sbjct: 1   MKIAVFCGSSNGASDVYKEGARQLGKELARRGITLVYGGASVGIMGAVADSVLEAGGEVI 60

Query: 61  GVIPTFLKDREIAHTNLSKLIVVENMPQRKGKMMSLGEAYIALPGGPGTLEEISEVISWS 120
           GV+P FL++ EI+H +L+KLIVVE M +RK KM  L + ++ALPGGPGTLEE  E+ +W+
Sbjct: 61  GVMPRFLEEPEISHPHLTKLIVVETMHERKAKMAELADGFLALPGGPGTLEEFFEIFTWA 120

Query: 121 RIGQNDSPCILYNINGYFNHLESMFDHMVSEGFLSQNDRNNVLFSDDIIEIEKFIKDYQS 180
           +IG +  PC L NIN YF+ L ++   HM +E FL +  R+  L   D I +      Y+
Sbjct: 121 QIGLHQKPCGLLNINHYFDPLVTLLHHMSNEQFLHEKYRSMALVHTDPILLLDQFSTYEP 180

Query: 181 PTIRKYS 187
           PT++ YS
Sbjct: 181 PTVKAYS 187
```

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

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

Example 548

A DNA sequence (GBSx0587) was identified in *S.agalactiae* <SEQ ID 1745> which encodes the amino acid sequence <SEQ ID 1746>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

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

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

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

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

Example 549

A DNA sequence (GBSx0588) was identified in *S.agalactiae* <SEQ ID 1747> which encodes the amino acid sequence <SEQ ID 1748>. This protein is predicted to be integrase. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3685(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 GENPEPT database:

```
>GP:AAF12706 GB:AF066865 integrase [bacteriophage TPW22]
 Identities = 106/377 (28%), Positives = 199/377 (52%), Gaps = 31/377 (8%)

Query: 4   ARYRRRGNQNLWAYEIREEGKTVAYNS----GFKTKKLAEAEAEPILQKLRTGSIITKNI 59
           A +R+RG    W + +  +     Y      G+KTKK AEA A+   ++L   S    +I
Sbjct: 2   ANFRKRGKT--WQFRLSYKDNNGEYKKFEKGGYKTKKEAEAAADEAKKRLNNHSEFDNDI 59

Query: 60  SLPELYQEWLDLKIMPSNRSDVTKKKYLSRKVTLEKLFGDKPISQIRPSEYQRIMNNYGQ 119
           SL + +++W  +   P + ++ T + Y      ++K   DKPI++I P+ YQ ++N
Sbjct: 60  SLYDFFEKWAKVYKKP-HVTEATWRTYKRTLNLIDKYIKDKPIAEITPTFYQAVLNKMSL 118

Query: 120 RVSRNFLGRLNTGVKQSLQMAIADKVMIEDFTQNVELFSTVKSQDADSKYLHSEKAYLDL 179
             +  L +    +K ++++A+ +KV+ E+F    +  S + ++  + KYLH+++ YL L
Sbjct: 119 LYRQESLDKFYFQIKSAMKIAVHEKVISENFADFTKAKSKLAARPVEEKYLHADE-YLKL 177

Query: 180 INAVKDKFNYKKSVVPYIIYFLLKTGMRYGELIALTWEDIDFDKGIFKTYRRFN-SETSQ 238
           +   ++K  Y      +  Y    TGMR+ EL+ LTW  +DFDK       R ++ S T+
Sbjct: 178 LAIAEEKMEYTSY---FACYLTAVTGMRFAELLGLTWSHVDFDKKEISIQRTWDYSITNN 234

Query: 239 FVPPKNKTSIRIVPVDNECLEILKNLKIEQNQSNKELGLQNTNNMVFQHFGYPNSVPSTN 298
           F   KN++S R +P+ ++ +++LK  K        KE   +N  + V  +         S N
Sbjct: 235 FAETKNESSKRKIPISSKTIKLLKKYK-------KEYWHENKYDRVIYNL-------SNN 280

Query: 299 GTNKVLRGIVQELNIEPIITTKGARHTYGSFLWHRGYDLGIIAKILGHKDISMLIEVYGH 358
           G NK ++ ++    + P       RH++ S+L ++G DL  ++K+LGH+++++ ++VY H
Sbjct: 281 GLNKTIK-VIAGRKVHP----HSLRHSFASYLIYKGIDLLTVSKLLGHENLNVTLKVYAH 335

Query: 359 TLEEKIQEEYNEIKQLW 375
            L+E  QE  + I++++
Sbjct: 336 QLKEMEQENNDVIRKIF 352
```

There is also homology to SEQ ID 578.

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

Example 550

A DNA sequence (GBSx0589) was identified in *S.agalactiae* <SEQ ID 1749> which encodes the amino acid sequence <SEQ ID 1750>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2710(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 551

A DNA sequence (GBSx0590) was identified in *S.agalactiae* <SEQ ID 1751> which encodes the amino acid sequence <SEQ ID 1752>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2534(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 GENPEPT database:

```
>GP:BAA06248 GB:D29979 ORF3 [Bacillus stearothermophilus]
 Identities = 81/263 (30%), Positives = 135/263 (50%), Gaps = 14/263 (5%)

Query: 65  MGVHVELKGQGCRQYEEFIEGNDNNWTSLVKRLI-DNNSNFTRLDIANDIFDESLNVQRL 123
           MG+HVE+ GQGCR +E        NW  L  RL+ +   N TRLD+A D F     +  L
Sbjct: 1   MGIHVEMTGQGCRLFELH---TSINWYELFYRLVYEYEVNITRLDVAVDDFKGYFKINTL 57

Query: 124 YEYSKKGLCITTARHAEYHEKFVIDSGELVGETVVFGARGNQQWCVYNKLMEQNGKLQTD 183
            +  K     +  + A + E  VI+ GE +G T+ FGA  +    +  +  E+N ++  D
Sbjct: 58  VKKLKDDEVTSRFKKARHIENIVIEGGETIGHTLYFGAPSSD---IQVRFYEKNVQMGMD 114

Query: 184 IDINSWVRAELRCWQEKANLIAHQL-NDMRPLASIYFEAINGHYRFVSPKARDKNKRRRE 242
           ID+  W R E++   ++A+++A  + +D+ PL  I    +  + +F + KA DKNK+R
Sbjct: 115 IDV--WNRTEIQLRDDRAHVVAQIIADDVLPLGEIVAGLLRNYIQFRTRKATDKNKKRWP 172

Query: 243 SVRWWQNYINTEEKTRLSIVREKPTLRQSEAWTDKQVSKTIAKVYMAKYEAYGIDQAEVF 302
             R+W N++   +  R++    K ++ +   W D QVSK+   +Y    E    ++ + F
Sbjct: 173 LARFWLNFLGDVQPLRIAKQMPKTSIEKKYRWIDSQVSKSFFMIYYCLNE----EEKQRF 228

Query: 303 LQDLLRRGVEKFTDNDEKEIEQY 325
           + D+L  G  K T  D + I Q+
Sbjct: 229 IDDVLAEGASKLTKADLQVINQF 251
```

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

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

Example 552

A DNA sequence (GBSx0591) was identified in *S.agalactiae* <SEQ ID 1753> which encodes the amino acid sequence <SEQ ID 1754>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2700(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 553

A DNA sequence (GBSx0592) was identified in *S.agalactiae* <SEQ ID 1755> which encodes the amino acid sequence <SEQ ID 1756>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3121(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1757> which encodes the amino acid sequence <SEQ ID 1758>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2913(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 19/52 (36%), Positives = 33/52 (62%)

Query: 8  FGPNLTRLRKERGISQVELSNQLQIGKQSISDYEKQKAFPTFANLDKIAEYF 59
          F  NL  L  ++ I Q+++ N+L I K +I+ Y K ++ PT  N+ K+A++F
Sbjct: 15 FSTNLNMLMAKKNIKQIDIHNKLGIPKSTITGYVKGRSLPTAGNVQKLADFF 66
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 554

A DNA sequence (GBSx0593) was identified in *S.agalactiae* <SEQ ID 1759> which encodes the amino acid sequence <SEQ ID 1760>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> 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>
```

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

```
>GP:AAA98584 GB:L44593 ORF536; putative [Lactococcus phage BK5-T]
 Identities = 248/532 (46%), Positives = 359/532 (66%), Gaps = 16/532 (3%)

Query: 1   MNFIEQISENNQFPIIFVGSGITQRYFENAPTWEKLLKDIWLELFDEESYYAK--AFELR 58
           MNFIE I +NNQFPIIFVGSG+T+RYF+N   WE+LL ++W  + +E+++Y +   FE
Sbjct: 1   MNFIENIKDNNQFPIIFVGSGVTKRYFKNGLKWEQLLLELWNLVEEEKAFYTQYHVFENL 60

Query: 59  ERFEN-----NDFDIYTNLASLLEKEVSKAFINGNIQVDNLDLKTAYELNISPFKQLVAN 113
            + +N      +F+I   +A +LE++++ AF +  + +DNL L  A+  +ISPF+Q +AN
Sbjct: 61  LKSKNLSKSDKEFEINLMMAGILEEKINNAFYSDELNIDNLTLAQAHTEHISPFRQCIAN 120

Query: 114 RFSNLKIREEKIEEIKQFSQMLSKARIIITTNYDNFIEECLKTINVSVKINVGNKGLFLK 173
```

```
             FSNL  ++   EEI  FS+ML KAR I+TTNYDNFIEEC    NVS+K+NVGN GLF+K
Sbjct: 121 TFSNLDRKKGFDEEIISFSKMLVKARFIVTTNYDNFIEECFSKRNVSIKVNVGNSGLFVK 180

Query: 174 SSDYGELYKIHGTVDDASTITITKEDYEKNVTKSALINAKILSNLVESPILFLGYSLTDE 233
           S+DYGELYKIHG+V + +TI IT EDY+ N +K AL+NAKILSNL ESPILF+GYSLTD+
Sbjct: 181 SNDYGELYKIHGSVKNPNTICITSEDYKNNESKLALVNAKILSNLTESPILFIGYSLTDK 240

Query: 234 NIRKLLTDFAENSPFDISESAQKIGVVEYLPDSESIETVVSSLPDLSVYYSCLKTDNFTN 293
           NIR+LLT ++EN P++ISE+A +IGVVEY PD   I+ +VS++PDL ++Y+ + TDN+
Sbjct: 241 NIRELLTSYSENLPYEISEAAARIGVVEYTPDKIEIQDIVSNIPDLGIHYTKISTDNYKK 300

Query: 294 IYRLISKINQGFLPSEIAKYENVFRKIIEVKGESKDLKTVLTSYEDLANLTEDEIRSKNI 353
           IY  IS+I QG+LPSEIAK+E  FRKIIEVKG+ K+L TVLTS+ D++ +  +E+++KNI
Sbjct: 301 IYDEISQIEQGYLPSEIAKFEGAFRKIIEVKGKEKELDTVLTSFIDISKINTEELKNKNI 360

Query: 354 VVAFGDERYIYKFPDFKEYVRSYFLDKETIPQEIVIRFIATQPVASHLPIKKYMFAMSEY 413
           VVAFGD +YIYK P +K+Y+R YF +    +   I + F+  +      +P KK+M  +  +
Sbjct: 361 VVAFGDSKYIYKMPTYKDYIREYFSNSMELDTRIALLFLKKRSANYPVPYKKHMGVIESW 420

Query: 414 --ISKDSNKYTENIKKRLSKEEELSLDDFTSSIGVPLL--HSKTLERQTEIVGILE-ADV 468
             I  D  +  E++K R+S   E  +  +     L   +   L + + I  ++  ++V
Sbjct: 421 GSIPNDLVQEVESLKTRISNFPESIVRTYSIKANKDLAKKYLPYLNKTSTIEDVMSLSNV 480

Query: 469 PDNVRYNFIATHIKNFPKEELFLLVEKIID----EGIFETSRRRFLKAFDLL 516
           P   + FI   I  F  EEL   + K ID    +GI  T  R+ + ++ ++
Sbjct: 481 PLYNKLRFILFKIDKFKVEELKDFIVKNIDMGEGKGISSTLYRKIVMSYSII 532
```

A related GBS gene <SEQ ID 8599> and protein <SEQ ID 8600> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:       1.55
GvH: Signal Score (-7.5): 0.27
     Possible site: 54
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   2.44 threshold:  0.0
   PERIPHERAL  Likelihood =  2.44     214
 modified ALOM score:  -0.99

*** 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>
```

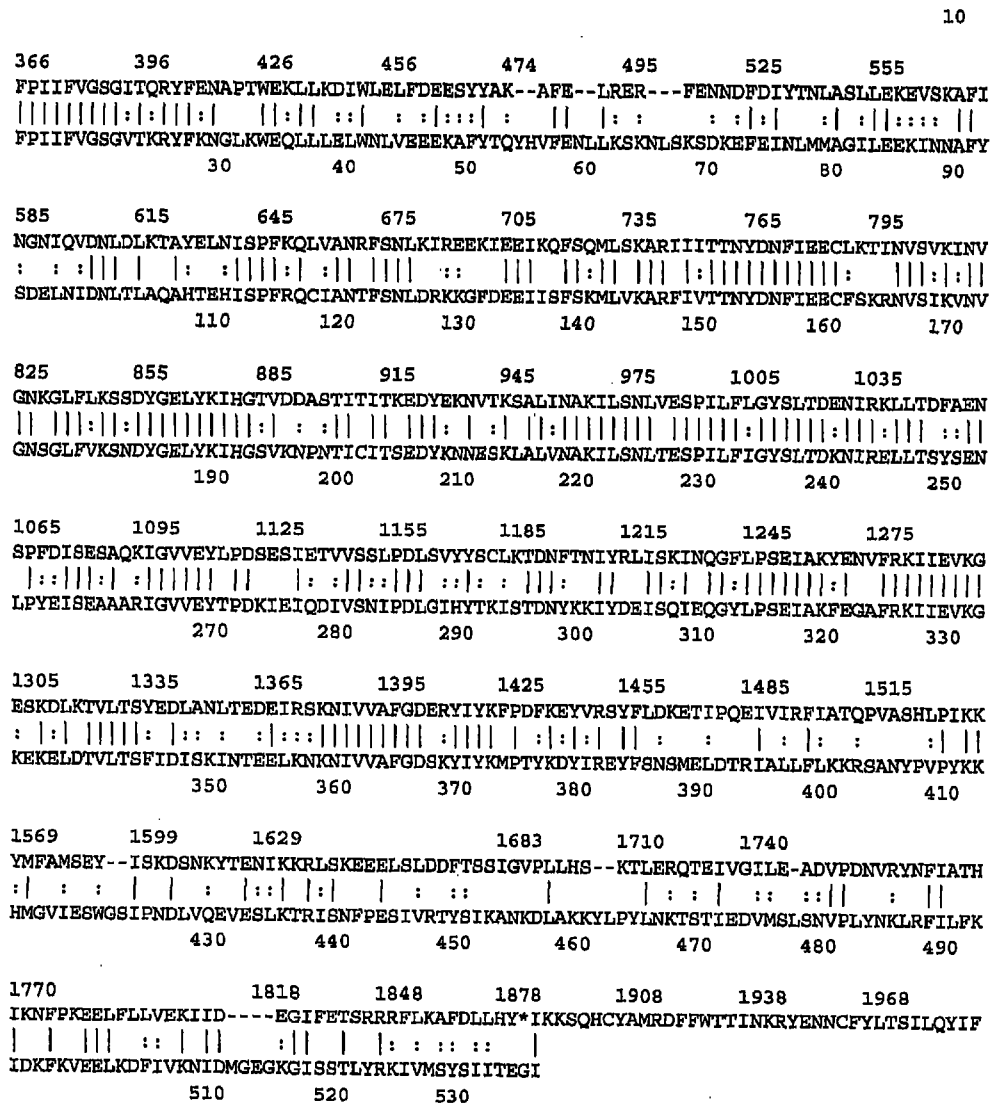
The protein has homology with the following sequences in the databases:

```
47.3/69.1% over 531aa

                                                                  Lactococcus lactis
  EGAD|36707| hypothetical protein Insert characterized
  GP|928833|gb|AAA98584.1||L44593 ORF536; putative {Lactococcus lactis phage BK5-T} Insert
characterized
  PIR|T13261|T13261 hypothetical protein 536 - phage BK5-T Insert characterized

ORF00184(301 - 1848 of 2154)
EGAD|36707|38110(1    -    532    of    536)    hypothetical    protein    {Lactococcus
lactis}GP|928833|gb|AAA98584.1||L44593  ORF536;  putative  {Lactococcus  lactis  phage  BK5-
T}PIR|T13261|T13261 hypothetical protein 536 - Lactococcus lactis phage BK5-T
%Match = 32.3
%Identity = 47.2  %Similarity = 69.0
Matches = 247  Mismatches = 155  Conservative Sub.s = 114

126       156       186       216       246       276       306       336
RMLILKAFYLAKFLKYYC*KK*CGTKRGQLYFRVYGLIIKINKMVSKML**D*QLNKLIINKR*GQELVNFIEQISENNQ
                                                                 :|||| | :|||
                                                                 MNFIENIKDNNQ
```



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

SEQ ID 8600 (GBS142) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 5; MW 54kDa). 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 6; MW 79.8kDa).

The GBS142-GST fusion product was purified (Figure 195, lane 3) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 249). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 555

A DNA sequence (GBSx0594) was identified in *S.agalactiae* <SEQ ID 1761> which encodes the amino acid sequence <SEQ ID 1762>. This protein is predicted to be integrase. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2933(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 GENPEPT database:

```
>GP:AAA98585 GB:L44593 integrase [Lactococcus phage BK5-T]
 Identities = 124/382 (32%), Positives = 202/382 (52%), Gaps = 21/382 (5%)

Query: 1   MATYRQRGKKKLWDYRIFNEKSELVA-SGSGFKTKREAMNEAMRIE---QQKLLVNSISS 56
           MATY++RGK   W Y I   K  L   +  GF TK +A  EAM IE   ++  +V+ I
Sbjct: 1   MATYQKRGKT--WQYSISRTKQGLPRLTKGGFSTKSDAQAEAMDIESKLKKGFIVDPIKQ 58

Query: 57  DITLYDL-WFEWYSLIIKPSNLAETTKNKYFTRGSVIRKLFGNQKVNKIKHSAYQRKLNT 115
           +I+ Y   W E Y    K + + E T   Y      ++    N  +++I  S+YQR LN
Sbjct: 59  EISEYFKDWMELY----KKNAIDEMTYKGYEQTLKYLKTYMPNVLISEITASSYQRALNK 114

Query: 116 YAEKYTKNHVRRLNSDIKKAIQFAKRDGVLLSDFTDGVVIAGRKFVKDADDKYLHSIFD- 174
           +AE + K   +  ++ ++ +IQ    +G L  DFT   V+ G    K   DK+++  FD
Sbjct: 115 FAETHAKASTKGFHTRVRASIQPLIEEGRLQKDFTTRAVVKGNGNDKAEQDKFVN--FDE 172

Query: 175 YKKVISYLENNLD--YSNSIVYYLLLVLFKTGLRVGEALALTWDDVNFEDLEIKTYR--R 230
           YK+++ Y  N L+  YS+  + +++ +   TG+R  EA  L WDD++F +  IK  R
Sbjct: 173 YKQLVDYFRNRLNPNYSSPTMLFIISI---TGMRASEAFGLVWDDIDFNNNTIKCRRTWN 229

Query: 231 FSGDKGTFSPPKTKTSIRTIPISQSLALILRDLKDDQQVMLKNLKIVNMNNQIFYDYRYG 290
           +   G F  PKT   IR I I      +L+D ++ Q+ + ++L I  +++ + Y
Sbjct: 230 YRNKVGGFKKPKTDAGIRDIVIDDDESMQLLKDFREQQKTLFESLGIKPIHDFVCYHPYRK 289

Query: 291 VSTNSAINKSLKNVLKILNINSKMTATGARHTYGSYLLAKGVDIWVVARLMGHKDITQLL 350
           + T SA+  +L + LK LNI++ +T  G RHT+ S LL  GVDI  V++ +GH  +
Sbjct: 290 IITLSALQNTLDHALKKLNISTPLTIHGLRHTHASVLLYHGVDIMTVSKRLGHASVAITQ 349

Query: 351 ETYGHVLTEVINKEYETVRSLV 372
           +TY H++ E+ NK+ + +  L+
Sbjct: 350 QTYIHIIKELENKDKDKIIELL 371
```

There is also homology to SEQ ID 578.

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

Example 556

A DNA sequence (GBSx0595) was identified in *S.agalactiae* <SEQ ID 1763> which encodes the amino acid sequence <SEQ ID 1764>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1603(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10209> which encodes amino acid sequence <SEQ ID 10210> was also identified.

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

```
>GP:BAB07266 GB:AP001519 unknown conserved protein in others
```

```
            [Bacillus halodurans]
 Identities = 26/71 (36%), Positives = 39/71 (54%), Gaps = 6/71 (8%)

Query: 37  WWDIDNLQELLGIGRSKLINDILLNPDIKKEVDLSINPNGFIVYPKGKGSRYKILATK-- 94
           WW + +L+E  G     L  +ILL+P  K  +D  I   GF+ YP+ KG R+  +A+
Sbjct: 4   WWSMQDLKERTGYSEDWLKENILLHPRYKPMLD--IENGGFVYYPEKKGERWCFIASSME 61

Query: 95  --ARKYFEDNF 103
             +KYF+D F
Sbjct: 62  EFLKKYFKDIF 72
```

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 557

A DNA sequence (GBSx0596) was identified in *S.agalactiae* <SEQ ID 1765> which encodes the amino acid sequence <SEQ ID 1766>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.88   Transmembrane   12 -  28 (  11 -  29)

----- Final Results -----
               bacterial membrane --- Certainty=0.2550(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 GENPEPT database:

```
>GP:AAB99663 GB:U67604 chromosome segretation protein (smc1)
            [Methanococcus jannaschii]
 Identities = 53/210 (25%), Positives = 95/210 (45%), Gaps = 33/210 (15%)

Query: 20  IFTNVGVLISNSRDNKAIQRELELLEEGQEKLVDEFSKISTNQYDKYV----------LI 69
           +F  +G+L  N      + + + +    + K++DE S I+     K            LI
Sbjct: 133 LFRRLGLLGDNVISQGDLLKIINISPIERRKIIDEISGIAEFDEKKKKAEEELKKARELI 192

Query: 70  Q------SNLSNNIEKNKQELVQKNSYVK--EDTKYIRDEMLIEKKSK-----EEVYNHV 116
           +      S + NN++K K+E      Y+K  E+ K  +  ++++K S      E + N +
Sbjct: 193 EMIDIRISEVENNLKKLKKEKEDAEKYIKLNEELKAAKYALILKKVSYLNVLLENIQNDI 252

Query: 117 KNGDKLIEKMAFANELILKFGEVSRENQMLGLKVNSLEEKIVDLSNQPKNDEISKLRKSI 176
           KN ++L       NE + K  E+  E + L L++N+     I++  N+  N+E+ +L KSI
Sbjct: 253 KNLEEL------KNEFLSKVREIDVEIENLKLRLNN----IINELNEKGNEEVLELHKSI 302

Query: 177 SSFERELSRFEDVGYSEAEEIKSTLRRILN 206
              E E+   + V  S   E+K     I N
Sbjct: 303 KELEVEIENDKKVLDSSINELKKVEVEIEN 332
```

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

SEQ ID 1766 (GBS315) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 42 (lane 4; MW 26.7kDa) and in Figure 239 (lane 5; MW 41kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 5; MW 52kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 558

A DNA sequence (GBSx0597) was identified in *S.agalactiae* <SEQ ID 1767> which encodes the amino acid sequence <SEQ ID 1768>. This protein is predicted to be surface protein. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.70   Transmembrane  229 - 245 ( 226 - 248)

----- Final Results -----
               bacterial membrane --- Certainty=0.4079(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 GENPEPT database:

```
>GP:CAA47097 GB:X66468 orf iota [Streptococcus pyogenes]
 Identities = 90/262 (34%), Positives = 138/262 (52%), Gaps = 26/262 (9%)

Query: 4   VKVLSLITV-SGLFLMAGNLSASADVVISGGDTIMLSGVDAGVSDSIMPPPSSINPV--- 59
           +K L+L+T+ S    L++  + + AD   S  D  +L+  D  V       P + ++PV
Sbjct: 1   MKKLALLTLFSTTLLVSAPIVSFADETASSSDINILADDDPVVPVEPTDPTTPVDPVDPV 60

Query: 60  -----------TDTTEPSAPTPSTDPI--TDTTEPSAPTPSTDPI--TDTTEPSAPTPST 104
                     T+ TEP+ PT  T+P   T+ TEP+ PT  T+P   T+ TEP+ PT  T
Sbjct: 61  DPVDPVDPVDPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPT 120

Query: 105 DQTTGTTDSS-TPSSSTTNPVDGITDNGTKPNAGIDKPSTNKPSDHSESSI--KPVTKPT 161
           + T  T  +  T  S  T P +      T+P    +     +PS  +E ++  KPV
Sbjct: 121 EPTEPTEPTEPTEPSKPTEPTE--PSKPTEPTEPTEPSKPTEPSKPTEPTVPNKPVDTNP 178

Query: 162 INQPITTVTGDQVIGTQDGKVLVQTPSGTQLK-DAAEVGGNVQKDGTVAIKKSDGKIEVL 220
           I  P+ T TG  ++  +D K ++Q   GT  K +A E+G +VQKDGTV +K SDGK++VL
Sbjct: 179 IENPVNTDTGVVIVAVEDSKPIIQLADGTTKKVEAKEIGADVQKDGTVTVKGSDGKMKVL 238

Query: 221 PKTGEGKTI-FTIVGLLLIAGA 241
           PKTGE   I  +++G L++ G+
Sbjct: 239 PKTGETANIALSVLGSLMVLGS 260
```

There is also homology to SEQ ID 760.

SEQ ID 1768 (GBS141) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 4; MW 35kDa). The GBS141-His fusion product was purified (Figure 194, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 295), which confirmed that the protein is immunoaccessible on GBS bacteria.

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

Example 559

A DNA sequence (GBSx0598) was identified in *S.agalactiae* <SEQ ID 1769> which encodes the amino acid sequence <SEQ ID 1770>. 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>
```

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

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

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

A related GBS gene <SEQ ID 8601> and protein <SEQ ID 8602> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:     14.39
GvH: Signal Score (-7.5): -1.23
     Possible site: 18
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   8.96 threshold:  0.0
   PERIPHERAL  Likelihood =  8.96     104
 modified ALOM score:  -2.29

*** 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 1770 (GBS17) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 2; MW 24kDa).

The His-fusion protein was purified as shown in Figure 189, lane 10.

Example 560

A DNA sequence (GBSx0599) was identified in *S.agalactiae* <SEQ ID 1771> which encodes the amino acid sequence <SEQ ID 1772>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> 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>
```

A related GBS gene <SEQ ID 10779> and protein <SEQ ID 10780> were also identified. A further related GBS nucleic acid sequence <SEQ ID 10957> which encodes amino acid sequence <SEQ ID 10958> was also identified.

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

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

SEQ ID 1772 (GBS643) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 129 (lane 2-4; MW 79kDa) and in Figure 186 (lane 2; MW 79kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 129 (lane 5-7; MW 54kDa) and in Figure 176 (lane 5; MW 54kDa).

GBS643-GST was purified as shown in Figure 236, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 561

A DNA sequence (GBSx0600) was identified in *S.agalactiae* <SEQ ID 1773> which encodes the amino acid sequence <SEQ ID 1774>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5815(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 562

A DNA sequence (GBSx0601) was identified in *S.agalactiae* <SEQ ID 1775> which encodes the amino acid sequence <SEQ ID 1776>. This protein is predicted to be membrane protein. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.32   Transmembrane  311 - 327 ( 282 - 332)
   INTEGRAL    Likelihood =-10.46   Transmembrane  293 - 309 ( 282 - 310)
   INTEGRAL    Likelihood = -8.55   Transmembrane  390 - 406 ( 388 - 410)
   INTEGRAL    Likelihood = -7.64   Transmembrane   49 -  65 (  40 -  69)
   INTEGRAL    Likelihood = -5.68   Transmembrane  100 - 116 (  98 - 122)
   INTEGRAL    Likelihood = -4.35   Transmembrane  130 - 146 ( 127 - 148)
   INTEGRAL    Likelihood = -3.88   Transmembrane  344 - 360 ( 342 - 363)

----- 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>
```

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

```
>GP:CAB70618 GB:AJ243106 membrane protein [Streptococcus thermophilus]
 Identities = 234/665 (35%), Positives = 379/665 (56%), Gaps = 59/665 (8%)

Query: 13  FAKVKDVDIFALKAYMEITH-GAETGAQSILLDVFVNFPFFLLNLIVGLFSVILRFFENF 71
           FAK+K VDIF+LK+YME T+ G+  GA  ++ ++FVN  FF+LN +VG FS+++R  E
Sbjct: 5   FAKLKGVDIFSLKSYMEPTNFGSFNGAWVLINELFVNLFFFILNAVVGFFSLLIRILEKI 64

Query: 72  SLYDTYKQTVYHSSQKLWENLSGN--GSYTS-SLLYLLVAISAFSIFISYLFSKGDFSKR 128
            LY TYK  V+H +  +W   +G+  G+ T+ SL+  L+ + AF +F  Y FSKG FS+
Sbjct: 65  DLYATYKTYVFHGASSIWHGFTGSNTGNITNKSLVGTLLLVLAFYLFYQYFFSKGSFSRT 124

Query: 129 LIHLFVVIILGMGYFGTIQSTSGGIYILDTVHQLAGSFSDAVTNLSLDNPSGGKTKITQK 188
           L+H+ +V++L +GYFGT+  TSGG+Y+LDTV+ ++   +  +  + +D       KI +
Sbjct: 125 LLHVCLVLLLALGYFGTVAGTSGGLYLLDTVNNVSKDVTKKIAGIKVDYAKDKSIKIGK- 183

Query: 189 SSVADNYVMKTSYTAYLFVNTGQLNGKFHNNQTGKEEKFDNEQVLGKYDKSGKFITPKQK 248
            S++D+Y+ +TSY AY+FVNTGQ NGK+ N+Q GKEE FD+ +VLG  DK+G F   K K
```

```
Sbjct: 184 -SMSDSYIAETSYKAYVFVNTGQENGKYKNSQDGKEEAFDDSKVLGTSDKNGNFKAVKAK 242

Query: 249 DILNYTDNLGDKATEGEEKNRWLSAVNDYLWIKSGYVILKIFEAVILAVPLILIQLIAFM 308
           +   Y D+LG+ A +  EKNRW+SA+ D+++ +  YVI KI EA +LAVP+ILIQL+  +
Sbjct: 243 ERSKYLDDLGEGANDDGEKNRWVSAMPDFIFTRVFYVIFKIVEAFVLAVPIILIQLLNVV 302

Query: 309 ADVLVIILMFIFPLALLVSFLPRMQDIIFNVLKVMFGAVSFPALAGFLTLIVFYTQTLIA 368
           A +LV+ ++ +FP+ LL+SF+PRMQ+++F VLKVMFG + FPA+   LTL++FY + +I
Sbjct: 303 AQILVLTMILLFPVVLLMSFVPRMQELVFGVLKVMFGGLIFPAITTLLTLLIFYIEKMIE 362

Query: 369 TFVKKKFTDGSLLSGSNFKGQAILFMLLITVFVQGCVFWGIWKYKETFLRLIIGSRASQV 428
             V   F DG L +  +     ++F LL++V  +G +++ IW++K   L+ I+GS+A  V
Sbjct: 363 NIVTNGF-DGVLKTLPSLLLFGLVFKLLVSVVSKGVIYFLIWRFKGQLLQFILGSKARMV 421

Query: 429 -------INQSVDKINEKAENLGITPKSIYERAHDMSSLAMMGAGYGVGTMMNAQ---DN 478
                  +   V K  E A  +   P      A  + +  + GAG+G G MMNA+     N
Sbjct: 422 ATDIGTKVEHGVTKSKEVASQV---PTRSLATAQHLGNFTLAGAGFGTGVMMNAKSHFQN 478

Query: 479 WNAFKERQQANLDDGQSKTNDADKYDEANADDTVISKEAELTNEGEYQSELPKEASKRIE 538
             +F  R++ +  +    +  +      + + +I            ++ P +  K I
Sbjct: 479 AGSFFTRKEPSQPETVMPSGPTEAPITPESPEPIIP----------PTQTPPDNFKTIG 527

Query: 539 QLGKESSYELSFISEGNSTEEILKNVKSDNHTFQEGDGDTSLTNQDMITNDIENHSNNYT 598
           +       +   +SEG  + E                              ++ +     +
Sbjct: 528 EEKPTPPSDSPIMSEGTPSSE----------------------------DEFQTLKEEWM 559

Query: 599 SPLKQRKLNKLEGELSQFNSDVSMTKNHGKNAFEKGFNASKTKEVRKQHNLERQSKVLEE 658
           SP KQ ++N LE  L  +    +M K  G NAF + +  + T++ + + N+ER+ ++ +
Sbjct: 560 SPFKQHRINTLERRLDAYKDPQAMYKAQGSNAFTRAYRKTLTRDDKIRANIERRDRLTQR 619

Query: 659 LEKLR 663
           L +LR
Sbjct: 620 LNQLR 624
```

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

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

Example 563

A DNA sequence (GBSx0602) was identified in *S.agalactiae* <SEQ ID 1777> which encodes the amino acid sequence <SEQ ID 1778>. This protein is predicted to be conjugative protein. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3714(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 GENPEPT database:

```
>GP:CAB70617 GB:AJ243106 conjugative protein [Streptococcus thermophilus]
 Identities = 515/757 (68%), Positives = 612/757 (80%), Gaps = 1/757 (0%)

Query: 1   MSDFEADLADDVKELGLETLDFTVDTLTHEMEIPYQFDWLIGVDLGKGQYNANIKEFIYN 60
           M DF   LADD +ELG E L +TVD LT EMEIPYQFDW+IGV L K  + A +K+  Y
Sbjct: 78  MRDFSEALADDSRELGEELLLYTVDRLTDEMEIPYQFDWVIGVTLRKQNHGATVKDLAYE 137

Query: 61  QFESIASNFASLAGYEVEVDEDWYKEHSEEELLVYSLLSTLKAKRLTDVDLFYYQRMQFL 120
            F   +   A   GYE  +   WY ++  +E  ++   S L+AKRLT+ +LFYYQRMQ+L
Sbjct: 138 SFNEFSEKIAKGLGYEYALSPTWYDDYRSDEFTIFQAFSVLRAKRLTNEELFYYQRMQYL 197
```

```
Query: 121 RYVPHTKSEVIANRNMLNVTDTLIKSLEGGFLKLESAYGSSFVSVLPVGRFSTIFNGFHL 180
           RY+PH K EV+ANR+  N+TDTLIK L+GGFL+LES YGSSFV++LPVG+F   FNGFHL
Sbjct: 198 RYIPHYKKEVLANRSQFNITDTLIKVLKGGFLELESPYGSSFVTILPVGKFPVQFNGFHL 257

Query: 181 GELVQRMSFPVELRFKAEFIDKTKLGGTMGRSNTRYDQIMKEAYNTNTVQQDDILMGAYS 240
           GE VQR++FPVELR KAEFID  K+ G MGRSNTRY  IM+EA NT+TVQQD+I+MG+ S
Sbjct: 258 GEFVQRLNFPVELRIKAEFIDTNKIKGRMGRSNTRYRNIMEEAENTDTVQQDEIIMGSIS 317

Query: 241 LKDLMKKVGNKEEIIEYGCYLVVAGSSLNQLKQRRYAILSYFDDMKVNVYEASHDTPYLF 300
           LKDLMKKVGNKE+IIEYG YL+V+ SS+NQL+QRR  IL+YFDDM V + EAS D PYLF
Sbjct: 318 LKDLMKKVGNKEDIIEYGAYLIVSASSVNQLRQRRQVILNYFDDMGVEISEASQDGPYLF 377

Query: 301 QALLYGQDLQKTTRKWNHLVTARGFSELMLFTNTQSGNRIGWYIGRVDNRLTAWDSIDEA 360
           QALLYG++LQK TR W H+VTARGFSELM FTNT SGNRIGWYIGRVDN +  WDSI +A
Sbjct: 378 QALLYGENLQKKTRTWTHMVTARGFSELMPFTNTSSGNRIGWYIGRVDNWIGRWDSIAKA 437

Query: 361 IMGSKNLVLFNATVANKEDVAGKVTKNPHVIITGATGQGKSYLAQMIFLHTAQQNVRVLY 420
           I  SKN+VL+NATV NKED+AGK+TKNPH+IITGATGQGKS+LAQ+IFL  A QNV+ LY
Sbjct: 438 IDSSKNIVLYNATVGNKEDIAGKITKNPHIIITGATGQGKSFLAQIIFLSVALQNVKTLY 497

Query: 421 VDPKRELRQHYLKVVSDPEYARKFPLRKKQIEETNFVTLDSSVKENHGVLDPIVILDKEG 480
           +DPKRELR HY +V++ PE+AR++P RKKQI+  NFVTLDSS+  NHGVLDPIV+LDKE
Sbjct: 498 IDPKRELRNHYQEVINSPEFARRYPERKKQIDNFNFVTLDSSLPSNHGVLDPIVVLDKEQ 557

Query: 481 ASSTAKNMLLYLLKNATEIKLDQTTALTEAISQVIAKREAGEVVGFNQVIEVLIDSESDE 540
           A   AKNML +LL+   ++ +DQ TA+TEAI+ ++ +R AGE VGF  V+E L ++ S E
Sbjct: 558 AVEVAKNMLEFLLQAVDDVTMDQKTAITEAINTIVERRVAGENVGFKHVLETLRNASSSE 617

Query: 541 VQSVGRYFKAIIQNSILELAFSDGDVAGLSYEERVTVLEVADLSLPKDGSDHISDHESNS 600
           + SVGRY  +I+ NSILELAFSDG   GL+YE RVT+LEV +L LPKD S  ISDHE NS
Sbjct: 618 IASVGRYLTSIVTNSILELAFSDGTTPGLNYESRVTILEVNNLKLPKDDSTKISDHERNS 677

Query: 601 IALMFALGAFCKHFGERSDDE-TVEIFDEAWVLMQSSEGKAVIKSMRRVGRSKYNVLMLV 659
           IALMFALGAFC HFGER+++E T+E FDEAW+LM+S+EGKAVIK+MRR+GRSK N L L+
Sbjct: 678 IALMFALGAFCTHFGERNENEDTIEFFDEAWILMKSAEGKAVIKNMRRIGRSKNNTLALI 737

Query: 660 SQSVHDAENDDDTTGFGTIFSFYEKSEREDILSHVGLEVTPKNLEWIDNMISGQCLYYDV 719
           +QSVHDAENDDDTTGFGTIF+FYEKSEREDIL HV LEVT  NLEWIDNMISGQCLYYDV
Sbjct: 738 TQSVHDAENDDDTTGFGTIFAFYEKSEREDILRHVNLEVTESNLEWIDNMISGQCLYYDV 797

Query: 720 YGNLNMISIHNIHPDIDPLLKPMKKTVSSHLENKYAS 756
           YGNLNMIS+HN+  DID LLKPMK TVSS LENKYAS
Sbjct: 798 YGNLNMISVHNLFEDIDMLLKPMKATVSSSLENKYAS 834
```

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

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

Example 564

A DNA sequence (GBSx0604) was identified in *S.agalactiae* <SEQ ID 1779> which encodes the amino acid sequence <SEQ ID 1780>. This protein is predicted to be ISL2 protein. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3469(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 GENPEPT database:

```
>GP:CAC18595 GB:AJ278419 IS1381 transposase [Streptococcus pneumoniae]
 Identities = 110/125 (88%), Positives = 119/125 (95%)

Query: 81  MNYEASKQLTDVRFKRLVGVQRTTFEEMLAVLKTAYQRKHAKGGRTPKLSLEDLLMATLQ 140
           MNYEASKQLTD RFKRLVGVQRTTFEEMLAVLKTAYQ KHAKGGR PKLSLEDLLMATLQ
Sbjct: 1   MNYEASKQLTDARFKRLVGVQRTTFEEMLAVLKTAYQLKHAKGGRKPKLSLEDLLMATLQ 60

Query: 141 YMREYRTYEQIAADFGIHESNLIRRSQWVESTLIQSGFTISKTHLSAEDTVIVDATEVKI 200
           Y+REYRTYE+IAADFG+HESNL+RRSQWVE TL+QSG TIS+T LS+EDTV++DATEVKI
Sbjct: 61  YVREYRTYEEIAADFGVHESNLLRRSQWVEVTLVQSGVTISRTPLSSEDTVMIDATEVKI 120

Query: 201 NRPKK 205
           NRPKK
Sbjct: 121 NRPKK 125
```

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

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

Example 565

A DNA sequence (GBSx0605) was identified in *S.agalactiae* <SEQ ID 1781> which encodes the amino acid sequence <SEQ ID 1782>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.58   Transmembrane   39 -  55 (  32 -  66)

----- 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>
```

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

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

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

Example 566

A DNA sequence (GBSx0606) was identified in *S.agalactiae* <SEQ ID 1783> which encodes the amino acid sequence <SEQ ID 1784>. This protein is predicted to be Cag-W. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.82   Transmembrane   50 -  66 (  49 -  66)
   INTEGRAL    Likelihood = -3.72   Transmembrane   25 -  41 (  23 -  45)

----- 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>
```

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

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

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

Example 567

A DNA sequence (GBSx0607) was identified in *S.agalactiae* <SEQ ID 1785> which encodes the amino acid sequence <SEQ ID 1786>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.80   Transmembrane   36 -  52 (  32 -  60)

----- 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>
```

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

```
>GP:CAB12298 GB:Z99106 similar to transposon protein [Bacillus subtilis]
 Identities = 68/339 (20%), Positives = 133/339 (39%), Gaps = 49/339 (14%)

Query: 16  KKEEGGKQPKTKEVKQRTANFIV--YGILGLLFIVGFFGSLRAIGLSNQVQHLKETVIAV 73
           K+ E  ++ K K  + R+    V  +  +G L  +     L +I   +Q+  +K+
Sbjct: 24  KRIERPEKDKQKVPRDRSKLIAVTLWSCVGSLLFICLLAVLLSINTRSQLNDMKDETNKP 83

Query: 74  EKKSKHKKTDDSLDISRIQYYMNNFVYYYINYS--QDTADQRKTELENY--------YSF 123
              K K     + ++  + +++ F+  Y+N    Q++ ++R   LE+Y        +
Sbjct: 84  TNDDKQK-----ISVTAAENFLSGFINEYMNVKNDQESIEKRMQSLESYMVKQEDNHFED 138

Query: 124 STASMTDDVRKSRTLQTQRLISVEKEKDYYIALMRIGYEV---------------------- 163
                D ++  R L+   L +V++     +   ++ YE
Sbjct: 139 EERFNVDGLKGDRELKGYSLYNVKEGDKNSLFQYKVTYENLYPVEKEVEKEVKDGKKKKK 198

Query: 164 --------DKKSYQMNLAVPFQMQRGLLAIVSQPYTVAEDLYLGKSKAFEKKTLDQVKEL 215
                  +K   QM L +P   +    A+ + PY     +Y  K     K   +   E
Sbjct: 199 VKEKVKTNEKYEKQMLLNIPVTNKGDSFAVSAVPYFT--QIYDLKGDIAFKGKEETRDEY 256

Query: 216 SKEQVSSIQKFLPVFFNKYALINKTDLKLLMKTPELMGKGFKVSELDLNNAIYYQEKKHQ 275
           + E+  SI+ FL  FF KYA   K ++  +MK PE +       E  + +   ++ KK
Sbjct: 257 AGEKKESIESFLQNFFEKYASEKKEEMVYMMKKPEALEGNLLFGE--VQSVKIFETKKGF 314

Query: 276 VVQLSVTFEDLVTGGTRSENFTLYLFKADNGWYVEEMYH 314
            V  +V F++        +E F+L + +     +YV ++ H
Sbjct: 315 EVFCAVRFKEKENDIPVNEKFSLEITENSGQFYVNKLKH 353
```

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

SEQ ID 1786 (GBS333d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 8-10; MW 58kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 11 & 13; MW 33kDa), in Figure 182 (lane 2; MW 33kDa) and in Figure 185 (lane 3; MW 58kDa).

GBS333d-GST was purified as shown in Figure 236, lane 2.

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

Example 568

A DNA sequence (GBSx0608) was identified in *S.agalactiae* <SEQ ID 1787> which encodes the amino acid sequence <SEQ ID 1788>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4177(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 GENPEPT database:

```
>GP:CAB38326 GB:Y17736 hypothetical protein [Streptomyces
         coelicolor A3(2)]
 Identities = 45/80 (56%), Positives = 56/80 (69%)

Query: 4  FTEEAWKDYVSWQQEDKKILKRINRLIEDIKRDPFEGIGKPEPLKYHYSGAWSRRITEEH 63
          FT   W+DYV W + D+K+ KRINRLI DI RDPF+G+GKPEPLK   SG WSRRI + H
Sbjct: 5  FTSHGWEDYVHWAESDRKVTKRINRLIADIARDPFKGVGKPEPLKGDLSGYWSRRIDDTH 64

Query: 64 RLIYMIEDGEIYFLSFRDHY 83
          RL+Y   D ++  +  R HY
Sbjct: 65 RLVYKPTDDQLVIVQARYHY 84
```

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

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

Example 569

A DNA sequence (GBSx0609) was identified in *S.agalactiae* <SEQ ID 1789> which encodes the amino acid sequence <SEQ ID 1790>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5669(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10211> which encodes amino acid sequence <SEQ ID 10212> was also identified.

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

```
>GP:AAD17306 GB:AF121418 putative Phd protein [Francisella
         tularensis subsp. novicida]
 Identities = 26/84 (30%), Positives = 45/84 (52%)

Query: 4  MEAIVYSHFRNNLKDYMKKVNDEFEPLIVVNKNPDENIVVLSQDSWESLQETIRLMENDY 63
          M+ + YS FRN L D M +V     P+IV   +  E +V++S + +++ +ET  LM +
Sbjct: 1  MQTVNYSTFRNELSDSMDRVTKNHSPMIVTRGSKKEAVVMMSLEDFKAYEETAYLMRSMN 60

Query: 64 LSHKVINGISQVKEKQVTKHGLIE 87
             ++ N I +V+     +  LIE
Sbjct: 61 NYKRLQNSIDEVESGLAIQKELIE 84
```

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 570

A DNA sequence (GBSx0610) was identified in *S.agalactiae* <SEQ ID 1791> which encodes the amino acid sequence <SEQ ID 1792>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2407(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 571

A DNA sequence (GBSx0611) was identified in *S.agalactiae* <SEQ ID 1793> which encodes the amino acid sequence <SEQ ID 1794>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1274(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10213> which encodes amino acid sequence <SEQ ID 10214> was also identified.

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

```
>GP:AAB60015 GB:U09422 ORF18 [Enterococcus faecalis]
 Identities = 41/140 (29%), Positives = 73/140 (51%), Gaps = 3/140 (2%)

Query: 23  FPVEMSELKLALGLREEDDLEYIIADSDCQL-LKEHDSIEMINQFVELVENVDSELVKAV 81
           FP++  E+K  +GL +E + EY I D +    + E+ SI  +N+  E+V  +  EL   +
Sbjct: 26  FPIDFEEVKEKIGLNDEYE-EYAIHDYELPFTVDEYTSIGELNRLWEMVSELPEELQSEL 84

Query: 82  HQVIGYTASDFVDYDFNFGDCCLLSDVTTRRELGEYYFDELGVQGVGKEALEMYFDHEAY 141
             ++ + +S   +   +  D  + SD     ++  YY +E G  G    +L+ Y D++AY
Sbjct: 85  SALLTHFSS-IEELSEHQEDIIIHSDCDDMYDVARYYIEETGALGEVPASLQNYIDYQAY 143

Query: 142 GRDIDLESQGGFSDYGYVEI 161
           GRD+DL      +++G  EI
Sbjct: 144 GRDLDLSGTFISTNHGIFEI 163
```

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 572

A DNA sequence (GBSx0612) was identified in *S.agalactiae* <SEQ ID 1795> which encodes the amino acid sequence <SEQ ID 1796>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1366(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 573

A DNA sequence (GBSx0613) was identified in *S.agalactiae* <SEQ ID 1797> which encodes the amino acid sequence <SEQ ID 1798>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1484(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 574

A DNA sequence (GBSx0614) was identified in *S.agalactiae* <SEQ ID 1799> which encodes the amino acid sequence <SEQ ID 1800>. This protein is predicted to be abortive phage resistance protein. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2205(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10215> which encodes amino acid sequence <SEQ ID 10216> was also identified.

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

```
>GP:AAB53710 GB:U94520 abortive phage resistance protein
          [Lactococcus lactis]
```

```
 Identities = 131/499 (26%), Positives = 210/499 (41%), Gaps = 97/499 (19%)

Query: 3   MFSKIEFKNFMSFSNLT------------------FDLLNRGKCKDIIAIYGENGSGKTN 44
            M     F+NF+SF  L+                  D+ N  K   +  IYG N SGK++
Sbjct: 1   MLVNFRFENFLSFDKLSTFSMAPGKSRQHMEDLIELDIKNNQKLLKLSTIYGANASGKSS 60

Query: 45  IVEAF---KLLVL-----SLQSMESLNENTRLQSLLKEQTNKE---ENQKTNFGDISEIL 93
             V+A    K L++      L    S N+NT   SL + +   E   E++  ++G  S IL
Sbjct: 61  FVDAIGISKSLIIRGFYNGLVLSNSYNKNTVDNSLNETKFEYEIVIEDKVYSYG-FSVIL 119

Query: 94  DKISFFTTFKGIAKNTHRIASEGNTILKYYFNIEKDNGYYLLEYNENNELVKEELVFKIK 153
               F + +     N  ++          Y    KDN      YN N+E     L    +
Sbjct: 120 SLKKFMSEWLYDITNDEKM---------IYTIDRKDN-----SYNINDEF----LNLDEQ 161

Query: 154 SNKGVHFSITNIDGLSQSLNKTIFKNTIFKDLTEQIEKYWGKHTFLSIFN--NYCLEV-- 209
            SN  +   I +    S + N  +F N++  D  + IE        F  +FN  N  LEV
Sbjct: 162 SNNRISIYIDD----SANDNTQLFLNSL-NDGKKTIESKDNSTIFKKVFNWFNNTLEVLG 216

Query: 210 ---------------NEEF---INEQVSINFQKVVDEFDKIFIWSGNFRGPFHSTELLLK 251
                          EEF   + + + +N   V+D          N   P    E +L
Sbjct: 217 PGDEARGSIASLTQEEEEFKEDLGKYLELNDTGVIDIVQVPVDNLSNV--PAKLQERILD 274

Query: 252 DISKGKIDKSEKEKLSYTEEIIYKYFSALYIDIKDVKYKQDAQGQEIKYELMIRKNIGGD 311
            +I+   I K +KE+     E I   F+ +    +++   Q+   Q   +EL   K+  G
Sbjct: 275 NITT-DIKKKKKER-----EDIEISFNTILNTSQNIYIIQNNDEQFEYFELKF-KHKNGT 327

Query: 312 LLDVPISLESQGTKNLLDLLKV-FNNVLDGKICIVDEIDSGIHDLLMNSILNDLK--GSV 368
            L    +S ES GT  L++L  V F+N  D K+ ++DEID  +H LL  + +   K   S+
Sbjct: 328 LYS--LSEESDGTVRLIELFSVLFHN--DEKVFVIDEIDRSLHPLLTYNFIESFKKQKSI 383

Query: 369 NGQLIFTTHDTTLL--KELSPSSAYFLNVDIKGNKVIISGNEADKKIGVNNNLEKLYLSG 426
            N QLI TTH+  +L  + L     +F++ + +GN  + S  E  ++    + ++   YL+G
Sbjct: 384 N-QLIVTTHEDYILNFELLRRDEVWFVDKNFEGNSSMFSLEEFKERF--DKDINTSYLNG 440

Query: 427 FFGAVPDPLDIDFSDLFLD 445
             +G +P+ L    FS+   D
Sbjct: 441 RYGGIPN-LSCLFSEFAKD 458
```

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 575

A DNA sequence (GBSx0615) was identified in *S.agalactiae* <SEQ ID 1801> which encodes the amino acid sequence <SEQ ID 1802>. This protein is predicted to be repressor (rstR-1). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3724(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 GENPEPT database:

```
>GP:AAB84427 GB:AF027868 transcription regulator [Bacillus subtilis]
 Identities = 31/81 (38%), Positives = 53/81 (65%), Gaps = 2/81 (2%)

Query: 9   QKLKELRKEKKLTQTELASKLNISQKSYSNWESGKAEPTLDNIIKLANILDVTVDYLLGR 68
            Q+L++LRK  KLT  +LA K+ I++ SY  +E+   +P LD ++ LA + DV+VDY+LG
Sbjct: 4   QRLRQLRKAHKLTMEQLAEKIGIAKSSYGGYEAESKKPPLDKLVILARLYDVSVDYILGL 63
```

```
Query: 69 SDNFSNTIVLSKNNMKSFSKR 89
          +D+    +   + N+K F ++
Sbjct: 64 TDDPDPKV--ERKNLKEFLEK 82
```

There is also homology to SEQ ID 1740.

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

Example 576

A DNA sequence (GBSx0616) was identified in *S.agalactiae* <SEQ ID 1803> which encodes the amino acid sequence <SEQ ID 1804>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3607(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 577

A DNA sequence (GBSx0617) was identified in *S.agalactiae* <SEQ ID 1805> which encodes the amino acid sequence <SEQ ID 1806>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0564(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10217> which encodes amino acid sequence <SEQ ID 10218> was also identified.

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

```
>GP:CAB12294 GB:Z99106 similar to transposon protein [Bacillus subtilis]
 Identities = 93/348 (26%), Positives = 164/348 (46%), Gaps = 28/348 (8%)

Query: 81  SRLQVMIDYVRITLKDVRDLEFFCRNFLHCAFKEFQPFESKLMNYNHLWKRGDIWIFDFA 140
           S L  M+DY+R++ K   D++      LH +       +S    Y   ++   I +F  A
Sbjct: 26  SPLVSMVDYIRVSFK-THDVDRIIEEVLHLSKDFMTEKQSGFYGYVGTYELDYIKVFYSA 84

Query: 141 DKHETGNFQITVQLSGRGCRQLELLMETEKFTWHDWLSYLRNSYRDDMNVTRFDIAIDEL 200
               G   + +++SG+GCRQ E  +E  K TW+D   + ++  +   + TRFD+AID+
Sbjct: 85  PDDNRG---VLIEMSGQGCRQFESFLECRKKTWYD---FFQDCMQQGGSFTRFDLAIDD- 137

Query: 201 YLGKDRENEQFHLSDMISKYYRHELDFESLRTWNYIGGGSLNFSDMEEIEQNRQGISLYF 260
                +   F + +++ K  + E      R  ++   GS + SD          G ++YF
Sbjct: 138 ------KKTYFSIPELLKKAQKGEC-ISRFRKSDF--NGSFDLSD-----GITGGTTIYF 183
```

```
Query: 261 GSRQSEMYFNFYEKRYEIAKQEGITVEEALEIFELWNRYEIRLSQSKANAAVDEFISGVP 320
           GS++SE Y  FYEK YE A++  I +EE  +    WNRYE+RL   +A  A+D  +
Sbjct: 184 GSKKSEAYLCFYEKNYEQAEKYNIPLEELGD----WNRYELRLKNERAQVAIDALLKTKD 239

Query: 321 IGEISRGLIVSKIDVYDGKNEY--GSFQADRKWQLMFGGVEPLKFVTKPEAYSIERTLRW 378
           +  I+  +I + +   D        ++    W   G V  L    KP+    +++  W
Sbjct: 240 LTLIAMQIINNYVRFVDADENITREHWKTSLFWSDFIGDVGRLPLYVKPQKDFYQKSRNW 299

Query: 379 LSDSVSPSLAMIREYDMIVDGDYLQTILNSGEVNERGEKILDSIKASL 426
           L +S +P++ M+ E D  +    L  ++   E+ ++ +K+LD   A +
Sbjct: 300 LRNSCAPTMKMVLEADEHLGKTDLSDMIAEAELADKHKKMLDVYMADV 347
```

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

A related GBS gene <SEQ ID 8603> and protein <SEQ ID 8604> were also identified. Analysis of this protein sequence reveals a RGD motif at residues 131-133.

The protein has homology with the following sequences in the databases:

```
29.4/54.5% over 342aa

                                                                    Bacillus subtilis
  EGAD|108511| hypothetical protein Insert characterized OMNI|NT01BS0566 conserved
hypothetical protein Insert characterized
  GP|1881297|dbj|BAA19324.1||AB001488 SIMILAR TO ORF20 OF ENTEROCOCCUS FAECALIS TRANSPOSON
TN916. Insert characterized
  GP|2632787|emb|CAB12294.1||Z99106 similar to transposon protein Insert characterized
  PIR|G69774|G69774 transposon-related protein homolog ydcR - Insert characterized

ORF00101(205 - 1581 of 1887)
EGAD|108511|BS0487(6 - 348 of 352) hypothetical protein (Bacillus subtilis) OMNI|NT01BS0566
conserved hypothetical protein GP|1881297|dbj|BAA19324.1||AB001488 SIMILAR TO ORF20 OF
ENTEROCOCCUS FAECALIS TRANSPOSON TN916. (Bacillus subtilis)
GP|2632787|emb|CAB12294.1||Z99106 similar to transposon protein (Bacillus subtilis)
PIR|G69774|G69774 transposon-related protein homolog ydcR - Bacillus subtilis
%Match = 9.7
%Identity = 29.3  %Similarity = 54.4
Matches = 103  Mismatches = 146  Conservative Sub.s = 88

153       183       213       243       273                   489       519
GVV*RSENGHAGHSAHTRALQRDLSILKPPFSNRGVRNEKFRILTPKNLYVSRVFR~~~~EQGKRKLTLEKYQEIKSHFG
                          : :|| :||||
                        MDELKQPPHANRGV----------------------------------------
                                  10

567       597       627       657       687       717       747
YLV--ENDS--SRLQVMIDYVRITLKDVRDLEFFCRNFLHCAFKEFQPFESKLMNYNHLWKRGDIWIFDFADKHETGNFQ
 :|  :|::  | |  |:||:|:::|   |::      || :      :| :  |   ::   | :|  |      |
VIVKEKNEAVESPLVSMVDYIRVSFK-THDVDRIIEEVLHLSKDFMTEKQSGFYGYVGTYELDYIKVFYSAPDDNRG---
              30        40        50        60        70        80        90

777       807       837       867       897       927       957       987
ITVQLSGRGCRQLELLMETEKFTWHDWLSYLRNSYRDDMNVTRFDIAIDELYLGKDRENEQFHLSDMISKYYRHELDFES
: :::||:||||:| ::|  | ||:|  ::::  :   : ||||:|||      |:    | : ::: |  : |
VLIEMSGQGCRQFESFLECRKKTWYD---FFQDCMQQGGSFTRFDLAID------DK-KTYFSIPELLKKAQKGEC-ISR
      100       110       120       130              140       150

1017      1047      1077      1107      1137      1167      1197      1227
LRTWNYIGGGSLNFSDMEEIEQNRQGISLYFGSRQSEMYFNFYEKRYEIAKQEGITVEEALEIFELWNRYEIRLSQSKAN
:|  ::   ||:::||         | ::||||::|| |: |||| || |::  | :||    :  |||||:||   :|
FRKSDF--NGSFDLSD-----GITGGTTIYFGSKKSEAYLCFYEKNYEQAEKYNIPLEE----LGDWNRYELRLKNERAQ
          170       180       190       200       210       220

1257      1287              1341      1371      1401      1431      1461
AAVDEFISGVPIGEISRGLIVSKIDVYDG-KN-EYGSFQADRKWQLMFGGVEPLKFVTKPEAYSIERTLRWLSDSVSPSL
 |:| ::   :  |:  :| : :   |  :|     ::     |    | |  | :  ||:    :::  || :| :|::
VAIDALLKTKDLTLIAMQIINNYVRFVDADENITREHWKTSLFWSDFIGDVGRLPLYVKPQKDFYQKSRNWLRNSCAPTM
```

```
          240       250       260       270       280       290       300

1491      1521      1551      1581      1611      1641      1671      1701
AMIREYDMIVDGDYLQTILNSGEVNERGEKILDSIKASLGIL*EVSFVLYSNREFAYCVNRRNNLDKMIDLLVFMIPDRE
 |: | |  :    |  ::   |: :: :|:||   | :  :
KMVLEADEHLGKTDLSDMIAEAELADKHKKMLDVYMADVADMVV
          320       330       340       350
```

SEQ ID 8604 (GBS294) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 167 (lane 6 & 7; MW 65kDa – thioredoxin fusion), in Figure 238 (lane 2; MW 65kDa) and in Figure 40 (lane 6; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 3; MW 76kDa).

Purified Thio-GBS294-His is shown in Figure 244, lane 2.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 578

A DNA sequence (GBSx0618) was identified in *S.agalactiae* <SEQ ID 1807> which encodes the amino acid sequence <SEQ ID 1808>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.61   Transmembrane   24 -  40 (  20 -  41)
   INTEGRAL    Likelihood = -1.97   Transmembrane   53 -  69 (  52 -  72)

----- 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 GENPEPT database:

```
>GP:AAB60012 GB:U09422 ORF21 [Enterococcus faecalis]
 Identities = 136/473 (28%), Positives = 228/473 (47%), Gaps = 40/473 (8%)

Query: 9   RGIKVKPYMRYMSYYL-FSFLFILFLTPVGVYSYYYLDL-------LKMMDKMSM----I 56
           RG +++P  + + ++   + L  +FL  VG++    +          L   DK+ +    I
Sbjct: 4   RGKRIRPSGKDLVFHFTIASLLPVFLLVVGLFHVKTIQQINWQDFNLSQADKIDIPYLII 63

Query: 57  SVGTGLFLAFFVSWYLTWFLQEANPLFNKLDRLKRMSKFLYENGYVYEKR-------KKS 109
           S    + +   V++    F +       +L   ++++K + EN +   ++        K S
Sbjct: 64  SFSVAILICLLVAFV---FKRVRYDTVKQLYHRQKLAKMILENKWYESEQVKTEGFFKDS 120

Query: 110 NKKTKTKYR-FPKVYVKQGKYDLSVSFEMAGGKFQKKFKDIGGELEDTFFMDFMEKTDDP 168
             +TK K   FPK+Y +     + +  E+  GK+Q +   +  +LE   + +  +K
Sbjct: 121 AGRTKEKITYFPKMYYRLKNGLIQIRVEITLGKYQDQLLHLEKKLESGLYCELTDKELKD 180

Query: 169 RFKIYKLAYSAFLSRITVKDVIWNKDKGIKLMDGYYWDFINDPHLLVAGGTGGGKTVLLR 228
            + Y L Y    SRI++ D +  KD  ++LM   +W++   PH+L+AGGTGGGKT  +
Sbjct: 181 SYVEYTLLYDTIASRISI-DEVEAKDGKLRLMKNVWWEYDKLPHMLIAGGTGGGKTYFIL 239

Query: 229 SILRCLAEI-GVCDICDPKRADFVTMSDLSAFEGRIAFEKADIIEKFENAVTIMFARYDF 287
           +++  L        I DPK AD   ++DL +   + + K D++   E     M  R +
Sbjct: 240 TLIEALLHTDSKLYILDPKNAD---LADLGSVMANVYYRKEDLLSCIETFYEEMMKRSE- 295

Query: 288 VRNEMKRLGHKDMKKFYDY-GLEPYFFVCDEYNALMSSLSYQEREIVDNAFIQYILLGRQ 346
              EMK++ +    K Y Y GL  +F + DEY A M  L  +E   V N   Q ++LGRQ
Sbjct: 296 ---EMKQMKNYKTGKNYAYLGLPAHFLIFDEYVAFMEMLGTKENTAVMNKLKQIVMLGRQ 352

Query: 347 VGCNAIIAMQKPSADDLPTKIRSNMMHHISVGRLDDGGYVMMFGDENRNKEFRFIKYLAG 406
            G   I+A Q+P A  L   IR      +++GR+ + GY MMFG + + K+F F+K
```

```
Sbjct: 353 AGFFLILACQRPDAKYLGDGIRDQFNFRVALGRMSEMGYGMMFGSDVQ-KDF-FLK---- 406

Query: 407 RRVYGRGYSAVFGEVAREFYSPLLPKNFSFYDAFEKINRHENPFDPTENQEVS 459
           R+ GRGY  V   V  EFY+PL+PK + F +  +K++        T   EV+
Sbjct: 407 -RIKGRGYVDVGTSVISEFYTPLVPKGYDFLEEIKKLSNSRQSTQATCEAEVA 458
```

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

A related GBS gene <SEQ ID 8605> and protein <SEQ ID 8606> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 8
McG: Discrim Score:    -10.05
GvH: Signal Score (-7.5): -3.42
     Possible site: 40
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value:  -3.61 threshold:  0.0
   INTEGRAL    Likelihood = -3.61   Transmembrane   24 -  40 (  20 -  41)
   INTEGRAL    Likelihood = -1.97   Transmembrane   53 -  69 (  52 -  72)
   PERIPHERAL  Likelihood =  1.01     224
 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:

```
29.9/52.7% over 456aa
                                                                Enterococcus faecalis
  EGAD|17035| hypothetical protein Insert characterized
  GP|532554|gb|AAB60012.1||U09422 ORF21 Insert characterized

ORF00100(319 - 1677 of 2316)
EGAD|17035|17250(2   -   458   of   461)   hypothetical   protein   {Enterococcus   faecalis}
GP|532554|gb|AAB60012.1||U09422 ORF21 {Enterococcus faecalis}
%Match = 11.2
%Identity = 29.9  %Similarity = 52.7
Matches = 135  Mismatches = 199  Conservative Sub.s = 103

207       237       267       297       327       357       384       414
FQVVCLKFLHHHLRKRMLQIMETHQKMKHLKLINKR*RRGNLARLIPQYRGIKVKPYMRYMSYYL-FSFLFILFLTPVGV
                                          : || :::|  : : :::  : |: :||  ||:
                                         MKQRGKRIRPSGKDLVFHFTIASLLPVFLLVVGL
                                               10        20        30

426       453       483       513                 570       600
Y------SYYYLDL-LKMMDKMSMISVGTGLFLAFFVSWYLTWFLQEAN-PLFNKLDRLKRMSKFLYENGYVYEKR----
:         : |: |   ||: :  :   : :| ::   : : ::        :|   ::::| : || : ||
FHVKTIQQINWQDFNLSQADKIDIPYLIISFSVAILICLLVAFVFKRVRYDTVKQLYHRQKLAKMILENKW-YESEQVKT
              50        60        70        80        90        100       110

636       663       693       723       753       783       813       843
----KKSNKKTKTKYR-FPKVYVKQGKYDLSVSFEMAGGKFQKKFKDIGGELEDTFFMDFMEKTDDPRFKIYKLAYSAFL
    | |  :|| |   |||:| :     : :  |:  ||:| ::  :  :|| :: :: :|     :  | | |
EGFFKDSAGRTKEKITYFPKMYYRLKNGLIQIRVEITLGKYQDQLLHLEKKLESGLYCELTDKELKDSYVEYTLLYDTIA
              130       140       150       160       170       180       190

873       903       933       963       993       1020      1050      1080
SRITVKDVIWNKDKGIKLMDGYYWDFINDPHLLVAGGTGGGKTVLLRSILRCLAEI-GVCDICDPKRADFVTMSDLSAFE
|||:: | :  ||  ::||   :|::   ||:|:|||||||| :: :::  |       | ||| ||   ::|| :
SRISI-DEVEAKDGKLRLMKNVWWEYDKLPHMLIAGGTGGGKTYFILTLIEALLHTDSKLYILDPKNAD---LADLGSVM
              210       220       230       240       250       260
```

```
1110       1140       1170       1200       1227       1257       1287       1317
GRIAFEKADIIEKFENAVTIMFARYDFVRNEMKRLGHKDMKKFYDY-GLEPYFFVCDEYNALMSSLSYQEREIVDNAFTQ
  : : | |::   |      |  |      |||:: :    | | | ||  :|:: ||| |:|  |  :|   | | : |
ANVYYRKEDLLSCIETFYEEMMKR----SEEMKQMKNYKTGKNYAYLGLPAHFLIFDEYVAFMEMLGTKENTAVMNKLKQ
         280        290        300        310        320        330        340

1347       1377       1407       1437       1467       1497       1527       1557
YILLGRQVGCNAIIAMQKPSADDLPTKIRSNMMHHISVGRLDDGGYVMMFGDENRNKEFRFIKYLAGRRVYGRGYSAVFG
 ::|||| |   |:| |:| | |   ||      :::||: : || |||| :   |:| |:| : ||      ||  |
IVMLGRQAGFFLILACQRPDAKYLGDGIRDQFNFRVALGRMSEMGYGMMFGSD-VQKDF-FLKRIKGR-----GYVDVGT
             360        370        380        390        400        410

1587       1617       1647       1677       1707       1737       1767       1797
EVAREFYSPLLPKNFSFYDAFEKINRHENPFDPTENQEVSKAILKDESLREFVEKTSENELLKGSVGFDFDDEMEEMENM
 |  |||:||:|| : | :  :|::       |   ||:
SVISEFYTPLVPKGYDFLEEIKKLSNSRQSTQATCEAEVAGVD
          430        440        450        460
```

SEQ ID 8606 (GBS216) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 42 (lane 3; MW 66.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 2; MW 91kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 579

A DNA sequence (GBSx0619) was identified in *S.agalactiae* <SEQ ID 1809> which encodes the amino acid sequence <SEQ ID 1810>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4095(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 580

A DNA sequence (GBSx0620) was identified in *S.agalactiae* <SEQ ID 1811> which encodes the amino acid sequence <SEQ ID 1812>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0944(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10219> which encodes amino acid sequence <SEQ ID 10220> was also identified.

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

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

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

Example 581

A DNA sequence (GBSx0621) was identified in *S.agalactiae* <SEQ ID 1813> which encodes the amino acid sequence <SEQ ID 1814>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.94   Transmembrane  810 - 826 ( 808 - 830)

----- Final Results -----
               bacterial membrane --- Certainty=0.2975(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 GENPEPT database:

```
!GB:D90354 surface protein antigen precursor [Strept...

>GP:BAA14368 GB:D90354 surface protein antigen precursor
             [Streptococcus sobrinus]
 Identities = 151/408 (37%), Positives = 219/408 (53%), Gaps = 27/408 (6%)

Query: 451  PSKAVIDEAGQSVNGKTVLPNAELNYVAKQDFSQYKGMTASQGKIAKNFVFIDDYKDDAL 510
            P K   +E G  ++GK+VL      Y    D  QYKG  +++  I K F ++DDY ++AL
Sbjct: 1162 PHKVNKNENGVVIDGKSVLAGTTNYYELTWDLDQYKGDKSAKETIQKGFFYVDDYPEEAL 1221

Query: 511  DGKSMKVNSIKASDGTDVSQL-LEMRHVLSTDTLDEKLQTLIKEAGISPVGEFYMWTAKD 569
            D   ++ + IK +D    +   + +    S +     +Q ++K+A I+P G F ++TA D
Sbjct: 1222 D---LRTDLIKLTDANGKAVTGVSVADYASLEAAPAAVQDMLKKANITPKGAFQVFTADD 1278

Query: 570  PQAFYKAYVQKGLDVTYNLSFKVKKEFTK--GQIQNGVAQIDFGNGYTGNIVVNDLTTPE 627
            PQAFY AYV  G D+T      VK E  K  G  +N   QIDFGNGY  NIV+N++
Sbjct: 1279 PQAFYDAYVVTGTDLTIVTPMTVKAEMGKIGGSYENKAYQIDFGNGYESNIVINNVPQIN 1338

Query: 628  IHKDV---LDKEDGKSINNGTVKLGDEVTYKLEGWVVPTGRSYDLFEYKFVDQLQRTPDL 684
              KDV   +D  D  +++  T+ L     Y+L G ++P   + +LFEY F D   +T D
Sbjct: 1339 PEKDVTLTMDPADSTNVDGQTIALNQVFNYRLIGGIIPADHAEELFEYSFSDDYDQTGDQ 1398

Query: 685  YLRD-KVVAKVDVTLKDGTVIKKGTNLGEYTETVYNKKTGLYELVFKKDFLEKVARSSEF 743
            Y    K  AKVD+TLKDGT+IK GT+L  YTE   ++  G  + FK+DFL  V+  S F
Sbjct: 1399 YTGQYKAFAKVDLTLKDGTIIKAGTDLTSYTEAQVDEANGQIVVTFKEDFLRSVSVDSAF 1458

Query: 744  GADDFVVVKRIKAGDVYNTADFFINGNKVKTETVVTHTPE--KPKPVEPQ---------- 791
             A+ ++ +KRI  G   NT    +NG    + TV T TPE  +P PV+P+
Sbjct: 1459 QAEVYLQMKRIAVGTFANTYVNTVNGITYSSNTVRTSTPEPKQPSPVDPKTTTTVVFQPR 1518

Query: 792  --KATPKAPAKG--LPQTGEASVAPLTALGAIILSA-IGLAGFKKRKE 834
              KA   AP  G  LP TG++S A L  LG + L+A   L G +++++
Sbjct: 1519 QGKAYQPAPPAGAQLPATGDSSNAYLPLLGLVSLTAGFSLLGLRRKQD 1566
 Identities = 75/242 (30%), Positives = 120/242 (48%), Gaps = 33/242 (13%)

Query: 11   SADQVTTQATTQTVTQNQAETVTSTQLDKAVATAKKAAVAVTTTAAVNHATTTDAQADLA 70
            S+   T+QA     T   +  V++++LD+A  +A++A V V+  A VN  T   +  D A
Sbjct: 73   SSQAETSQAQAGQKTGAMSVDVSTSELDEAAKSAQEAGVTVSQDATVNKGTVETS--DEA 130

Query: 71   NQTQT-VKDVTAKAQANTQAIKDATAENAKIDAENKAESQRVSQLNAQTKAKID---AEN 126
            NQ +T +KD   +K  A+   I+  T +    A N+AE+ R++Q NA  KA+ +   A N
Sbjct: 131  NQKETEIKDDYSKQAAD---IQKTTEDYKAAVAANQAETDRITQENAAKKAQYEQDLAAN 187
```

```
Query: 127 KDAQAKADATNAQLQKDYQAKLAKIKSVEAYNAGVRQRNKDAQA--------------KA 172
           K    +    NAQ + DY+AKLA+ +   A    V+Q N D+QA               +
Sbjct: 188 KAEVERITNENAQAKADYEAKLAQYQKDLA---AVQQANNDSQAAYAAAKEAYDKELARV 244

Query: 173 DATNAQLQKDYQAKLA---LYNQALKAKAEADKQSINNVAFDIKAQ----AKGVDNAEYG 225
            A NA  +K+Y+  LA     N+ +KA+  A +Q       D +A+     K +  A+ G
Sbjct: 245 QAANAAAKKEYEEALAANTTKNEQIKAENAAIQQRNAQAKADYEAKLAQYEKDLAAAQSG 304

Query: 226 NS 227
           N+
Sbjct: 305 NA 306
 Identities = 63/223 (28%), Positives = 100/223 (44%), Gaps = 31/223 (13%)

Query: 2   ITTLQTSQVSADQVTTQATTQTVTQNQAETVTSTQLDKAVATAK-----------KAAVA 50
           +  +Q +  +A +    +A      T+N+      +  + +  A AK           K   A
Sbjct: 241 LARVQAANAAAKKEYEEALAANTTKNEQIKAENAAIQQRNAQAKADYEAKLAQYEKDLAA 300

Query: 51  VTTTAAVNHATTTDAQADLANQTQTVKDVTAKA-QANTQAIKDATAENAKIDAENKAESQ 109
             + A N A     +A    +   V+   A A QA  QA+   TA+NA+I AEN+A  Q
Sbjct: 301 AQSGNATNEADYQAKKAAYEQELARVQAANAAAKQAYEQALAANTAKNAQITAENEAIQQ 360

Query: 110 RVSQLNAQTKAKIDAENKDAQAKADATNAQLQKDYQAKLA----KIKSVEAYNAGVRQRN 165
           R +Q  A  +AK+    KD  A A + NA  + DYQ KLA    ++  V+A NA  +Q
Sbjct: 361 RNAQAKANYEAKLAQYQKDL-AAAQSGNAANEADYQEKLAAYEKELARVQAANAAAKQEY 419

Query: 166 KDAQAKADATNAQL--------------QKDYQAKLALYNQAL 194
           +    +A+A NA++               + DY+ KL+ Y + L
Sbjct: 420 EQKVQEANAKNAEITEANRAIRERNAKAKTDYELKLSKYQEEL 462
 Identities = 75/243 (30%), Positives = 101/243 (40%), Gaps = 56/243 (23%)

Query: 8   SQVSAD-QVTTQATTQTVTQNQAETVTSTQLDKAVATAKKAAVAVTTTAAVNHATTTDAQ 66
           S+ +AD Q TT+     V  NQAET   TQ + A   A+         A V   T  +AQ
Sbjct: 142 SKQAADIQKTTEDYKAAVAANQAETDRITQ-ENAAKKAQYEQDLAANKAEVERITNENAQ 200

Query: 67  ADL---ANQTQTVKDVTAKAQANT--------------------------------QAIK 91
           A     A    Q  KD+ A  QAN                                 +A+
Sbjct: 201 AKADYEAKLAQYQKDLAAVQQANNDSQAAYAAAKEAYDKELARVQAANAAAKKEYEEALA 260

Query: 92  DATAENAKIDAENKAESQRVSQLNAQTKAKIDAENKDAQAKADATNAQLQKDYQAKLA-- 149
             T +N +I AEN A  QR +Q  A  +AK+    KD  A A + NA  + DYQAK A
Sbjct: 261 ANTTKNEQIKAENAAIQQRNAQAKADYEAKLAQYEKDL-AAAQSGNATNEADYQAKKAAY 319

Query: 150 --KIKSVEAYNAGVRQRNKDAQAKADATNAQL--------------QKDYQAKLALYNQA 193
             ++  V+A NA  +Q  + A A   A NAQ+               + +Y+AKLA Y +
Sbjct: 320 EQELARVQAANAAAKQAYEQALAANTAKNAQITAENEAIQQRNAQAKANYEAKLAQYQKD 379

Query: 194 LKA 196
           L A
Sbjct: 380 LAA 382
```

There is also homology to SEQ ID 598.

SEQ ID 1814 (GBS191) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 176 (lane 2; MW 91kDa).

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

Example 582

A DNA sequence (GBSx0622) was identified in *S.agalactiae* <SEQ ID 1815> which encodes the amino acid sequence <SEQ ID 1816>. This protein is predicted to be TnpA. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2935(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10221> which encodes amino acid sequence <SEQ ID 10222> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9921> which encodes amino acid sequence <SEQ ID 9922> was also identified.

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

```
>GP:AAC82523 GB:AF027768 TnpA [Serratia marcescens]
 Identities = 168/385 (43%), Positives = 232/385 (59%), Gaps = 13/385 (3%)

Query: 26  MMFKVEAVGPPERCPECGFD-KLYKHSSRNQLIMDLPIRLKRVGLHLNRRRYKCRECGST 84
           M F+V+ V   P   C ECG    +  +    R+     DLPI   KRV L + RRRY CR C +T
Sbjct: 1   MHFQVD-VPDPIACEECGVQGEFVRFGKRDVPYRDLPIHGKRVTLWVVRRRYTCRACKTT 59

Query: 85  IS------VDEKRSMTKRLLKSIQEQSMSKTFVEVAESVGVDEKTIRNVFKDYVALKERE 138
                  VD  R MT RL + ++++S +  +  VA    G+DEKT+R++F         R
Sbjct: 60  FRPQLPEMVDGFR-MTLRLHEYVEKESFNHPYTFVAAQTGLDEKTVRDIFNARAEFLGRW 118

Query: 139 YQFETPKWLGIDEIHIIRRPRLVLTNIERRTIYDIKPNRNKETVIQRLSEISDRTYIEYV 198
           ++FETP+ LGIDE+++ +R R +LTNIE RT+ D+    R ++ V    L ++ DR  +E V
Sbjct: 119 HRFETPRILGIDELYLNKRYRCILTNIEERTLLDLLATRRQDVVTNYLMKLKDRQKVEIV 178

Query: 199 TMDMWKPYKDAVNTILPQAKVVVDKFHVVRMANQALDNVRKSLKAHMSQKERRTLMRERF 258
           +MDMW PY+ AV  +LPQA++VVDKFHVVRMAN AL+ VRK L+  +   + RTL  +R
Sbjct: 179 SMDMWNPYRAAVKAVLPQARIVVDKFHVVRMANDALERVRKGLRKELKPSQSRTLKGDRK 238

Query: 259 ILLKRKHDLNERESFLLDTWLGNLPALKEAYELKEEFYWIWDTPDPDEGHLRYSQWRHRC 318
           ILLKR H++++RE  +++TW G  P L  AYE KE FY IWD     +      +W
Sbjct: 239 ILLKRAHEVSDRERLIMETWTGAFPQLLAAYEHKERFYGIWDATTRLQAEAALDEWI-AT 297

Query: 319 MSSNSKDAYKDLVRAVDNWHVEIFNYF--DKRLTNAYTESINSIIRQVERMGRGYSFDAL 376
           +    K+ + DLVRAV NW  E   YF  D  +TNAYTESIN + +   R GRGYSF+ +
Sbjct: 298 IPKGQKEVWSDLVRAVGNWREETMTYFETDMPVTNAYTESINRLAKDKNREGRGYSFEVM 357

Query: 377 RAKILFNEKLHKKRKPRFNSSAFNK 401
           RA++L+  K HKK+ P    S F K
Sbjct: 358 RARMLYTTK-HKKKAPTAKVSPFYK 381
```

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 583

A DNA sequence (GBSx0623) was identified in *S.agalactiae* <SEQ ID 1817> which encodes the amino acid sequence <SEQ ID 1818>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2115(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 GENPEPT database:

```
>GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]
 Identities = 411/546 (75%), Positives = 483/546 (88%)

Query: 1   MNKFKVNISGMTCTGCEKHVESALEKIGAKNIESSYRRGEAVFELPDDIEVESAIKAIDE 60
            M K++V++ GMTCTGCE+HV  ALE +GA  IE  +RRGEAVFELP+ + VE+A KAI +
Sbjct: 1   MKKYRVDVQGMTCTGCEEHVAVALENMGATGIEVDFRRGEAVFELPNALGVETAKKAISD 60

Query: 61  ANYQAGEIEEVSSLENVALINEDNYDLLIIGSGAAAFSSAIKAIEYGAKVGMIERGTVGG 120
            A YQ G+ EEV S E V L NE +YD +IIGSG AAFSSAI+A++YGAKV MIERGT+GG
Sbjct: 61  AKYQPGKAEEVQSQEMVQLGNEGDYDYIIIGSGGAAFSSAIEAVKYGAKVAMIERGTIGG 120

Query: 121 TCVNIGCVPSKTLLRAGEINHLSKDNPFIGLQTSAGEVDLASLITQKDKLVSELRNQKYM 180
            TCVNIGCVPSKTLLRAGEINHL+K+NPF+GL TSAGEVDLA LI QK++LV+ELRN KY+
Sbjct: 121 TCVNIGCVPSKTLLRAGEINHLAKNNPFVGLHTSAGEVDLAPLIKQKNELVTELRNSKYV 180

Query: 181 DLIDEYNFDLIKGEAKFVDASTVEVNGTKLSAKRFLIATGASPSLPQISGLEKMDYLTST 240
            DLID+Y F+LI+GEAKFVD  TVEVNG  +SAKRFLIATGASP+ P I GL ++DYLTST
Sbjct: 181 DLIDDYGFELIEGEAKFVDEKTVEVNGAPISAKRFLIATGASPAKPNIPGLNEVDYLTST 240

Query: 241 TLLELKKIPKRLTVIGSGYIGMELGQLFHHLGSEITLMQRSERLLKEYDPEISESVEKAL 300
            +LLELKK+PKRL VIGSGYIGMELGQLFH+LGSE+TL+QRSERLLKEYDPEISESVEK+L
Sbjct: 241 SLLELKKVPKRLVVIGSGYIGMELGQLFHNLGSEVTLIQRSERLLKEYDPEISESVEKSL 300

Query: 301 IEQGINLVKGATFERVEQSGEIKRVYVTVNGSREVIESDQLLVATGRKPNTDSLNLSAAG 360
            +EQGINLVKGAT+ER+EQ+G+IK+V+V VNG + +IE+DQLLVATGR PNT +LNL AAG
Sbjct: 301 VEQGINLVKGATYERIEQNGDIKKVHVEVNGKKRIIEADQLLVATGRTPNTATLNLRAAG 360

Query: 361 VETGKNNEILINDFGQTSNEKIYAAGDVTLGPQFVYVAAYEGGIITDNAIGGLNKKIDLS 420
            VE G   EI+I+D+ +T+N +IYAAGDVTLGPQFVYVAAY+GG+   NAIGGLNKK++L
Sbjct: 361 VEIGSRGEIIIDDYSRTTNTRIYAAGDVTLGPQFVYVAAYQGGVAAPNAIGGLNKKLNLE 420

Query: 421 VVPAVTFTNPTVATVGLTEEQAKEKGYDVKTSVLPLGAVPRAIVNRETTGVFKLVADAET 480
            VVP VTFT P +ATVGLTE+QAKE GY+VKTSVLPL AVPRA+VNRETTGVFKLVAD++T
Sbjct: 421 VVPGVTFTAPAIATVGLTEQQAKENGYEVKTSVLPLDAVPRALVNRETTGVFKLVADSKT 480

Query: 481 LKVLGVHIVSENAGDVIYAASLAVKFGLTIEDLTETLAPYLTMAEGLKLVALTFDKDISK 540
            +KVLG H+V+ENAGDVIYAA+LAVKFGLT++D+ ETLAPYLTMAEGLKL ALTFDKDISK
Sbjct: 481 MKVLGAHVVAENAGDVIYAATLAVKFGLTVDDIRETLAPYLTMAEGLKLAALTFDKDISK 540

Query: 541 LSCCAG 546
            LSCCAG
Sbjct: 541 LSCCAG 546
```

There is also homology to SEQ ID 1820.

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

Example 584

A DNA sequence (GBSx0624) was identified in *S.agalactiae* <SEQ ID 1821> which encodes the amino acid sequence <SEQ ID 1822>. This protein is predicted to be regulatory protein. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4529(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 GENPEPT database:

```
>GP:AAA83973 GB:AF138877 mercury resistance operon negative
          regulator MerR1 [Bacillus sp. RC607]
 Identities = 84/129 (65%), Positives = 105/129 (81%)

Query: 1   MIYRISEFADKCGVNKETIRYYERKNLLQEPHRTEAGYRIYSYDDVKRVGFIKRIQELGF 60
           M +RI E ADKCGVNKETIRYYER  L+ EP RTE GYR+YS   V R+ FIKR+QELGF
Sbjct: 1   MKFRIGELADKCGVNKETIRYYERLGLIPEPERTEKGYRMYSQQTVDRLHFIKRMQELGF 60

Query: 61  SLSEIYKLLGVVDKDEVRCQDMFEFVSKKQKEVQKQIEDLKRIETMLDDLKQRCPDEKKL 120
           +L+EI KLLGVVD+DE +C+DM++F   K +++Q++IEDLKRIE ML DLK+RCP+ K +
Sbjct: 61  TLNEIDKLLGVVDRDEAKCRDMYDFTILKIEDIQRKIEDLKRIERMLMDLKERCPENKDI 120

Query: 121 HSCPIIETL 129
           + CPIIETL
Sbjct: 121 YECPIIETL 129
```

There is also homology to SEQ ID 1712.

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

Example 585

A DNA sequence (GBSx0625) was identified in *S.agalactiae* <SEQ ID 1823> which encodes the amino acid sequence <SEQ ID 1824>. This protein is predicted to be Nramp metal ion transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.85   Transmembrane  175 - 191 ( 169 - 201)
   INTEGRAL    Likelihood =-11.94   Transmembrane  150 - 166 ( 132 - 173)
   INTEGRAL    Likelihood = -9.45   Transmembrane  491 - 507 ( 481 - 509)
   INTEGRAL    Likelihood = -8.92   Transmembrane  375 - 391 ( 374 - 396)
   INTEGRAL    Likelihood = -8.39   Transmembrane   72 -  88 (  69 -  93)
   INTEGRAL    Likelihood = -7.96   Transmembrane  280 - 296 ( 274 - 299)
   INTEGRAL    Likelihood = -7.17   Transmembrane  413 - 429 ( 411 - 431)
   INTEGRAL    Likelihood = -6.79   Transmembrane  327 - 343 ( 322 - 346)
   INTEGRAL    Likelihood = -3.40   Transmembrane  444 - 460 ( 443 - 462)
   INTEGRAL    Likelihood = -3.24   Transmembrane  132 - 148 ( 132 - 149)
   INTEGRAL    Likelihood = -0.96   Transmembrane  115 - 131 ( 114 - 131)

----- Final Results -----
              bacterial membrane --- Certainty=0.6540(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 GENPEPT database:

```
>GP:AAF83825 GB:AE003939 manganese transport protein [Xylella
          fastidiosa]
 Identities = 185/450 (41%), Positives = 278/450 (61%), Gaps = 29/450 (6%)

Query: 16  ANGPSLEEINGTIEVPKDLSFFKTLLAYSGPGALVAVGYMDPGNWSTSITGGQNFQYLLI 75
           ++ PSL E++ ++ V +   +   LLA+ GPG +V+VGYMDPGNW+T + GG  F Y+L+
Sbjct: 35  SDSPSLGEMHASVAVSRRGHWGFRLLAFLGPGYMVSVGYMDPGNWATGLAGGSRFGYMLL 94

Query: 76  SIILMSSLIAMLLQYMSAKLGIVTQMDLAQAIRARTSKQLGIVLWILTELAIMATDIAEV 135
           S+IL+S+++A++LQ ++A+LGI + MDLAQA RAR S+   + LW++ ELAI+A D+AEV
Sbjct: 95  SVILLSNVMAIVLQALAARLGIASDMDLAQACRARYSRGTTLALWVVCELAIIACDLAEV 154

Query: 136 IGGAIALYLLFHIPLAIAVFITVFDVLLLLLLTKIGFRKIEALVVALILVIFLVFAYQVA 195
```

```
          IG AIAL LL  +P+   V IT  DV+L+LLL   GFR +EA V+AL+LVIF  F  Q+
Sbjct: 155 IGTAIALNLLLGVPIIWGVVITAVDVVLVLLLMHRGFRALEAFVIALLLVIFGCFVVQIV 214

Query: 196 LSHPIWTDIFKGLVPTSEAFSTSHTVNGQTPLSGALGIIGATVMPHNLYLHSSVVQSRKL 255
           L+ P   ++  G VP  +       V     L  A+GI+GATVMPHNLYLHSS+VQ+R
Sbjct: 215 LAAPPLQEVLGGFVPRWQ------VVADPQALYLAIGIVGATVMPHNLYLHSSIVQTRAY 268

Query: 256 DHNNKKDIAR--AIRFSTFDSNIQLTVAFFVNSLLLIMGVAVFKTGSVTDPSFFGLFKAL 313
               +  + R  A+R++  DS + L +A F+N+ +LI+  AVF       D
Sbjct: 269 P---RTPVGRRSALRWAVADSTLALMLALFINASILILAAAVFHAQHHFD---------- 315

Query: 314 SNSTIMSNSILAHIASSGILSLLFAIALLASGQNSTITGTLTGQIIMEGFIHMKVPIWFR 373
                +  +LA +   G+ + LFA ALLASG NST+T TL GQI+MEGF+ +++  W R
Sbjct: 316 VEEIEQAYQLLAPVLGVGVAATLFATALLASGINSTVTATLAGQIVMEGFLRLRLRPWLR 375

Query: 374 RIITRLISVIPVMICVLVTSGRSTVEEHIAINNLMNNSQVFLAFALPFSMLPLLIFTNSK 433
           R++TR ++++PV++ V +   + T          L+  SQV L+  LPF+++PLL     +
Sbjct: 376 RVLTRGLAIVPVIVVVALYGEQGT-------GRLLLLSQVILSMQLPFAVIPLLRCVADR 428

Query: 434 VEMDDDFKNTWIIKILGWLSVIGLIYLNMK 463
             M       W++ ++ WL    ++ LN+K
Sbjct: 429 KVMGALVAPRWLM-VVAWLIAGVIVVLNVK 457
```

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

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

Example 586

A DNA sequence (GBSx0626) was identified in *S.agalactiae* <SEQ ID 1825> which encodes the amino acid sequence <SEQ ID 1826>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2590(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 587

A DNA sequence (GBSx0627) was identified in *S.agalactiae* <SEQ ID 1827> which encodes the amino acid sequence <SEQ ID 1828>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.82   Transmembrane  212 - 228 ( 204 - 233)
   INTEGRAL    Likelihood = -8.39   Transmembrane   98 - 114 (  94 - 125)
   INTEGRAL    Likelihood = -7.22   Transmembrane  132 - 148 ( 122 - 154)
   INTEGRAL    Likelihood = -6.42   Transmembrane  159 - 175 ( 155 - 188)
   INTEGRAL    Likelihood = -4.78   Transmembrane   54 -  70 (  51 -  72)
   INTEGRAL    Likelihood = -2.97   Transmembrane   18 -  34 (  15 -  36)

----- Final Results -----
```

```
bacterial membrane --- Certainty=0.4927(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 GENPEPT database:

```
>GP:CAB16051 GB:Z99124 yydJ [Bacillus subtilis]
 Identities = 97/239 (40%), Positives = 154/239 (63%), Gaps = 3/239 (1%)

Query: 4   LEFRKSIRGRTLFYIISTVALTYVLGYILPVGIDKIRHLTLGEFYFSTYTVFTQFGFLIF 63
           LEF+KSI  + +  + +      ++LGY L VGIDK+ ++T   F+FS+YTV TQFG ++F
Sbjct: 3   LEFKKSISNKVIIILGAMFVFLFLLGYFLLVGIDKVSNVTPEMFFFSSYTVATQFGLMLF 62

Query: 64  GFVIVYFFNKDYSDKCILYHYFSGYHLTKYFYTKLLVLFSEFFIAIIVCNILASLLWGYS 123
            FVI +F N++YS+K IL++   G ++  +FY K+ VLF E F  I +  ++ SL++ +
Sbjct: 63  SFVIAFFINREYSNKNILFYKLIGENIYTFFYKKIAVLFLECFAFITLGLLIISLMY-HD 121

Query: 124 LFYFLTTTILFSLVVLQYLLVVSTISILFSNMLVSIGVTIFYWITSIILVAIGG-IFKVS 182
             +F     LFS V+LQY+L++ TIS+L  N+L+SIGV+I YW+TS+ILVAI    F
Sbjct: 122 FSHFALLLFLFSAVILQYILIIGTISVLCPNILISIGVSIVYWMTSVILVAISNKTFGFI 181

Query: 183 AIFDASNSLYKIIGK-LFSHPMTIDLTDFFIIVPYMICLSVISFLIVCLSNRRWLLNGM 240
           A F+A N++Y  I + L S  MT+   D   I+ Y++ + +I+ +++  S  RW+  G+
Sbjct: 182 APFEAGNTMYPRIERVLQSDNMTLGSNDVLFIILYLVSIIIINAIVLRFSKTRWIKMGL 240
```

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

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

Example 588

A DNA sequence (GBSx0628) was identified in *S.agalactiae* <SEQ ID 1829> which encodes the amino acid sequence <SEQ ID 1830>. This protein is predicted to be antibiotic epidermin immunity protein F. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2901(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 GENPEPT database:

```
>GP:CAB16052 GB:Z99124 similar to ABC transporter (ATP-binding
          protein) [Bacillus subtilis]
 Identities = 100/209 (47%), Positives = 150/209 (70%), Gaps = 4/209 (1%)

Query: 1   MFINNYTLKIGNRILLENTNLDFEEGEINHLLGRNGSGKSQLAKDFIINRGNYFSNDIYE 60
           M I NYTLK+  + LL++T+L F  G+INH++G+NG GKSQLAKDF++N       DI +
Sbjct: 1   MNIANYTLKVKGKTLLQDTDLHFSSGKINHVVGKNGVGKSQLAKDFLLNNSKRIGRDIRQ 60

Query: 61  DTLIISSYSNLPSDVT----INDLERTIPWKLSKEIYQLLNINQISKTVKLKQLSDGQKQ 116
           +  +ISS SN+P+DV+     ++ L +    K+  +I  LLN++ I   V +K LSDGQKQ
Sbjct: 61  NVSLISSSSNIPNDVSKDFLLHFLSKKFDAKMIDKIAYLLNLDNIDGKVLIKNLSDGQKQ 120

Query: 117 KVKLLVLLSLDKHIIILDEITNALDKKSVDEINVFLQNYIQYYPEKIIINISHDINNIRS 176
           K+KLL  L  DK+II+LDEITN+LDKK+V EI+ FL  YIQ  PEKIIINI+HD++++++
Sbjct: 121 KLKLLSFLLEDKNIIVLDEITNSLDKKTVIEIHGFLNKYIQENPEKIIINITHDLSDLKA 180

Query: 177 LKGNYFLIDNQKICKVDTLDDAISWYLGE 205
           ++G+Y++ ++Q+I +  ++D  I  Y+ E
Sbjct: 181 IEGDYYIFNHQEIQQYHSVDKLIEVYINE 209
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1831> which encodes the amino acid sequence <SEQ ID 1832>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2760(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 49/174 (28%), Positives = 82/174 (46%), Gaps = 27/174 (15%)

Query: 3   INNYTLKIGNRILLENTNLDFEEGEINHLLGRNGSGKSQLAK----------DFIINRGN 52
           I N     G R +L N N++  +G++  L+G NG+GKS + K            II  G
Sbjct: 23  IQNLKKSYGKRTILNNVNMNIPKGKVYALIGPNGAGKSTIMKILTGLVSKTSGSIIFEGR 82

Query: 53  YFS-------NDIYEDTLI---ISSYSNLPSDVTINDL-ERTIPWKLSKEIYQLLNINQI 101
            +S         I E+  +   +S+Y N+    T+  + E TI   L+K     + +  I
Sbjct: 83  EWSRRDLRKIGSIIEEPPLYKNLSAYDNMKVVTTMLGVSESTILPLLNK-----VGLGNI 137

Query: 102 SKTVKLKQLSDGQKQKVKLLVLLSLDKHIIILDEITNALDKKSVDEINVFLQNY 155
            K   +KQ S G KQ++ + + L     ++ILDE TN LD   + E+   ++++
Sbjct: 138 DKR-PVKQFSLGMKQRLGIAISLINSPKLLILDEPTNGLDPIGIQELREIIESF 190
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 589

A DNA sequence (GBSx0629) was identified in *S.agalactiae* <SEQ ID 1833> which encodes the amino acid sequence <SEQ ID 1834>. This protein is predicted to be aminoglycoside 6-adenylyltransferase. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1780(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 GENPEPT database:

```
>GP:CAA29839 GB:X06627 ORF (str) [Staphylococcus aureus]
 Identities = 91/289 (31%), Positives = 146/289 (50%), Gaps = 14/289 (4%)

Query: 1   MRDEQEIYNLVLNIANQDKRIEAVLLNGSRANPNVPKDDFQDYDIVFVTNFIEDIISDTN 60
           MR E+EI NLV   A Q   ++ + L GSR N N+ KD FQDYD  F  + IE    + +
Sbjct: 1   MRTEKEILNLVSEFAYQRSNVKIIALEGSRTNENIKKDKFQDYDFAFFVSDIEYFTHEES 60

Query: 61  YHKKFGDILIMQKPNE---FRNKTEYNCFAYLMQFQDLTRIDLRLIKPEFLEDYLDDA-- 115
           +   FG++L +QKP +   F     +Y  ++Y+M F+D  ++D+ LI  + L  Y  D+
Sbjct: 61  WLSLFGELLFIQKPEDMELFPPDLDYG-YSYIMYFKDGIKMDITLINLKDLNRYFSDSDG 119

Query: 116 FSKVLLDKKNKYLDYNFERSSLYETKQLSEDEINKILNEIYWVSTYVVKGIARNDIIYSE 175
             K+L+DK N          S Y  K+ +E E     NE + VSTYV KG+ R +I+++
Sbjct: 120 LVKILVDKDNLVTQEIVPDDSNYWLKKPTEREFYDCCNEFWSVSTYVAKGVFRREILFAL 179

Query: 176 FMISNPIKNAFIKLLKQKILIEKELDSLSFGKLDKDILQYITDKD--QLLKIFSNKSLKD 233
             +N ++   ++++  I  + D  S GK  K I +Y+TDK+  LL  F    +
```

```
Sbjct: 180 DHFNNILRPELLRMISWYIGFNRGFD-FSLGKNYKFINKYLTDKEFNMLLATFEMNGYRK 238

Query: 234 IEANLRFLLDETNQMAKYISINRKLNLNQGEYQSAMKFMNIFLSNSYQN 282
           + +       ++ KY S N+   L    Y +  K +  F+ N+Y+N
Sbjct: 239 TYQSFKLCC----ELFKYYS-NKVSCLGNYNYPNYEKNIENFIRNNYEN 282
```

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

A related GBS gene <SEQ ID 8607> and protein <SEQ ID 8608> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:      -5.26
GvH: Signal Score (-7.5): -6.14
     Possible site: 33
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 0 value:   6.10 threshold:  0.0
   PERIPHERAL  Likelihood =  6.10     151
 modified ALOM score:  -1.72

*** Reasoning Step: 3

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1780(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:

```
31.0/53.4% over 281aa
 Staphylococcus aureus
  EGAD|9462| streptomycin resistance protein Insert characterized
  SP|P12055|STR_STAAU STREPTOMYCIN RESISTANCE PROTEIN. Insert characterized
  GP|46644|emb|CAA29839.1||X06627 ORF (str) Insert characterized
  PIR|S00938|S00938 str protein - plasmid pS194 Insert characterized

ORF00399(301 - 1146 of 1452)
EGAD|9462|9267(1 - 282 of 282) streptomycin resistance protein {Staphylococcus aureus}
SP|P12055|STR_STAAU STREPTOMYCIN RESISTANCE PROTEIN. GP|46644|emb|CAA29839.1||X06627 ORF
(str) {Staphylococcus aureus} PIR|S00938|S00938 str protein - Staphylococcus aureus plasmid
pS194
%Match = 12.8
%Identity = 31.0  %Similarity = 53.4
Matches = 87  Mismatches = 125  Conservative Sub.s = 63

117       147       177       207       237       267       297       327
**LMTY*H*TVENIWNHNQLLRKI*N*ILGGRKGMSMLI*VYDYMLREKYKGNIKVLEXTW*YKVK*EVAIMRDEQEIYN
                                                                || |:|| |
                                                               MRTEKEILN

357       387       417       447       477       507                 558
LVLNIANQDKRIEAVLLNGSRANPNVPKDDFQDYDIVFVTNFIEDIISDTNYHKKFGDILIMQKPNEFR---NKTEYNCF
||   | |   :: : | ||| | |: || |||||  |  : ||    : ::   ||::| :||| :       :|  :
LVSEFAYQRSNVKIIALEGSRTNENIKKDKFQDYDFAFFVSDIEYFTHEESWLSLFGELLFIQKPEDMELFPPDLDYG-Y
         20        30        40        50        60        70        80

588       618                 672       702       732       762       792
AYLMQFQDLTRIDLRLIKPEFLEDYLDDA--FSKVLLDKKNKYLDYNFERSSLYETKQLSEDEINKILNEIYWVSTYVVK
:|:| |:|  ::|: || : |  |: |:  : |:|:|| |          | |  |: :| |     || : ||||| |
SYIMYFKDGIKMDITLINLKDLNRYFSDSDGLVKILVDKDNLVTQEIVPDDSNYWLKKPTEREFYDCCNEFWSVSTYVAK
         100       110       120       130       140       150       160

822       852       882       912       942       966       996       1026
GIARNDIIYSEFMISNPIKNAFIKLLKQKILIEKELDSLSFGKLDKDILQYITDKD--QLLKIFSNKSLKDIEANLRFLL
|: | :|:::    :| ::  :::::   |   : :| :|:||  | | :|:|||:   ||  |     :    ::::
GVFRREILFALDHFNNILRPELLRMISWYIGFNRGFD-FSLGKNYKFINKYLTDKEFNMLLATFEMNGYRKTYQSFKLCC
         180       190       200       210       220       230       240
```

```
1056      1086      1116      1146      1176      1206      1236      1266
DETNQMAKYISINRKLNLNQGEYQSAMKFMNIFLSNSYQNFN*YYCVKDNRL*LSKLNYHS*RFSRKIINNFGDK*WDKS
:     :  :|      |     | :  | :  |: |:|:|
ELFKYYSNKVS-----CLGNYNYPNYEKNIENFIRNNYEN
               260       270       280
```

SEQ ID 1834 (GBS46) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 6; MW 34.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 3; MW 59.8kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 590

A DNA sequence (GBSx0630) was identified in *S.agalactiae* <SEQ ID 1835> which encodes the amino acid sequence <SEQ ID 1836>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1179(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 591

A DNA sequence (GBSx0631) was identified in *S.agalactiae* <SEQ ID 1837> which encodes the amino acid sequence <SEQ ID 1838>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.81   Transmembrane  177 - 193 ( 177 - 194)
   INTEGRAL    Likelihood = -0.27   Transmembrane  129 - 145 ( 129 - 145)

----- Final Results -----
              bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8609> which encodes amino acid sequence <SEQ ID 8610> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:    -19.59
GvH: Signal Score (-7.5): -4.49
     Possible site: 44
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value:  -2.81 threshold:  0.0
   INTEGRAL    Likelihood = -2.81   Transmembrane  172 - 188 ( 172 - 189)
   INTEGRAL    Likelihood = -0.27   Transmembrane  124 - 140 ( 124 - 140)
   PERIPHERAL  Likelihood =  8.01       30
```

```
 modified ALOM score:   1.06

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 592

A DNA sequence (GBSx0632) was identified in *S.agalactiae* <SEQ ID 1839> which encodes the amino acid sequence <SEQ ID 1840>. 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>
```

A related GBS nucleic acid sequence <SEQ ID 10223> which encodes amino acid sequence <SEQ ID 10224> was also identified.

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

```
>GP:CAB49414 GB:AJ248284 hypothetical protein [Pyrococcus abyssi]
 Identities = 29/86 (33%), Positives = 52/86 (59%), Gaps = 4/86 (4%)

Query: 14  TYYILLALFE--EAHGYAIMQKVEEMSGGDVRIAAGTMYGAIENLLKQKWIKSIPSD--D 69
           +Y ILL L E  + HGYAI +++EE++ G +  + G +Y  ++ L K K ++   ++
Sbjct: 19  SYLILLILNENEKLHGYAIRKRLEELTDGKLVPSEGALYSILKMLKKYKLVEDYWAEVGG 78

Query: 70  RRRKVYIITETGKEIVELETNRLRKL 95
           R R+ Y ITE GKE+++     +R++
Sbjct: 79  RVRRYYQITELGKEVLDEIKEEIREI 104
```

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 593

A DNA sequence (GBSx0633) was identified in *S.agalactiae* <SEQ ID 1841> which encodes the amino acid sequence <SEQ ID 1842>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0510(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10225> which encodes amino acid sequence <SEQ ID 10226> was also identified.

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

```
>GP:AAF22299 GB:AF185571 putative N-acetyltransferase Camello 2
          [Homo sapiens]
 Identities = 32/110 (29%), Positives = 54/110 (49%), Gaps = 4/110 (3%)

Query: 67  IKMAEQDDIFQIENYYQNRKGQ-FWIALENEKVVGSIALLRIDDKTAVLKKFFTYPKYRG 125
           + +A + D+  I   Y +  G  FW+A   EKVVG++  L +DD T   K+   +
Sbjct: 86  VDIALRTDMSDITKSYLSECGSCFWVAESEEKVVGTVGALPVDDPTLREKRLQLFHLSVD 145

Query: 126 NPVR---LGRKLFERFMLFARASKFTRIVLDTPEKEKRSHFFYENQGFKQ 172
           N  R   + + L    + FAR   ++ +VLDT   +   Y++ GFK+
Sbjct: 146 NEHRGQGIAKALVRTVLQFARDQGYSEVVLDTSNIQLSAMGLYQSLGFKK 195
```

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 594

A DNA sequence (GBSx0634) was identified in *S.agalactiae* <SEQ ID 1843> which encodes the amino acid sequence <SEQ ID 1844>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-11.94   Transmembrane  159 - 175 ( 151 - 180)
   INTEGRAL    Likelihood =-11.62   Transmembrane  231 - 247 ( 225 - 251)
   INTEGRAL    Likelihood = -9.98   Transmembrane  182 - 198 ( 177 - 203)
   INTEGRAL    Likelihood = -7.11   Transmembrane  118 - 134 ( 106 - 136)
   INTEGRAL    Likelihood = -1.49   Transmembrane   74 -  90 (  74 -  93)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 10227> which encodes amino acid sequence <SEQ ID 10228> was also identified.

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

```
>GP:CAB15891 GB:Z99123 yxlG [Bacillus subtilis]
 Identities = 42/188 (22%), Positives = 94/188 (49%), Gaps = 4/188 (2%)

Query: 1   MKSLAVMLKKEWMENVRTYKVISILITCSIFGILGPLTALMMPDIMA--GILPKKLQGAI 58
           MK +  +L+KEW+E  ++ K+I + I   I G+  PLT   MP+I+A  G LP  ++ +
Sbjct: 1   MKVMMALLQKEWLEGWKSGKLIWLPIAMMIVGLTQPLTIYYMPEIIAHGGNLPDGMKISF 60

Query: 59  PEPTYIDSYIQYFKNMNQLGLVILVFLFSSTLTQEFSKGTLINLVTKGLAKKVIILAKFI 118
             P+  +       N LG+ +++F    ++  E ++G    ++++ +     I++K++
Sbjct: 61  TMPSGSEVMVSTLSQFNTLGMALVIFSVMGSVANERNQGVTALIMSRPVTAAHYIVSKWL 120

Query: 119 VITLLWTVSYLLSVVIHFSYTLYYFSNEGSHKLMVYGATWFIGILFI-SLILFFSVLFRK 177
           + +++  +S+     + + Y    F +    +       + + ++FI +  L  S +FR
Sbjct: 121 IQSVIGIMSFAAGYGLAYYYVRLLFEDASFSRFAASLGLYALWVIFIVTAGLAGSTIFR- 179

Query: 178 TLGGLLGC 185
           ++G    C
Sbjct: 180 SVGAAAAC 187
```

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 595

A DNA sequence (GBSx0635) was identified in *S.agalactiae* <SEQ ID 1845> which encodes the amino acid sequence <SEQ ID 1846>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3431(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10229> which encodes amino acid sequence <SEQ ID 10230> was also identified.

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

```
>GP:CAB12736 GB:Z99108 similar to ABC transporter (ATP-binding
          protein) [Bacillus subtilis]
 Identities = 105/299 (35%), Positives = 175/299 (58%), Gaps = 11/299 (3%)

Query: 4   ISFQNVTKSFGPKKILNNVSFDLEENMIYGFVGPNGAGKTTTIKMILGLLKFDTGFITIF 63
           +  +NVTK+    + I++++SF + E  ++GF+GPNGAGKTTTI+M++GL+K   G + I
Sbjct: 5   LELKNVTKNIRGRTIIDDLSFTIREGEVFGFLGPNGAGKTTTIRMMVGLMKLSKGDVLIC 64

Query: 64  GKKVNFGRTDTNQLIGYLPDVPEYYDYMTALEYLDLCSGLARSKHKLSNKELLRSVGLDD 123
           G+ +         + IG + + PE Y +++   + L    + +    K    E++  VGL D
Sbjct: 65  GQSITKEYAKAIKHIGAIVENPELYKFLSGYKNLQQFARMVKGVTKEKIDEVVELVGLTD 124

Query: 124 N-HQKIATYSRGMKQRLGLAQALVHDPKIIICDEPTSALDPKGRQDILDIISNLRGEK-- 180
             H K+ TYS GM+QRLGLAQ L+HDPK++I DEPT+ LDP G ++I D +  L  E+
Sbjct: 125 RIHDKVKTYSLGMRQRLGLAQCLLHDPKVLILDEPTNGLDPAGIREIRDHLKKLTRERGM 184

Query: 181 TVIFSTHILSDVEKICDHVLVLTKCGIYSLEELKGKKSEENYSVRILIKVTKSEAKVLSH 240
            VI S+H+LS++E +CD + +L K  +  ++ +K +  +EN +    ++   SEA  + +
Sbjct: 185 AVIVSSHLLSEMELMCDRIAILQKGKLIDIQNVKDENIDENDTYFFQVE-QPSEAATVLN 243

Query: 241 NYQIEKKDNEYALTLKGSKMDNKADLLAGFYQDLVSLKISPSAIEVIDNSLEELYLEVT 299
            Y +  K N   + L   ++    +L       LV  +I    ++VI  SLE+ +LE+T
Sbjct: 244 QYDLLSKTNGVEIKLAKEEVPAVIEL-------LVMQQIRIYEVKVITKSLEDRFLEMT 295
```

There is also homology to SEQ ID 686.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 596

A DNA sequence (GBSx0636) was identified in *S.agalactiae* <SEQ ID 1847> which encodes the amino acid sequence <SEQ ID 1848>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4040(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 GENPEPT database:

```
>GP:AAB71491 GB:U53767 ORF6 [Bacillus pumilus]
 Identities = 39/134 (29%), Positives = 71/134 (52%), Gaps = 16/134 (11%)

Query: 2   LGENIYLQRTQIGMTQENLSDYLHLTKTTISKWENNQAKPDIDYLILMANLFDISLDDLV 61
           LG NI  +R  + ++QE +++ L +++   ISKWE NQ++P +D LI +A LFD  + +LV
Sbjct: 4   LGSNISNKRKSLKLSQEYVAEQLGVSRQAISKWETNQSEPSMDNLIRLAELFDSDIKELV 63

Query: 62  GYQKTLSDDQRNQLIKDLKIKANVLSERDFFQEVKELSKQFPNDFKTLLIMINM--VLSN 119
                S +Q ++  KDL+ +              K++  Q    F  +L++I+    +
Sbjct: 64  ------SPEQYSEEQKDLETRIE--------HGQKDIKMQMSAVFGRILMLISFFGYIGA 109

Query: 120 LTNLNDSEMKEWSL 133
           L +L+  ++  W L
Sbjct: 110 LFDLSSYQLPIWXL 123
```

There is also homology to SEQ ID 1740.

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

Example 597

A DNA sequence (GBSx0637) was identified in *S.agalactiae* <SEQ ID 1849> which encodes the amino acid sequence <SEQ ID 1850>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.59   Transmembrane  152 - 168 ( 145 - 173)
   INTEGRAL    Likelihood = -9.71   Transmembrane    7 -  23 (   3 -  27)
   INTEGRAL    Likelihood = -6.95   Transmembrane  125 - 141 ( 122 - 146)
   INTEGRAL    Likelihood = -4.51   Transmembrane   85 - 101 (  83 - 102)
   INTEGRAL    Likelihood = -3.35   Transmembrane   55 -  71 (  54 -  75)

----- Final Results -----
              bacterial membrane --- Certainty=0.6434(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 GENPEPT database:

```
>GP:CAA79986 GB:Z21972 ORF2 [Bacillus megaterium]
 Identities = 51/186 (27%), Positives = 106/186 (56%), Gaps = 5/186 (2%)

Query: 5   SFFQCVILLVSFLVLTLAVKSQSDMISYLDNITSAFFQSIRNPDLTNLMTIISTVVSPLT 64
           +F   V+ L+ F  +   + S ++ + + +++ S   Q   +P LT++M   + + S  +
Sbjct: 10  AFIISVLSLIGFSFMAFTI-SANEYLKFDEDVIS-LVQGWESPLLTDIMKFFTYIGSTAS 67

Query: 65  TSLIALVILGYQY-FLNQRIAVWLFM-LFFGTNALALLLKDIIARHRP-MNQLVFDSGYS 121
             +++LVIL + Y  L  R+ + LF  + G+  L L++K    R RP +++L+   GYS
Sbjct: 68  LIILSLVILFFLYRILKHRLELVLFTAVMVGSPLLNLMVKLFFQRARPDLHRLIDIGGYS 127

Query: 122 FPSGHTISAFLLMILVLVVARQRLRRVLSQVVFVIFALVILASVIFSRLYLENHFLTDIL 181
           FPSGH ++AF L  ++  +  + +++++ ++F+++++ S+  SR+YL  H+ +DI+
Sbjct: 128 FPSGHAMNAFSLYGILTFLLWRHITARWARILLILFSMLMILSIGISRIYLGVHYPSDII 187

Query: 182 GSLLLG 187
             L G
Sbjct: 188 AGYLAG 193
```

There is also homology to SEQ ID 1852.

A related GBS gene <SEQ ID 8611> and protein <SEQ ID 8612> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:     11.91
GvH: Signal Score (-7.5): -4.6
     Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 5 value: -13.59 threshold:  0.0
   INTEGRAL    Likelihood =-13.59   Transmembrane  152 - 168 ( 145 - 173)
   INTEGRAL    Likelihood = -9.71   Transmembrane    7 -  23 (   3 -  27)
   INTEGRAL    Likelihood = -6.95   Transmembrane  125 - 141 ( 122 - 146)
   INTEGRAL    Likelihood = -4.51   Transmembrane   85 - 101 (  83 - 102)
   INTEGRAL    Likelihood = -3.35   Transmembrane   55 -  71 (  54 -  75)
   PERIPHERAL  Likelihood =  1.16     184
 modified ALOM score:   3.22

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6434(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:

```
ORF01359(313 - 864 of 1212)
EGAD|16772|16959(10   -   194   of   216)   hypothetical   protein   (Bacillus   megaterium)
GP|288301|emb|CAA79986.1||Z21972 ORF2 (Bacillus megaterium) PIR|S32217|S32217 hypothetical
protein 2 - Bacillus megaterium
%Match = 9.5
%Identity = 28.2  %Similarity = 60.1
Matches = 53  Mismatches = 68  Conservative Sub.s = 60

66        96        126       156       186       216       246       276
SFFIEFTHPFLIICNIHYSLRFKYIVAILLY**KFER*LIGKVRIWYFF*FVNSHI*T*KVSAYFKHFLNILNHNV*RFI

306       336       366       396       426       456       486       516
SLLK*GYVVNKKSFFQCVILLVSFLVLTLAVKSQSDMISYLDNITSAFFQSIRNPDLTNLMTIISTVVSPLTTSLIALVI
            :|    |: |: |  : : : | :: : : :::  :: |   :| ||::|   : : |  :  :::|||
   MKLKQQLTIAFIISVLSLIGFSFMAFTI-SANEYLKFDEDV-ISLVQGWESPLLTDIMKFFTYIGSTASLIILSLVI
             10        20        30        40        50        60        70

543       570       600       630       657       687       714       744
LGYQY-FLNQRIAVWLFM-LFFGTNALALLLKDIIARHRP-MNQLVFDSGYSFPSGHTISAFLLM-ILVLVVARQRLRRV
| : |  | :|: : ||  :  |:  | |::|    | || :::|:  |||||||| ::|| |  || ::: |: :
LFFLYRILKHRLELVLFTAVMVGSPLLNLMVKLFFQRARPDLHRLIDIGGYSFPSGHAMNAFSLYGILTFLLWRH-ITAR
              90        100       110       120       130       140       150

774       804       834       864       894       924       954       984
LSQVVFVIFALVILASVIFSRLYLENHFLTDILGSLLLGASSYYGLSAIVSLKELQ*K**LPMNYKRAFLKGSFIIHYFS
 :::::::|:::::  |:  ||:||  |: :||:   | |
WARILLILFSMLMILSIGISRIYLGVHYPSDIIAGYLAGGCWIAISIWFFQRYQDRRKNKDR
              170       180       190       200       210
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 598

A DNA sequence (GBSx0638) was identified in *S.agalactiae* <SEQ ID 1853> which encodes the amino acid sequence <SEQ ID 1854>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4288(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 GENPEPT database:

```
>GP:CAB15470 GB:Z99121 yvdC [Bacillus subtilis]
 Identities = 53/96 (55%), Positives = 70/96 (72%)

Query: 1  MDITDYQKWVSEFYKKRNWYQYNSFIRSNFLSEEVGELAQAIRKYEIGRDRPDETEQTDL 60
          M + D +KW+ EFY+KR W +Y  FIR  FL EE GELA+A+R YEIGRDRPDE E +
Sbjct: 1  MQLADAEKWMKEFYEKRGWTEYGPFIRVGFLMEEAGELARAVRAYEIGRDRPDEKESSRA 60

Query: 61 ENLNDIKEELGDVLDNIFILADQYNISLEEIISAHR 96
          E   ++ EE+GDV+ NI ILAD Y +SLE+++ AH+
Sbjct: 61 EQKQELIEEMGDVIGNIAILADMYGVSLEDVMKAHQ 96
```

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

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

Example 599

A DNA sequence (GBSx0639) was identified in *S.agalactiae* <SEQ ID 1855> which encodes the amino acid sequence <SEQ ID 1856>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0635(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 GENPEPT database:

```
>GP:BAB06803 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 83/186 (44%), Positives = 117/186 (62%)

Query: 1   MRITIFCGASTGENPVYSEKTVALAQWMAQNKHSLVYGGGKVGLMGVMADTVIANGGYTT 60
           M+I +FCG+S G + VY E     L + +A+    +LVYGG  VG+MG +AD+V+  GG
Sbjct: 1   MKIAVFCGSSNGASDVYKEGARQLGKELARRGITLVYGGASVGIMGAVADSVLEAGGEVI 60

Query: 61  GVIPTFLRDREIAHENLSELIIVNNMPERKAKMMLLGDAFIALPGGPGTLEEISEVISWS 120
           GV+P FL + EI+H +L++LI+V  M ERKAKM  L D F+ALPGGPGTLEE  E+ +W+
Sbjct: 61  GVMPRFLEEPEISHPHLTKLIVVETMHERKAKMAELADGFLALPGGPGTLEEFFEIFTWA 120

Query: 121 RIGQNDNPCILYNVNGYFNDLKNMFDHMVGEGFLSLEDRENVLFSDDITEIEDFITNYKV 180
           +IG +  PC L N+N YF+ L  +  HM  E FL  + R   L   D   + D  + Y+
Sbjct: 121 QIGLHQKPCGLLNINHYFDPLVTLLHHMSNEQFLHEKYRSMALVHTDPILLLDQFSTYEP 180

Query: 181 PSTRQY 186
           P+ + Y
Sbjct: 181 PTVKAY 186
```

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

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

Example 600

A DNA sequence (GBSx0640) was identified in *S.agalactiae* <SEQ ID 1857> which encodes the amino acid sequence <SEQ ID 1858>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.86   Transmembrane  222 - 238 ( 214 - 239)
   INTEGRAL    Likelihood = -6.69   Transmembrane   39 -  55 (  36 -  58)
   INTEGRAL    Likelihood = -4.25   Transmembrane  266 - 282 ( 266 - 284)
   INTEGRAL    Likelihood = -1.28   Transmembrane  166 - 182 ( 166 - 182)
   INTEGRAL    Likelihood = -1.01   Transmembrane  190 - 206 ( 190 - 206)
   INTEGRAL    Likelihood = -0.96   Transmembrane   70 -  86 (  70 -  86)

----- Final Results -----
               bacterial membrane --- Certainty=0.4142(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 GENPEPT database:

```
>GP:CAB12420 GB:Z99107 ydiL [Bacillus subtilis]
 Identities = 40/132 (30%), Positives = 63/132 (47%), Gaps = 8/132 (6%)

Query: 107 ESQNYDATFNI-----LMISYSVVVGPFFEEVLYRGIVLNLL-SKYGKWFAIITSGILFG 160
           ES+N  A  ++      LMI  S +VGP  EE+++R I+   L  K   +FA + S ++FG
Sbjct: 114 ESENTQAILDVIQAVPLMIIVSSIVGPILEEIIFRKIIFGALYEKTNFFFAGLISSVIFG 173

Query: 161 LMHQDISQLLTTSIAGIIMGFI-AYHYSFKVALLLHICNNFIVEIFTQLSTVNELYGTYF 219
           ++H D+  LL  +  G    F+ A      V +  H+  N  V +  QL  V
Sbjct: 174 IVHADLKHLLLYTAMGFTFAFLYARTKRIWVPIFAHLMMNTFV-VIMQLEPVRNYLEQQS 232

Query: 220 ENILLILAILFI 231
             + LI+  LF+
Sbjct: 233 TQMQLIIGGLFL 244
```

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

A related GBS gene <SEQ ID 8613> and protein <SEQ ID 8614> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:     12.52
GvH: Signal Score (-7.5): -1.74
     Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 2 value:  -6.69 threshold:  0.0
   INTEGRAL    Likelihood = -6.69   Transmembrane   39 -  55 (  36 -  58)
   INTEGRAL    Likelihood = -0.96   Transmembrane   70 -  86 (  70 -  86)
   PERIPHERAL  Likelihood =  4.56      21
 modified ALOM score:   1.84

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3675(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:

```
Query:    10 LIGLILLAQAIVLSLATTLFAEILQNDVWIGIASTLIALLIPCF 53
             L+ L LL ++++LS++       +L   +W+ +A+ L+A ++ CF
Sbjct:    21 LLCLCLLVRSLLLSVSLYSALILLVLILWVTVATPLLAFVVSCF 64
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 601

A DNA sequence (GBSx0641) was identified in *S.agalactiae* <SEQ ID 1859> which encodes the amino acid sequence <SEQ ID 1860>. This protein is predicted to be capa protein. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.80   Transmembrane   27 -  43 (  22 -  50)

----- Final Results -----
               bacterial membrane --- Certainty=0.6519(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9385> which encodes amino acid sequence <SEQ ID 9386> was also identified.

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

```
>GP:AAF13661 GB:AF188935 pXO2-56 [Bacillus anthracis]
 Identities = 68/224 (30%), Positives = 118/224 (52%), Gaps = 10/224 (4%)

Query: 95  FKEVKSWIESADLAIGDYEGTISSE----YPLAGYPL-FNAPNEIATTMKETGYDVVDLA 149
           F+ V  +++++D   G++E  +  E    Y  A   +  +A  E    +KE G+ V++LA
Sbjct: 87  FRHVSPYLKNSDYVSGNFEHPVLLEDKKNYQKADKNIHLSAKEETVKAVKEAGFTVLNLA 146

Query: 150 HNHILDSQLAGAINTVKTFNRLGLDTIGVYLKDRNKEDILIKHVNGIKIAILGYSYGY-N 208
           +NH+ D    G  +T+K F    LD +G     ++ ++I+ ++VNG+++A LG++  +
Sbjct: 147 NNHMTDYGAKGTKDTIKAFKEADLDYVGAGENFKDVKNIVYQNVNGVRVATLGFTDAFVA 206

Query: 209 GMEANVSKSDYEKHMSDLDTKKIKQDIKKAEKEADITIVMPQMGIEYQKKPTTEQVMLYH 268
           G  A   +        D+  K+I +     +  AD+ +V    G EY  KP+  Q  L
Sbjct: 207 GAIATKEQPGSLSMNPDVLLKQISKAKDPKKGNADLVVVNTHWGEEYDNKPSPRQEALAK 266

Query: 269 SMIKWGADIIFGGHPHVVEPSEVIKKDGQKKFIIYSMGNFISNQ 312
           +M+  GADII G HPHV++  +V K+      I YS+GNF+ +Q
Sbjct: 267 AMVDAGADIIVGHHPHVLQSFDVYKQG----IIFYSLGNFVFDQ 306
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1861> which encodes the amino acid sequence <SEQ ID 1862>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.05   Transmembrane   44 -  60 (  40 -  68)

----- Final Results -----
               bacterial membrane --- Certainty=0.5819(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 9119> which encodes the amino acid sequence <SEQ ID 9120>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty= 0.582(Affirmative) < succ>
                bacterial outside --- Certainty= 0.000(Not Clear) < succ>
```

```
bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 232/334 (69%), Positives = 273/334 (81%), Gaps = 4/334 (1%)

Query: 24  YQKTLIFCVAVIIAIFILGLSKDLAQSKGQKVANNNT----VKTARVVANGDILLHDVLY 79
           Y+KT+   VA+I+A+ + GL  DL   +  ++A   +    VKTARVVANGDIL+HD+LY
Sbjct: 40  YKKTMATVVALIVALLLFGLIYDLLGVQKNELAAQKSAQPKVKTARVVANGDILIHDILY 99

Query: 80  ASARQPDGTYNFTPYFKEVKSWIESADLAIGDYEGTISSEYPLAGYPLFNAPNEIATTMK 139
            SAR+ D TY+FTPYF+ VK WI  ADLAIGDYEGTIS +YPLAGYPLFNAP EIA  +K
Sbjct: 100 MSARKADDTYDFTPYFEYVKDWISGADLAIGDYEGTISPDYPLAGYPLFNAPEEIAGALK 159

Query: 140 ETGYDVVDLAHNHILDSQLAGAINTVKTFNRLGLDTIGVYLKDRNKEDILIKHVNGIKIA 199
            TGYDVVDLAHNHILDSQL GA+NT K F++LG+D+IG+Y KDR+KE  LIK+VNGIKIA
Sbjct: 160 NTGYDVVDLAHNHILDSQLDGALNTKKVFHQLGIDSIGIYDKDRSKESFLIKNVNGIKIA 219

Query: 200 ILGYSYGYNGMEANVSKSDYEKHMSDLDTKKIKQDIKKAEKEADITIVMPQMGIEYQKKP 259
           ILGYSYGYNGMEA +S+ DYEKHMSDLD  KIK++++ AEK+AD+TIVMPQMG EY  +P
Sbjct: 220 ILGYSYGYNGMEATLSQEDYEKHMSDLDEAKIKKELQLAEKKADVTIVMPQMGTEYALEP 279

Query: 260 TTEQVMLYHSMIKWGADIIFGGHPHVVEPSEVIKKDGQKKFIIYSMGNFISNQRLETVDD 319
           T EQ  LYH MI WGAD++ GGHPHV+EPSE + K  QKKFIIYSMGNFISNQRLETVDD
Sbjct: 280 TAEQKELYHKMIDWGADVVLGGHPHVIEPSETVIKGRQKKFIIYSMGNFISNQRLETVDD 339

Query: 320 IWTERGLLMDVTIEKKGQKTVIKKVKAHPTLVEA 353
           IWTERGLLMD+T EKK  KT IK V+AHPT+V A
Sbjct: 340 IWTERGLLMDLTFEKKDNKTKIKTVEAHPTMVLA 373
```

A related GBS gene <SEQ ID 8615> and protein <SEQ ID 8616> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 7
SRCFLG: 0
McG: Length of UR:   18
     Peak Value of UR:   3.83
     Net Charge of CR: 2
McG: Discrim Score:     15.36
GvH: Signal Score (-7.5): -1.52
     Possible site: 32
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 33
ALOM program   count: 0 value:   4.35 threshold:  0.0
   PERIPHERAL  Likelihood =  4.35     170
 modified ALOM score:  -1.37

*** Reasoning Step: 3

Rule gpo1

----- 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:

```
30.6/53.3% over 230aa

                                                                  Bacillus anthracis
  EGAD|20151| capa protein Insert characterized
  SP|P19579|CAPA_BACAN CAPA PROTEIN. Edit characterized
  GP|142633|gb|AAA22288.1||M24150 46 Kd encapsulation protein CapA Insert characterized
  PIR|C30091|C30091 capA protein - Insert characterized

ORF02075(574 - 1257 of 1734)
```

```
EGAD|20151|20674(83 - 313 of 411) capa protein (Bacillus anthracis) SP|P19579|CAPA_BACAN
CAPA PROTEIN. GP|142633|gb|AAA22288.1||M24150 46 Kd encapsulation protein CapA (Bacillus
anthracis) PIR|C30091|C30091 capA protein - Bacillus anthracis
%Match = 8.9
%Identity = 30.6  %Similarity = 53.3
Matches = 70  Mismatches = 102  Conservative Sub.s = 52

468       498       528       558       585       615       645       663
LAQSKGQKVANNNTVKTARVVANGDILLHDVLYASARQPDGTYNFTPY-FKEVKSWIESADLAIGDYEGTI----SSEYP
:| :  |:      ||          :: |::     :       | | |: |  :::::|   |::|  :       |
IAATWVQRTEAVAPVKHRENEKLTMTMVGDIMMGRHVKEIVNRYGTDYVFRHVSPYLKNSDYVSGNFEHPVLLEDKKNYQ
          50        60        70        80        90       100       110

690       720       750       780       810       840       870       900
LAGYPL-FNAPNEIATTMKETGYDVVDLAHNHILDSQLAGAINTVKTFNRLGLDTIGVYLKDRNKEDILIKHVNGIKIAI
 |   : ::|  |    :|| |: |::||:||: |    | :|:| |    || :|     :: ::|: ::|||:::|
KADKNIHLSAKEETVKAVKEAGFTVLNLANNHMTDYGAKGTKDTIKAFKEADLDYVGAGENFKDVKNIVYQNVNGVRVAT
          130       140       150       160       170       180       190

927       957       987      1017      1047      1077      1107      1137
LGYSYGY-NGMEANVSKSDYEKHMSDLDTKKIKQDIKKAEKEADITIVMPQMGIEYQKKPTTEQVMLYHSMIKWGADIIF
||::  :  |  |   :        |:  |:| :     :  ||: :|  : | ||  ||:  |  |  :|:  |||||
LGFTDAFVAGAIATKEQPGSLSMNPDVLLKQISKAKDPKKGNADLVVVNTHWGEEYDNKPSPRQEALAKAMVDAGADIIV
          210       220       230       240       250       260       270

1167      1197      1227      1257      1287      1317      1347      1377
GGHPHVVEPSEVIKKDGQKKFIIYSMGNFISNQRLETVDDIWTERGLLMDVTIEKKGQKTVIKKVKAHPTLVEAKPNGRY
| ||||::  :| |    :  | ||:|||: :|      |  :  | |          :     :  :  |    |
GHHPHVLQSFDVYK----QGIIFYSLGNFVFDQGWTRTKDSALVQYHLRDNGTAILDVVPLNIQEGSPKPVASALDKNRV
          290           300       310       320       330       340       350
```

SEQ ID 8616 (GBS289) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 5; MW 40kDa), in Figure 181 (lane 6; MW 47kDa), in Figure 169 (lane 13 & 14; MW 54.5kDa – thioredoxin fusion) and in Figure 239 (lane 3; MW 54.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 5; MW 65kDa).

SEQ ID 8616 (GBS289L) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 126 (lane 2; MW 72kDa) and in Figure 184 (lane 5; MW 72kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 126 (lane 5-7; MW 47kDa).

GBS289L-His was purified as shown in Figure 234, lane 9-10. Purified GBS289L-GST is shown in Figure 245, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 602

A DNA sequence (GBSx0642) was identified in *S.agalactiae* <SEQ ID 1863> which encodes the amino acid sequence <SEQ ID 1864>. This protein is predicted to be thiamin biosynthesis protein ThiI (thiI). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

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

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

A related GBS nucleic acid sequence <SEQ ID 9971> which encodes amino acid sequence <SEQ ID 9972> was also identified.

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

```
>GP:AAC00308 GB:AF008220 YtbJ [Bacillus subtilis]
 Identities = 184/354 (51%), Positives = 249/354 (69%)

Query: 11  MQYSEIMIRYGELSTKKKNRMRFINKLKNNMEHVLSIYPDVSVKTDRDRGHVYLNGTDYH 70
           M Y  I+IR+GE+STK KNR  FI +LK N+  VL  YP++   ++RDR  + LNG D
Sbjct: 1   MNYDHILIRFGEISTKGKNRKSFIERLKQNIRLVLKDYPNLKYFSNRDRMTITLNGEDPE 60

Query: 71  EVAESLKEIFGIQAFSPSFKVEKNVDTLVKAVQEIMTSVYKDGMTFKITAKRSDHSFELD 130
            +   LK++FGIQ+FS + K +  +D +     + +   YK G TFK+  KR+   FELD
Sbjct: 61  ALFPHLKQVFGIQSFSLAIKCDSRLDDIKATALKAIKDQYKPGDTFKVATKRAYKQFELD 120

Query: 131 SRALNHTLGDAVFSVLPNIKAQMKQPDINLKVEIRDEAAYISYEDIRGAGGLPVGTSGKG 190
           +  +N  +G  +      +   ++ PDI L++EIR+EA +++  D +GAGGLPVG++GK
Sbjct: 121 TNQMNAEIGGHILRNTEGLTVDVRNPDIPLRIEIREEATFLTIRDEKGAGGLPVGSAGKA 180

Query: 191 MLMLSGGIDSPVAGYLALKRGVDIEAVHFASPPYTSPGALKKAHDLTRKLTKFGGNIQFI 250
           MLMLSGG DSPVAG+ A+KRG+ +EAVHF SPPYTS  A +K  DL + L++FGG++
Sbjct: 181 MLMLSGGFDSPVAGFYAMKRGLSVEAVHFFSPPYTSERAKQKVMDLAKCLSRFGGSMTLH 240

Query: 251 EVPFTEIQEEIKAKAPEAYLMTLTRRFMMRITDRIREDRNGLVIINGESLGQVASQTLES 310
            VPFT+ QE I+ + PE Y MT TRR M++I DRIRE RNGL II GESLGQVASQTLES
Sbjct: 241 IVPFTKTQELIQKQIPENYTMTATRRLMLQIADRIREKRNGLAIITGESLGQVASQTLES 300

Query: 311 MQAINAVTATPIIRPVVTMDKLEIIDIAQKIDTFDISIQPFEDCCTIFAPDRPK 364
           M AINAVT+TPI+RP++ MDK EII+ +++I T++ SIQPFEDCCTIF   +P+
Sbjct: 301 MYAINAVTSTPILRPLIAMDKTEIIEKSREIGTYETSIQPFEDCCTIFTTAKPR 354
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1865> which encodes the amino acid sequence <SEQ ID 1866>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4897(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 316/404 (78%), Positives = 362/404 (89%)

Query: 11  MQYSEIMIRYGELSTKKKNRMRFINKLKNNMEHVLSIYPDVSVKTDRDRGHVYLNGTDYH 70
           M YSEIM+R+GELSTK KNRMRFINKLKNN++ VL+ +P ++V++DRDR HV LNGTDY
Sbjct: 1   MDYSEIMVRHGELSTKGKNRMRFINKLKNNIQDVLAPFPAITVRSDRDRTHVSLNGTDYQ 60

Query: 71  EVAESLKEIFGIQAFSPSFKVEKNVDTLVKAVQEIMTSVYKDGMTFKITAKRSDHSFELD 130
            + E+LK +FG+QA SP +K+EK+V  LV AVQ+IMTS+Y+DG+TFKI  KRSDH+FELD
Sbjct: 61  PIVEALKLVFGVQALSPVYKLEKSVPLLVTAVQDIMTSLYRDGLTFKIATKRSDHAFELD 120

Query: 131 SRALNHTLGDAVFSVLPNIKAQMKQPDINLKVEIRDEAAYISYEDIRGAGGLPVGTSGKG 190
           SR LN  LG AVF VLPNI+AQMK PD+ LKVEIRDEAAYISYE+I+GAGGLPVGTSGKG
Sbjct: 121 SRELNSLLGGAVFEVLPNIQAQMKHPDVTLKVEIRDEAAYISYEEIKGAGGLPVGTSGKG 180

Query: 191 MLMLSGGIDSPVAGYLALKRGVDIEAVHFASPPYTSPGALKKAHDLTRKLTKFGGNIQFI 250
           MLMLSGGIDSPVAGYLALKRG+DIE VHFASPPYTSPGAL KA DLTR+LT+FGGNIQFI
Sbjct: 181 MLMLSGGIDSPVAGYLALKRGLDIEVVHFASPPYTSPGALAKAQDLTRRLTRFGGNIQFI 240

Query: 251 EVPFTEIQEEIKAKAPEAYLMTLTRRFMMRITDRIREDRNGLVIINGESLGQVASQTLES 310
```

```
            EVPFTEIQEEIK KAPEAYLMTLTRRFMMRITD IRE R GLVI+NGESLGQVASQTLES
Sbjct: 241  EVPFTEIQEEIKNKAPEAYLMTLTRRFMMRITDAIREQRKGLVIVNGESLGQVASQTLES 300

Query: 311  MQAINAVTATPIIRPVVTMDKLEIIDIAQKIDTFDISIQPFEDCCTIFAPDRPKTNPKIK 370
            MQAINAVT+TPIIRPVVTMDKLEII++AQ IDTFDISIQPFEDCCTIFAPDRPKTNPK+
Sbjct: 301  MQAINAVTSTPIIRPVVTMDKLEIIEMAQAIDTFDISIQPFEDCCTIFAPDRPKTNPKLG 360

Query: 371  NTEQYEKRMDVEGLVERAVAGIMVTTIQPQADSDDVDDLIDDLL 414
            N E+YE+  D++GLV+RAV+GI+VT I P+  +D+V++LID LL
Sbjct: 361  NAEKYEECFDIDGLVQRAVSGIVVTEITPEIVNDEVENLIDALL 404
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 603

A DNA sequence (GBSx0643) was identified in *S.agalactiae* <SEQ ID 1867> which encodes the amino acid sequence <SEQ ID 1868>. This protein is predicted to be nifs protein homolog , fragment. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.27   Transmembrane  131 - 147 ( 131 - 147)

----- 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>
```

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

```
>GP:CAA43493 GB:X61190 nifS-like gene [Lactobacillus delbrueckii]
 Identities = 177/353 (50%), Positives = 234/353 (66%), Gaps = 1/353 (0%)

Query: 14   PEVLRTYQEVASKIYGNPSSLHELGTTSSRILEASRKQIASLLELKANEIFFTSGGTEAD 73
            P+ L TY +V +KI+GNPSSLH+LG  +  +LEASRKQ+A LL +  +EI+FTSGGTE++
Sbjct: 3    PKALETYSQVVTKIWGNPSSLHKLGDRAHGLLEASRKQVADLLGVNTDEIYFTSGGTESN 62

Query: 74   NWVIKGLAFEKQHFGNHIIVSDIEHPAVKESAKWLGEYGFEIDYAPVDDKGFVDVEALVK 133
            N  IKG A+ K+ FG HII S +EH +V  +   L   GF +   PVD +G V+ E L
Sbjct: 63   NTAIKGTAWAKREFGKHIITSSVEHASVANTFTELENLGFRVTRLPVDKEGRVNPEDLKA 122

Query: 134  LIKPETILISIMAINNEIGSIQPIKAISDLLSDKPTISFHVDAVQAIGKIPTKDYLTERV 193
             +  +T L+SIM +NNEIG+IQPIK IS++L+D P I FHVD VQA+GK       T RV
Sbjct: 123  ALDKDTTLVSIMGVNNEIGTIQPIKEISEILADYPNIHFHVDNVQALGKGIWDQVFTSRV 182

Query: 194  DFASFSSHKFHGVRGVGFLYIKEGKRISPLLTGGGQETDLRSTTENVAGIAATAKALRMV 253
            D  SFSSHKFHG RG+G LY K G+ + PL  GGGQE  LRS TEN+A IAA AKA R++
Sbjct: 183  DMMSFSSHKFHGPRGIGILYKKRGRMLMPLCEGGGQEKGLRSGTENLAAIAAMAKAARLL 242

Query: 254  MDKEVVAIPKISKMKTIIHDELAKYEDITLFSG-KEDFSPNIITFGIKGVRGEVLVHAFE 312
            +  E     +   +K  I   LA    I +FS  K DF+P+I+ F ++G+RGE LVH  E
Sbjct: 243  LTDEKEKADREYAIKEKISKYLAGKPGIHIFSPLKADFAPHILCFALEGIRGETLVHTLE 302

Query: 313  GHDIFISTTSACSSKAGKPAGTLIAMGISTKLAQTAVRISLDDDNDMGQVEQF 365
              DI+ISTTSAC+SK   A TL+AM     +A +AVR+S D+ N + + ++F
Sbjct: 303  DQDIYISTTSACASKKADEASTLVAMKTPDAIATSAVRLSFDESNTLEEADEF 355
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1869> which encodes the amino acid sequence <SEQ ID 1870>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3067(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 268/370 (72%), Positives = 322/370 (86%)

Query: 1   MIYFDNSATTIPYPEVLRTYQEVASKIYGNPSSLHELGTTSSRILEASRKQIASLLELKA 60
           MIYFDN+ATTIPY E L+TYQEVA+KIYGNPSSLH+LGT +SRILEASRKQIA LL +K+
Sbjct: 1   MIYFDNAATTIPYGEALKTYQEVATKIYGNPSSLHQLGTNASRILEASRKQIAGLLGVKS 60

Query: 61  NEIFFTSGGTEADNWVIKGLAFEKQHFGNHIIVSDIEHPAVKESAKWLGEYGFEIDYAPV 120
            EIFFTSGGTE+ NW IKG+AFEK  FG HII+S IEHPAV ES KWL   GFE+ YAPV
Sbjct: 61  EEIFFTSGGTESANWAIKGIAFEKNAFGKHIIISAIEHPAVSESVKWLLTQGFEVSYAPV 120

Query: 121 DDKGFVDVEALVKLIKPETILISIMAINNEIGSIQPIKAISDLLSDKPTISFHVDAVQAI 180
             +G VDV AL +LI+P+TILISIMA+NNE+G+IQPI+AIS+LL+++PTI+FHVDAVQAI
Sbjct: 121 TTQGVVDVNALAELIRPDTILISIMAVNNEMGAIQPIRAISNLLANQPTITFHVDAVQAI 180

Query: 181 GKIPTKDYLTERVDFASFSSHKFHGVRGVGFLYIKEGKRISPLLTGGGQETDLRSTTENV 240
           GKIP  DY+T RVD ASFS HKFH VRGVGFLY K GKR++PLL+GGGQE +LRSTTENV
Sbjct: 181 GKIPLCDYMTNRVDLASFSGHKFHSVRGVGFLYKKAGKRLNPLLSGGGQEQELRSTTENV 240

Query: 241 AGIAATAKALRMVMDKEVVAIPKISKMKTIIHDELAKYEDITLFSGKEDFSPNIITFGIK 300
           AGIA+ AKALR+V +K+V  +PK++ M+ +I+  L+ Y D+T+FS +E F+PNI+TFGI+
Sbjct: 241 AGIASMAKALRIVTEKQVSVLPKLTAMRDVIYKSLSAYPDVTVFSAQEGFAPNILTFGIR 300

Query: 301 GVRGEVLVHAFEGHDIFISTTSACSSKAGKPAGTLIAMGISTKLAQTAVRISLDDDNDMG 360
            GVRGEV+VHAFE ++I+ISTTSACSSKAG+PAG+L+AMGI  K AQTAVRISLDDDNDMG
Sbjct: 301 GVRGEVIVHAFEKYEIYISTTSACSSKAGEPAGSLVAMGIPVKTAQTAVRISLDDDNDMG 360

Query: 361 QVEQFLTIFK 370
           QVEQFLTIF+
Sbjct: 361 QVEQFLTIFQ 370
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 604

A DNA sequence (GBSx0644) was identified in *S.agalactiae* <SEQ ID 1871> which encodes the amino acid sequence <SEQ ID 1872>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1539(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 605

A DNA sequence (GBSx0645) was identified in *S.agalactiae* <SEQ ID [illegible] which encodes the amino acid sequence <SEQ ID 1874>. This protein is predicted to be glutathione reductase (gor). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood = -4.25   Transmembrane  170 - 186 ( 169 - 187)

----- 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>
```

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

```
>GP:BAA76640 GB:AB019579 glutathione reductase (GR) [Streptococcus mutans]
 Identities = 274/450 (60%), Positives = 346/450 (76%), Gaps = 1/450 (0%)

Query: 1   MSKQYDYIVIGGGSAGSGTANRAAMYGAKVLLIEGGQVGGTCVNLGCVPKKIMWYGAQVS 60
           M+KQYDYIVIGGGS G  +ANRAAM+GAKV+L EG QVGGTCVN+GCVPKK+MWYGAQV+
Sbjct: 1   MTKQYDYIVIGGGSGGIASANRAAMHGAKVILFEGKQVGGTCVNVGCVPKKVMWYGAQVA 60

Query: 61  ETLHKYSSGYGFEVNNLNFDFTTLKANRDAYVQRSRQSYAANFERNGVEKIDGFARFIDN 120
           ET++ Y++ YGF+V    F F  LK NR AY+ R + SY   F+ NGVE++  +A F+D
Sbjct: 61  ETINNYAADYGFDVTTQTFHFDALKQNRQAYIDRIQDSYERGFDSNGVERVYSYATFVDA 120

Query: 121 HTIEVNGQQYKAPHITIATGGHPLYPDIIGSELGETSDDFFGWETLPDSILIVGAGYIAA 180
           HT+EV G+ Y APHI IATGGH L PDI GSE G TSD FF  + +P    +VGAGYIA
Sbjct: 121 HTVEVAGEHYTAPHILIATGGHALLPDIPGSEYGITSDGFFELDAIPKRTAVVGAGYIAV 180

Query: 181 ELAGVVNELGVETHLAFRKDHILRGFDDMVTSEVMAEMEKSGISLHANHVPKSLKRDEGG 240
           E++GV++ LG ETHL  R+D  LR FD  +   ++ EM+K G  LH   VPK + ++
Sbjct: 181 EISGVLHALGGETHLFVRRDRPLRKFDKEIVGTLVDEMKKDGPHLHTFSVPKEVIKNTDN 240

Query: 241 KLIFEAENGKTLVVDRVIWAIGRGPNV-DMGLENTDIVLNDKGYIKADEFENTSVDGVYA 299
            L    ENG+   VD +IWAIGR  N     LE T + L+ +G+I  D FENT+V+G+YA
Sbjct: 241 SLTLILENGEEYTVDTLIWAIGRAANTKGFNLEVTGVTLDSRGFIATDAFENTNVEGLYA 300

Query: 300 IGDVNGKIALTPVAIAAGRRLSERLFNHKDNEKLDYHNVPSVIFTHPVIGTVGLSEAAAI 359
           +GDVNGK+ LTPVA+ AGR+LSERLFNHK   K+DY +V +VIF+HPVIG++GLSE  A+
Sbjct: 301 LGDVNGKLELTPVAVKAGRQLSERLFNHKPQAKMDYKDVATVIFSHPVIGSIGLSEEVAL 360

Query: 360 EQFGEDNIKVYTSTFTSMYTAVTTNRQAVKMKLITLGKEEKVIGLHGVGYGIDEMIQGFS 419
           +Q+GE+N+  VY STFTSMYTAVT++RQA KMKL+T+G++EK++GLHG+GYG+DEMIQGF+
Sbjct: 361 DQYGEENVTVYRSTFTSMYTAVTSHRQACKMKLVTVGEDEKIVGLHGIGYGVDEMIQGFA 420

Query: 420 VAIKMGATKADFDDTVAIHPTGSEEFVTMR 449
           VAIKMGATKADFD+TVAIHPTGSEEFVTMR
Sbjct: 421 VAIKMGATKADFDNTVAIHPTGSEEFVTMR 450
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1875> which encodes the amino acid sequence <SEQ ID 1876>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood = -1.33   Transmembrane  173 - 189 ( 173 - 191)

----- 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>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 268/446 (60%), Positives = 340/446 (76%), Gaps = 1/446 (0%)

Query: 5   YDYIVIGGGSAGSGTANRAAMYGAKVLLIEGGQVGGTCVNLGCVPKKIMWYGAQVSETLH 64
           YDYIVIGGGSAG  +ANRAAM+GAKVLL EG ++GGTCVNLGCVPKK+MWYGAQV++ L
Sbjct: 8   YDYIVIGGGSAGIASANRAAMHGAKVLLAEGKEIGGTCVNLGCVPKKVMWYGAQVADILG 67

Query: 65  KYSSGYGFEVNNLNFDFTTLKANRDAYVQRSRQSYAANFERNGVEKIDGFARFIDNHTIE 124
            Y+  YGF+     FDF  LKANR AY+ R   SY   FE+NGV++I  +A F D HT+E
Sbjct: 68  TYAKDYGFDFKEKAFDFKQLKANRQAYIDRIHASYERGFEQNGVDRIYDYAVFKDAHTVE 127

Query: 125 VNGQQYKAPHITIATGGHPLYPDIIGSELGETSDDFFGWETLPDSILIVGAGYIAAELAG 184
           + GQ Y APHI IATGGHP++PDI G++ G +SD FF  + +P    +VGAGYIA ELAG
Sbjct: 128 IAGQLYTAPHILIATGGHPVFPDIEGAQYGISSDGFFALDEVPKRTAVVGAGYIAVELAG 187

Query: 185 VVNELGVETHLAFRKDHILRGFDDMVTSEVMAEMEKSGISLHANHVPKSLKRDEGGKLIF 244
           V++ LG +T L  R D  LR FD  +   ++ EM  +G  LH +     + ++    L
Sbjct: 188 VLHALGSKTDLFIRHDRPLRSFDKTIVDVLVDEMAVNGPRLHTHAEVAKVVKNTDESLTL 247

Query: 245 EAENGKTLVVDRVIWAIGRGPNVD-MGLENTDIVLNDKGYIKADEFENTSVDGVYAIGDV 303
             ++G+ + VD++IWAIGR PN++   L+ T + LNDKGYI+ D +ENTSV G+YA+GDV
Sbjct: 248 YLKDGQEVEVDQLIWAIGRKPNLEGFSLDKTGVTLNDKGYIETDAYENTSVKGIYAVGDV 307

Query: 304 NGKIALTPVAIAAGRRLSERLFNHKDNEKLDYHNVPSVIFTHPVIGTVGLSEAAAIEQFG 363
           NGK+ALTPVA+AAGRRLSERLFN K +EKLDY NV +VIF+HPVIG+VGLSE AA++Q+G
Sbjct: 308 NGKLALTPVAVAAGRRLSERLFNGKTDEKLDYQNVATVIFSHPVIGSVGLSEEAAVKQYG 367

Query: 364 EDNIKVYTSTFTSMYTAVTTNRQAVKMKLITLGKEEKVIGLHGVGYGIDEMIQGFSVAIK 423
           ++ +K Y S FTSM+TA+T +RQ   MKL+T+G  EK++GLHG+GYG+DEMIQGF+VAIK
Sbjct: 368 QEAVKTYQSRFTSMFTAITNHRQPCLMKLVTVGDTEKIVGLHGIGYGVDEMIQGFAVAIK 427

Query: 424 MGATKADFDDTVAIHPTGSEEFVTMR 449
           MGATKADFD+TVAIHPTGSEEFVTMR
Sbjct: 428 MGATKADFDNTVAIHPTGSEEFVTMR 453
```

SEQ ID 1874 (GBS417) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 5; MW 53kDa).

GBS417-His was purified as shown in Figure 216, lane 2.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 606

A DNA sequence (GBSx0646) was identified in *S.agalactiae* <SEQ ID 1877> which encodes the amino acid sequence <SEQ ID 1878>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3122(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 GENPEPT database:

```
>GP:AAC62417 GB:AF084104 hypothetical protein [Bacillus firmus]
 Identities = 33/110 (30%), Positives = 66/110 (60%)

Query: 1   MANVYDLANELERAVRALPEYQAVLTAKSAIESDADAQVLWQDFLATQSKVQEMMQSGQM 60
           M+NVYD A+EL++A+    E+ A+ +    IE+D  A+ + ++F   Q ++Q+    G
Sbjct: 1   MSNVYDKAHELKKAIAESEEFSALKSMHEEIEADEIAKKMLENFRNLQLELQQKQMQGIQ 60

Query: 61  PSQEEQDEMSKLGEKIESNDLLKVYFDQQQRLSVYMSDIEKIVFAPMQDL 110
```

```
            ++EE  +  +  E ++ ++L+    + +QRLSV + DI KI+  P++++
Sbjct: 61   ITEEEAQKAQQQFELVQQHELISKLMEAEQRLSVIIGDINKIITEPLEEI 110
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1879> which encodes the amino acid sequence <SEQ ID 1880>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4058(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 68/108 (62%), Positives = 86/108 (78%)

Query: 4    VYDLANELERAVRALPEYQAVLTAKSAIESDADAQVLWQDFLATQSKVQEMMQSGQMPSQ 63
            +YD AN+LERAVRALPEYQ VL  K AI++D  A  L+ +F+A Q K+Q MMQSGQMP+
Sbjct: 5    IYDYANQLERAVRALPEYQKVLEVKEAIQADVSASELFDEFVAMQEKIQGMMQSGQMPTA 64

Query: 64   EEQDEMSKLGEKIESNDLLKVYFDQQQRLSVYMSDIEKIVFAPMQDLM 111
            EEQ  + +L +KIE+ND LK YF+ QQ LSVYMSDIE+IVFAP++DL+
Sbjct: 65   EEQTSIQELSQKIEANDQLKAYFEAQQALSVYMSDIERIVFAPLKDLV 112
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 607

A DNA sequence (GBSx0647) was identified in *S.agalactiae* <SEQ ID 1881> which encodes the amino acid sequence <SEQ ID 1882>. This protein is predicted to be chorismate synthase (aroC). Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.67   Transmembrane  343 - 359 ( 341 - 364)

----- Final Results -----
               bacterial membrane --- Certainty=0.2869(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 GENPEPT database:

```
>GP:BAB05375 GB:AP001512 chorismate synthase [Bacillus halodurans]
 Identities = 227/381 (59%), Positives = 282/381 (73%), Gaps = 2/381 (0%)

Query: 1    MRYLTAGESHGPSLTAIIEGIPAGLKLSAKDINEDLKRRQGGYGRGNRMKIETDQVIISS 60
            MRYLTAGESHGP LT IIEG PA L+L A DIN DL RRQGG+GRG RM+IE DQV I
Sbjct: 1    MRYLTAGESHGPQLTTIIEGAPAQLELVADDINVDLARRQGGHGRGRRMQIEKDQVQIVG 60

Query: 61   GVRHGKTLGSPITLTVTNKDHSKWLDIMSVEDI--EERLKQKRRIKHPRPGHADLVGGIK 118
            G+RHGKT G+PI L V NKD   W  IM  E +  +E  + KR+I  PRPGHADL G IK
Sbjct: 61   GIRHGKTTGAPIALVVENKDWKHWTKIMGAEPLTGDEEKEIKRKITRPRPGHADLNGAIK 120

Query: 119  YRFDDLRNALERSSARETTMRVAIGAIAKRILKEIGIEIANHIVVFGGKEITVPDKLTVQ 178
            Y   D+RN LERSSARETT+RVA GA+AK+IL+  GIE+ +H++  GG +         +
Sbjct: 121  YGHRDMRNVLERSSARETTVRVAAGAVAKKILRTFGIEVGSHVLEIGGVKAEKTSYDQLS 180

Query: 179  QIKVLSSQSQVAIVNPSFEQEIKDYIDSVKKAGDTIGGVVETIVGGVPVGLGSYVHWDRK 238
             +K L+  S V  ++   EQE+   ID  K+ GD+IGGVVE IV GVP+GLGS+VH+DRK
Sbjct: 181  NLKELAEASPVRCLDKEAEQEMIAAIDQAKENGDSIGGVVEVIVEGVPIGLGSHVHYDRK 240
```

```
Query: 239 LDAKIAQAVVSINAFKGVEFGLGFKSGFLKGSQVMDSISWTKDQGYIRQSNNLGGFEGGM 298
           LDAKIA AV+SINAFKGVEFG+GF++    GS+V D I+W +++GY R+SNNLGGFEGGM
Sbjct: 241 LDAKIAAAVMSINAFKGVEFGIGFEAASKPGSEVHDEIAWDEERGYYRKSNNLGGFEGGM 300

Query: 299 TNGEPIIVRGVMKPIPTLYKPLMSVDIDTHEPYRATVERSDPTALPAAGVVMEAVVATVL 358
           TNG PI+VRGVMKPIPTLYKPL SVDI T EP+ A++ERSD  A+PAA VV EAVVA  +
Sbjct: 301 TNGMPIVVRGVMKPIPTLYKPLQSVDIATKEPFAASIERSDSCAVPAAAVVAEAVVAWEV 360

Query: 359 VTEVLEKFSSDNMYELKEAVK 379
              +LE+F +D + E+++ ++
Sbjct: 361 ANALLERFGADQVEEIEKNIR 381
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1883> which encodes the amino acid sequence <SEQ ID 1884>. Analysis of this protein sequence reveals the following:

```
     Possible site: 15
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.75   Transmembrane  342 - 358 ( 342 - 359)
   INTEGRAL    Likelihood = -0.16   Transmembrane  155 - 171 ( 155 - 171)

----- Final Results -----
               bacterial membrane --- Certainty=0.1298(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:

```
>GP:BAB05375 GB:AP001512 chorismate synthase [Bacillus halodurans]
 Identities = 213/390 (54%), Positives = 277/390 (70%), Gaps = 2/390 (0%)

Query: 1   LRYLTAGESHGPSLTAIIEGIPAGLTLHPADIDHELQRRQGGYGRGARMSIETDRVQISS 60
           +RYLTAGESHGP LT IIEG PA L L   DI+ +L RRQGG+GRG RM IE D+VQI
Sbjct: 1   MRYLTAGESHGPQLTTIIEGAPAQLELVADDINVDLARRQGGHGRGRRMQIEKDQVQIVG 60

Query: 61  GVRHGKTTGAPITLTVINKDHQKWLDVMAVGDI--EETLKLKRRVKHPRPGHADLVGGIK 118
           G+RHGKTTGAPI L V NKD + W  +M    +  +E  ++KR++  PRPGHADL G IK
Sbjct: 61  GIRHGKTTGAPIALVVENKDWKHWTKIMGAEPLTGDEEKEIKRKITRPRPGHADLNGAIK 120

Query: 119 YHFNDLRDALERSSARETTMRVAVGAVAKRILAELGIDMLHHILIFGGITITIPSKLSFR 178
           Y   D+R+ LERSSARETT+RVA GAVAK+IL  GI++  H+L  GG+     S
Sbjct: 121 YGHRDMRNVLERSSARETTVRVAAGAVAKKILRTFGIEVGSHVLEIGGVKAEKTSYDQLS 180

Query: 179 ELQERALHSELSIVNPKQEEEIKTYIDKIKKEGDTIGGIIETIVQGVPAGLGSYVQWDKK 238
            L+E A  S +  ++ + E+E+   ID+ K+ GD+IGG++E IV+GVP GLGS+V +D+K
Sbjct: 181 NLKELAEASPVRCLDKEAEQEMIAAIDQAKENGDSIGGVVEVIVEGVPIGLGSHVHYDRK 240

Query: 239 LDAKLAQAVLSINAFKGVEFGAGFDMGFQKGSQVMDEITWTPTQGYGRQTNHLGGFEGGM 298
           LDAK+A AV+SINAFKGVEFG GF+   + GS+V DEI W   +GY R++N+LGGFEGGM
Sbjct: 241 LDAKIAAAVMSINAFKGVEFGIGFEAASKPGSEVHDEIAWDEERGYYRKSNNLGGFEGGM 300

Query: 299 TTGQPLVVKGVMKPIPTLYKPLMSVDIDSHEPYKATVERSDPTALPAAGVIMENVVATVL 358
           T G P+VV+GVMKPIPTLYKPL SVDI + EP+ A++ERSD  A+PAA V+ E VVA  +
Sbjct: 301 TNGMPIVVRGVMKPIPTLYKPLQSVDIATKEPFAASIERSDSCAVPAAAVVAEAVVAWEV 360

Query: 359 AKEILETFSSTTMSELQKAFSDYRAYVKQF 388
           A  +LE F +  + E++K   ++   + F
Sbjct: 361 ANALLERFGADQVEEIEKNIREFNEKARLF 390
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 284/388 (73%), Positives = 333/388 (85%)

Query: 1   MRYLTAGESHGPSLTAIIEGIPAGLKLSAKDINEDLKRRQGGYGRGNRMKIETDQVIISS 60
           +RYLTAGESHGPSLTAIIEGIPAGL L   DI+ +L+RRQGGYGRG RM IETD+V ISS
Sbjct: 1   LRYLTAGESHGPSLTAIIEGIPAGLTLHPADIDHELQRRQGGYGRGARMSIETDRVQISS 60
```

```
Query: 61  GVRHGKTLGSPITLTVTNKDHSKWLDIMSVEDIEERLKQKRRIKHPRPGHADLVGGIKYR 120
           GVRHGKT G+PITLTV NKDH KWLD+M+V DIEE LK KRR+KHPRPGHADLVGGIKY
Sbjct: 61  GVRHGKTTGAPITLTVINKDHQKWLDVMAVGDIEETLKLKRRVKHPRPGHADLVGGIKYH 120

Query: 121 FDDLRNALERSSARETTMRVAIGAIAKRILKEIGIEIANHIVVFGGKEITVPDKLTVQQI 180
           F+DLR+ALERSSARETTMRVA+GA+AKRIL E+GI++ +HI++FGG  IT+P KL+ +++
Sbjct: 121 FNDLRDALERSSARETTMRVAVGAVAKRILAELGIDMLHHILIFGGITITIPSKLSFREL 180

Query: 181 KVLSSQSQVAIVNPSFEQEIKDYIDSVKKAGDTIGGVVETIVGGVPVGLGSYVHWDRKLD 240
           +  +  S+++IVNP  E+EIK YID +KK GDTIGG++ETIV GVP GLGSYV WD+KLD
Sbjct: 181 QERALHSELSIVNPKQEEEIKTYIDKIKKEGDTIGGIIETIVQGVPAGLGSYVQWDKKLD 240

Query: 241 AKIAQAVVSINAFKGVEFGLGFKSGFLKGSQVMDSISWTKDQGYIRQSNNLGGFEGGMTN 300
           AK+AQAV+SINAFKGVEFG GF  GF KGSQVMD I+WT  QGY RQ+N+LGGFEGGMT
Sbjct: 241 AKLAQAVLSINAFKGVEFGAGFDMGFQKGSQVMDEITWTPTQGYGRQTNHLGGFEGGMTT 300

Query: 301 GEPIIVRGVMKPIPTLYKPLMSVDIDTHEPYRATVERSDPTALPAAGVVMEAVVATVLVT 360
           G+P++V+GVMKPIPTLYKPLMSVDID+HEPY+ATVERSDPTALPAAGV+ME VVATVL
Sbjct: 301 GQPLVVKGVMKPIPTLYKPLMSVDIDSHEPYKATVERSDPTALPAAGVIMENVVATVLAK 360

Query: 361 EVLEKFSSDNMYELKEAVKLYRNYVDHF 388
           E+LE FSS  M EL++A   YR YV  F
Sbjct: 361 EILETFSSTTMSELQKAFSDYRAYVKQF 388
```

A related GBS gene <SEQ ID 8617> and protein <SEQ ID 8618> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:      -2.42
GvH: Signal Score (-7.5): -3.23
     Possible site: 15
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -4.67 threshold:  0.0
   INTEGRAL    Likelihood = -4.67   Transmembrane  343 - 359 ( 341 - 364)
   PERIPHERAL  Likelihood =  0.69     214
 modified ALOM score:   1.43

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2869(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:

```
57.7/73.8% over 354aa
                                                                  Bacillus subtilis
  EGAD|20299| chorismate synthase Insert characterized
  SP|P31104|AROC_BACSU CHORISMATE SYNTHASE (EC 4.6.1.4) (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE
PHOSPHOLYASE)
 (VEGETATIVE PROTEIN 216) (VEG216). Edit characterized
  GP|143806|gb|AAA20859.1||M80245 AroF Insert characterized
  GP|2634689|emb|CAB14187.1||Z99115 chorismate synthase Insert characterized
  PIR|C69590|C69590 chorismate synthase aroF - Insert characterized

ORF00121(301 - 1359 of 1719)
EGAD|20299|BS2267(1     -     355     of     368)     chorismate     synthase     (Bacillus
subtilis}SP|P31104|AROC_BACSU CHORISMATE SYNTHASE (EC 4.6.1.4) (5-ENOLPYRUVYLSHIKIMATE-3-
PHOSPHATE PHOSPHOLYASE) (VEGETATIVE PROTEIN 216) (VEG216).GP|143806|gb|AAA20859.1||M80245
AroF {Bacillus subtilis}GP|2634689|emb|CAB14187.1||Z99115 chorismate synthase {Bacillus
subtilis}PIR|C69590|C69590 chorismate synthase aroF - Bacillus subtilis
%Match = 35.0
%Identity = 57.6  %Similarity = 73.7
Matches = 204  Mismatches = 92  Conservative Sub.s = 57

75       105       135       165       195       225       255       285
```

```
IQLSRVAERKNLMPRGISQDIYNMCLKFGLPVHYAEWDKDVLFDILSHDKKASGQFIKIVILPQLGSATVHQIPLEEMRD

315       345       375       405       435       465       495       525
YLEK*MRYLTAGESHGPSLTAIIEGIPAGLKLSAKDINEDLKRRQGGYGRGNRMKIETDQVIISSGVRHGKTLGSPITLT
     |||||||||||| || ||||:|||| :: :||| :| ||| |:||| ||:|| ||  | ||||| :|||||| |
     MRYLTAGESHGPQLTTIIEGVPAGLYITEEDINFELARRQKGHGRGRRMQIEKDQAKIMSGVRHARTLGSPIALV
               10        20        30        40        50        60        70

555                 609       639       669       699       729       759
VTNKDHSKWLDIMSVEDI--EERLKQKRRIKHPRPGHADLVGGIKYRFDDLRNALERSSARETTMRVAIGAIAKRILKEI
| | |   | ||    | :| : ||:| |||||||| | |||   |:|| |||||||||:||| ||:||:|| |:
VENNDWKHWTKIMGAAPITEDEEKEMKRQISRPRPGHADLNGAIKYNHRDMRNVLERSSARETTVRVAAGAVAKKILSEL
               90        100       110       120       130       140       150

789       819       849       879       909       939       969       999
GIEIANHIVVFGGKEITVPDKLTVQQIKVLSSQSQVAIVNPSFEQEIKDYIDSVKKAGDTIGGVVETIVGGVPVGLGSYV
||::| |::  |  :       ::: :: :: :| |   :    :::   || |  ||:|||:|| || |:|||:||||
GIKVAGHVLQIGAVKAEKTGYTSIEDLQRVTEESPVRCYDEEAGKKMMAAIDEAKANGDSIGGIVEVIVEGMPVGVGSYV
               170       180       190       200       210       220       230

1029      1059      1089      1119      1149      1179      1209      1239
HWDRKLDAKIAQAVVSINAFKGVEFGLGFKSGFLKGSQVMDSISWTKDQGYIRQSNNLGGFEGGMTNGEPIIVRGVMKPI
|:|||||:|:| ||:||||||||||:||::    ||:| | | | :::|| | :| |||:||||| | ||:|||||||||
HYDRKLDSKLAAAVLSINAFKGVEFGIGFEAAGRNGSEVHDEIIWDEEKGYTRATNRLGGLEGGMTTGMPIVVRGVMKPI
               250       260       270       280       290       300       310

1269      1299      1329      1359      1389      1419      1449      1479
PTLYKPLMSVDIDTHEPYRATVERSDPTALPAAGVVMEAVVATVLVTEVLEKFSSDNMYN*KKL*NYIAIMLIIFK*KLV
|||||||  ||||:| ||: |::|||||  |:||| || ||:    |   :
PTLYKPLKSVDIETKEPFSASIERSDSCAVPAASVVAEALSLGKLQPSLNNSD
               330       340       350       360
```

SEQ ID 8618 (GBS192) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 4; MW 44kDa).

GBS192-His was purified as shown in Figure 196, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 608

A DNA sequence (GBSx0648) was identified in *S.agalactiae* <SEQ ID 1885> which encodes the amino acid sequence <SEQ ID 1886>. This protein is predicted to be 3-dehydroquinate synthase (aroB). Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood = -3.82   Transmembrane   99 - 115 (  98 - 116)

----- 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>
```

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

```
>GP:BAA18068 GB:D90911 3-dehydroquinate synthase [Synechocystis sp.]
 Identities = 138/351 (39%), Positives = 200/351 (56%), Gaps = 4/351 (1%)

Query: 3   VEVDLPNHPYHIKIEEGCFSEAGDWVSHLWQKQMITIITDSNVEILYGESLVNQLKKQGF 62
           + V LP  PY ++I  G  +   D ++ L   + I ++++  +   YGE ++  L++ G+
Sbjct: 5   IPVPLPQSPYQVQIVPGGLAAIADHLAPLGLGKKIMVVSNPEIYDYYGEVVIQALQRAGY 64
```

```
Query: 63  TVHVFSFAAGEASKTLEVANRIYAFLAKHHMTRSDGIIALGGGVVGDLAAFVASTYMRGI 122
           V      AGE  KTL   N +Y    + ++ R+  +++LGGGV+GD+  F A+T++RGI
Sbjct: 65  EVFQHLIPAGETHKTLASINELYDVAFQANLERNSTLLSLGGGVIGDMTGFGAATWLRGI 124

Query: 123 HFLQIPTSLTAQVDSSIGGKTGVNTSFAKNMVGTFAQPDGVLIDPVTLKTLGNRELVEGM 182
           +F+Q+PTSL A VD+SIGGKTGVN    KN++G F QP  V IDPV LKTL  RE   GM
Sbjct: 125 NFVQVPTSLLAMVDASIGGKTGVNHPQGKNLIGAFYQPRLVYIDPVVLKTLPEREFRAGM 184

Query: 183 GEVIKYGLIDDIKLWHILEEMD--GTIDSILDNALA-IIYHSCQVKRKHVLADQYDKGLR 239
           EVIKYG+I D +L+  LEE +   +ID + D  L  II  SCQ K   V  D+ + GLR
Sbjct: 185 AEVIKYGVIWDSELFTALEEAEDLSSIDRLFDELLTKIIQRSCQAKVDVVSQDEKEAGLR 244

Query: 240 MHLNFGHTIGHAIEVHAGYGEIMHGEAVAIGMIQLSRVAERKNLMPRGISQDIYNMCLKF 299
             LN+GHT+GH +E   GYG I HGEAVAIGM   +++A    L  + +      + LK
Sbjct: 245 AILNYGHTVGHGVESLTGYGVINHGEAVAIGMEAAAKIAHYLGLCDQSLGDRQRQLLLKT 304

Query: 300 GLPVHY-AEWDKDVLFDILSHDKKASGQFIKIVILPQLGSATVHQIPLEEM 349
           LP          + L   L HDKK     ++ ++   +G  T+     +E+
Sbjct: 305 KLPTEMPPTLAVENLLASLLHDKKVKAGKVRFILPTAIGQVTISDAVTDEV 355
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1887> which encodes the amino acid sequence <SEQ ID 1888>. Analysis of this protein sequence reveals the following:

```
     Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.43   Transmembrane   97 - 113 (  97 - 114)

----- Final Results -----
               bacterial membrane --- Certainty=0.1171(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:

```
>GP:BAA18068 GB:D90911 3-dehydroquinate synthase [Synechocystis sp.]
 Identities = 123/349 (35%), Positives = 190/349 (54%), Gaps = 9/349 (2%)

Query: 1   MPQTLHVHSRVKDYDILFTDHVLKTLADCLGERKQ-RKLLFITDQTVYHLYQTLFEEFAQ 59
           M  T+ V      Y +      L  +AD L      +K++ +++  +Y  Y  +  +  Q
Sbjct: 1   MATTIPVPLPQSPYQVQIVPGGLAAIADHLAPLGLGKKIMVVSNPEIYDYYGEVVIQALQ 60

Query: 60  Q--YNAFVHVCPPGGQSKSLERVSAIYDQLIAENFSKKDMIVTIGGGVVGDLGGFVAATY 117
           +  Y  F H+ P G   K+L  ++ +YD     N  +   ++++GGGV+GD+ GF AAT+
Sbjct: 61  RAGYEVFQHLIPAGETHKTLASINELYDVAFQANLERNSTLLSLGGGVIGDMTGFGAATW 120

Query: 118 YRGIPYIQIPTTLLSQVDSSIGGKVGVHFKGLTNMIGSIYPPEAIIISTTFLETLPQREF 177
           RGI ++Q+PT+LL+ VD+SIGGK GV+     N+IG+ Y P  + I    L+TLP+REF
Sbjct: 121 LRGINFVQVPTSLLAMVDASIGGKTGVNHPQGKNLIGAFYQPRLVYIDPVVLKTLPEREF 180

Query: 178 SCGISEMLKIGFIHDRPLFQQLRDFQ-----KETDKQGLERLIYQSISNKKRIVEQDEFE 232
           + G++E++K G I D  LF  L + +          + L ++I +S   K  +V QDE E
Sbjct: 181 RAGMAEVIKYGVIWDSELFTALEEAEDLSSIDRLPDELLTKIIQRSCQAKVDVVSQDEKE 240

Query: 233 NGLRMSLNFGHTLGHAIESLCHHDFYHHGEAIAIGMVVDAKLAVSKGLLPKEDLDSLLQV 292
           + GLR  LN+GHT+GH +ESL   +HGEA+AIGM   AK+A   GL  +  D   Q+
Sbjct: 241 AGLRAILNYGHTVGHGVESLTGYGVINHGEAVAIGMEAAAKIAHYLGLCDQSLGDRQRQL 300

Query: 293 FERYQLPTTLERADVSATSLFDVFKTDKKNSEQHIIFILPTETGFTTLA 341
           F  +LPT +    ++  +L      DKK     + FILPT  G  T++
Sbjct: 301 LLKTKLPTEMP-PTLAVENLLASLLHDKKVKAGKVRFILPTAIGQVTIS 348
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 121/332 (36%), Positives = 182/332 (54%), Gaps = 7/332 (2%)

Query: 12  YHIKIEEGCFSEAGDWVSHLWQKQMITIITDSNVEILYGESLVNQLKKQGFTVHVFSFAA 71
           Y I   +      D +    Q++++  ITD  V  LY ++L  +  +Q +   V
```

```
Sbjct: 14  YDILFTDHVLKTLADCLGERKQRKLL-FITDQTVYHLY-QTLFEEFAQQ-YNAFVHVCPP 70

Query: 72  GEASKTLEVANRIYAFLAKHHMTRSDGIIALGGGVVGDLAAFVASTYMRGIHFLQIPTSL 131
           G  SK+LE  + IY  L    + ++ D I+ +GGGVVGDL  FVA+TY RGI ++QIPT+L
Sbjct: 71  GGQSKSLERVSAIYDQLIAENFSKKDMIVTIGGGVVGDLGGFVAATYYRGIPYIQIPTTL 130

Query: 132 TAQVDSSIGGKTGVNTSFAKNMVGTFAQPDGVLIDPVTLKTLGNRELVEGMGEVIKYGLI 191
            +QVDSSIGGK GV+     NM+G+   P+ ++I    L+TL  RE   G+ E++K G I
Sbjct: 131 LSQVDSSIGGKVGVHFKGLTNMIGSIYPPEAIIISTTFLETLPQREFSCGISEMLKIGFI 190

Query: 192 DDIKLWHILEEMDGTIDSILDNALAIIYHSCQVKRKHVLADQYDKGLRMHLNFGHTIGHA 251
            D  L+  L +     D         +IY S   K++ V  D+++ GLRM LNFGHT+GHA
Sbjct: 191 HDRPLFQQLRDFQKETDK--QGLERLIYQSISNKKRIVEQDEFENGLRMSLNFGHTLGHA 248

Query: 252 IEVHAGYGEIMHGEAVAIGMIQLSRVAERKNLMPRGISQDIYNMCLKFGLP--VHYAEWD 309
           IE    +    HGEA+AIGM+  +++A  K L+P+     +  ++ LP  +  A+
Sbjct: 249 IESLCHHDFYHHGEAIAIGMVVDAKLAVSKGLLPKEDLDSLLQVFERYQLPTTLERADVS 308

Query: 310 KDVLFDILSHDKKASGQFIKIVILPQLGSATV 341
              LFD+   DKK S Q I  ++  + G  T+
Sbjct: 309 ATSLFDVFKTDKKNSEQHIIFILPTETGFTTL 340
```

SEQ ID 1886 (GBS336) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 2; MW 42.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 5; MW 68kDa).

The GBS336-GST fusion product was purified (Figure 209, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 310), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 609

A DNA sequence (GBSx0649) was identified in *S.agalactiae* <SEQ ID 1889> which encodes the amino acid sequence <SEQ ID 1890>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

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

A related GBS nucleic acid sequence <SEQ ID 9973> which encodes amino acid sequence <SEQ ID 9974> was also identified.

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

```
>GP:CAB14240 GB:Z99116 3-dehydroquinate dehydratase [Bacillus subtilis]
 Identities = 70/233 (30%), Positives = 127/233 (54%), Gaps = 12/233 (5%)

Query: 2   KIVVPVMPRSLEEA-QEIDLSKFDSVDIIEWRADALPK----DDIINVAPAIFEKFAGHE 56
           KI++P+M ++ ++    E +  K  + DI+EWR D    K     + +  +    + +
Sbjct: 17  KIIIPLMGKTEKQILNEAEAVKLLNPDIVEWRVDVFEKANDREAVTKLISKLRKSLEDKL 76

Query: 57  IIFTLRTTREGGNIVLSDAEYVELIQKINSIYNPDYIDFEYFSHKEVFQEMLEFPN---- 112
            +FT RT +EGG++ + ++ Y+ L++       + D ID E FS      + ++
Sbjct: 77  FLFTFRTHKEGGSMEMDESSYLALLESAIQTKDIDLIDIELFSGDANVKALVSLAEENNV 136
```

```
Query: 113 -LVLSYHNFQETP--ENIMEIFSELTALAPRVVKIAVMPKNEQDVLDVMNYTRGFKTINP 169
            +V+S H+F++TP  + I+    ++  L   + K+AVMP +  D+L +++ T  KTI
Sbjct: 137 YVVMSNHDFEKTPVKDEIISRLRKMQDLGAHIPKMAVMPNDTGDLLTLLDATYTMKTIYA 196

Query: 170 DQVYATVSMSKIGRISRFAGDVTGSSWTFAYLDSSIAPGQITISEMKRVKALL 222
           D+   T+SM+  G ISR +G+V GS+ TF   + + APGQI +SE++ V  +L
Sbjct: 197 DRPIITMSMAATGLISRLSGEVFGSACTFGAGEEASAPGQIPVSELRSVLDIL 249
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1891> which encodes the amino acid sequence <SEQ ID 1892>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3248(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 160/225 (71%), Positives = 198/225 (87%)

Query: 1   MKIVVPVMPRSLEEAQEIDLSKFDSVDIIEWRADALPKDDIINVAPAIFEKFAGHEIIFT 60
           M+IV PVMPR  +EAQ ID+SK++ V++IEWRAD LPKD+I+ VAPAIFEKFAG EIIFT
Sbjct: 1   MRIVAPVMPRHFDEAQAIDISKYEDVNLIEWRADFLPKDEIVAVAPAIFEKFAGKEIIFT 60

Query: 61  LRTTREGGNIVLSDAEYVELIQKINSIYNPDYIDFEYFSHKEVFQEMLEFPNLVLSYHNF 120
           LRT +EGGNI LS  EYV++I++IN+IYNPDYIDFEYF+HK VFQEML+FPNL+LSYHNF
Sbjct: 61  LRTVQEGGNITLSSQEYVDIIKEINAIYNPDYIDFEYFTHKSVFQEMLDFPNLILSYHNF 120

Query: 121 QETPENIMEIFSELTALAPRVVKIAVMPKNEQDVLDVMNYTRGFKTINPDQVYATVSMSK 180
           +ETPEN+ME FSE+T LAPRVVKIAVMP++EQDVLD+MNYTRGFKT+NP+Q +AT+SM K
Sbjct: 121 EETPENLMEAFSEMTKLAPRVVKIAVMPQSEQDVLDLMNYTRGFKTLNPEQEFATISMGK 180

Query: 181 IGRISRFAGDVTGSSWTFAYLDSSIAPGQITISEMKRVKALLDAD 225
           +GR+SRFAGDV GSSWT+  LD     PGQ+T+++MKR+  +L+ D
Sbjct: 181 LGRLSRFAGDVIGSSWTYVSLDHVSGPGQVTLNDMKRIIEVLEMD 225
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 610

A DNA sequence (GBSx0650) was identified in *S.agalactiae* <SEQ ID 1893> which encodes the amino acid sequence <SEQ ID 1894>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1195(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 611

A DNA sequence (GBSx0651) was identified in *S.agalactiae* <SEQ ID 1895> which encodes the amino acid sequence <SEQ ID 1896>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3431(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 GENPEPT database:

```
>GP:CAB15862 GB:Z99123 alternate gene name: ipa-19d-similar to
          hypothetical proteins [Bacillus subtilis]
 Identities = 161/396 (40%), Positives = 235/396 (58%), Gaps = 11/396 (2%)

Query: 1   MNKLKVNSVVERKIKSGAQLLEKKDFDTSLVNQ----LVQLFSQSN-QFLGMAYLSPQNK 55
           M  L +       KIK G  L+EK+    S  +     LV + S+S  +FL   Y   QNK
Sbjct: 1   MKLLTLKKAHAAKIKKGYPLIEKEALAGSAGHMKEGDLVDIVSESGGEFLARGYYGLQNK 60

Query: 56  GIGWLLSRQVFD-FNHDYFVSLFEKSREKRQKFEKSSQTTAYRLFNQDGDNFGGLTIDFY 114
           G+GW L+R   +  +  +F+S   K+ + R K  ++  TTA+RLFN +GD  GG+TID+Y
Sbjct: 61  GVGWTLTRNKHEQIDQAFFLSKLTKAAQARAKLFEAQDTTAFRLFNGEGDGVGGVTIDYY 120

Query: 115 SDYALFSWYNEFVYTNRQMIVAAFKQVYPNIKGAYEKIRFKGLDF---ESAHLYGQEAPE 171
             Y L  WY++ +YT + M+++A  ++  + K  YEK RF         +   + G+
Sbjct: 121 DGYLLIQWYSKGIYTFKDMLISALDEMDLDYKAIYEKKRFDTAGQYVEDDDFVKGRRGEF 180

Query: 172 SFLILENNIKYSVFLNDGLMTGIFLDQHDVRKALATNLSEGKKVLNMFSYTAAFSVAAAV 231
             +I EN I+Y+V LN+G MTGIFLDQ  VRKA+    ++GK VLN FSYT AFSVAAA+
Sbjct: 181 PIIIQENGIQYAVDLNEGAMTGIFLDQRHVRKAIRDRYAKGKTVLNTFSYTGAFSVAAAL 240

Query: 232 GGALETTSVDLAKRSRELSKAHFDANQIVTDNHRFIVMDVFEYYKYAKRKHLSYDVIVID 291
           GGA +TTSVD+A RS   +   F  N++  + H   VMDVF Y+ YA +K L +D+I++D
Sbjct: 241 GGAEKTTSVDVANRSLAKTIEQFSVNKLDYEAHDIKVMDVFNYFSYAAKKDLRFDLIILD 300

Query: 292 PPSFARNKKQTFSVTKDYYKLIEQALDILTPGGTIIASTNAANLTVSQFKKQLEKGFGKA 351
           PPSFAR KK+TFS  KDY  L+++ + I    G I+ASTN++   + +FK  ++  F +
Sbjct: 301 PPSFARTKKRTFSAAKDYKNLLKETIAITADKGVIVASTNSSAFGMKKFKGFIDAAFKET 360

Query: 352 SHNYISLQQ--LPEDFTINDKDQQSNYLKVFTIKVK 385
           +  Y  +++  LPEDF       + NYLKV  ++ K
Sbjct: 361 NERYTIIEEFTLPEDFKTISAFPEGNYLKVVLLQKK 396
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1897> which encodes the amino acid sequence <SEQ ID 1898>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2699(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 259/386 (67%), Positives = 315/386 (81%), Gaps = 1/386 (0%)

Query: 1   MNKLKVNSVVERKIKSGAQLLEKKDFDT-SLVNQLVQLFSQSNQFLGMAYLSPQNKGIGW 59
           MNKL ++S VE+K+ +G QLL++KDF      NQLVQL ++SN+ +G AY+S QNKGIGW
Sbjct: 1   MNKLYIDSFVEKKLTAGVQLLDEKDFSNIKEKNQLVQLVTKSNRPIGTAYISKQNKGIGW 60
```

```
Query: 60  LLSRQVFDFNHDYFVSLFEKSREKRQKFEKSSQTTAYRLFNQDGDNFGGLTIDFYSDYAL 119
           L  +  D +  YFVSLF  ++ KRQ F +S +T AYRLFNQ+GD FGG+TID Y D+A+
Sbjct: 61  YLGPEKIDLSISYFVSLFSVAKAKRQDFAQSDETNAYRLFNQEGDGFGGVTIDLYKDFAV 120

Query: 120 FSWYNEFVYTNRQMIVAAFKQVYPNIKGAYEKIRFKGLDFESAHLYGQEAPESFLILENN 179
           FSWYN FVY  ++MI+ AF+QV+P +KGAYEK RFKG D E+AHLYG+ A E+F ILEN
Sbjct: 121 FSWYNAFVYDKKEMIMEAFQQVFPEVKGAYEKCRFKGPDTETAHLYGELAQETFSILENG 180

Query: 180 IKYSVFLNDGLMTGIFLDQHDVRKALATNLSEGKKVLNMFSYTAAFSVAAAVGGALETTS 239
           I Y VFLN+GLMTGIFLDQHDVR+AL   L+ GK +LN+FSYTAAFSVAAA+GGA+ETTS
Sbjct: 181 IAYQVFLNEGLMTGIFLDQHDVRRALVDGLAMGKSLLNLFSYTAAFSVAAAMGGAIETTS 240

Query: 240 VDLAKRSRELSKAHFDANQIVTDNHRFIVMDVFEYYKYAKRKHLSYDVIVIDPPSFARNK 299
           VDLAKRSRELS AHF+ NQ+   +H F+VMDVFEY+KYAKRK L +DVIVIDPPSFARNK
Sbjct: 241 VDLAKRSRELSLAHFEHNQLNLASHHFVVMDVFEYFKYAKRKKLIFDVIVIDPPSFARNK 300

Query: 300 KQTFSVTKDYYKLIEQALDILTPGGTIIASTNAANLTVSQFKKQLEKGFGKASHNYISLQ 359
           KQTFSV++DY+KLI +ALDIL+P GTIIASTNAAN+TVSQFKKQ+ KGFG      ++LQ
Sbjct: 301 KQTFSVSRDYHKLITEALDILSPKGTIIASTNAANMTVSQFKKQIIKGFGSRRPESMTLQ 360

Query: 360 QLPEDFTINDKDQQSNYLKVFTIKVK 385
           QLP DFTIN  D++SNYLKVFTIKV+
Sbjct: 361 QLPSDFTINKADERSNYLKVFTIKVR 386
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 612

A DNA sequence (GBSx0652) was identified in *S.agalactiae* <SEQ ID 1899> which encodes the amino acid sequence <SEQ ID 1900>. This protein is predicted to be minimal change nephritis transmembrane glycoprotein. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.85   Transmembrane  129 - 145 ( 126 - 152)
   INTEGRAL    Likelihood = -4.88   Transmembrane   48 -  64 (  46 -  69)
   INTEGRAL    Likelihood = -4.83   Transmembrane   75 -  91 (  74 -  97)
   INTEGRAL    Likelihood = -4.62   Transmembrane   16 -  32 (  15 -  34)
   INTEGRAL    Likelihood = -2.28   Transmembrane  163 - 179 ( 163 - 182)

----- Final Results -----
              bacterial membrane --- Certainty=0.3739(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 GENPEPT database:

```
>GP:CAB12545 GB:Z99107 alternate gene name: yetP-similar to
          hypothetical proteins [Bacillus subtilis]
 Identities = 299/676 (44%), Positives = 415/676 (61%), Gaps = 33/676 (4%)

Query: 2   KKIKDFASRAINTRLGFILLLVVIYWLKTIWAYHTDFNLGLENSYQLFLTIINPIPLGLL 61
           KK++    +  + +L F +L V+++W KT  +Y T+FNLG++ + Q  L I NP    +
Sbjct: 9   KKVEVAMKKLFSYKLSFFVLAVILFWAKTYLSYKTEFNLGVKGTTQEILLIFNPFSSAVF 68

Query: 62  IIGLALYVKRTKAFYITAFITYAIVNILLIANAIYYREFSDFITVSAVLASSKTSAGLGD 121
            +GLAL  K  K+  I   I + ++  +L AN ++YR F DF+T   +  S       +GD
Sbjct: 69  FLGLALLAKGRKSAIIMLIIDF-LMTFVLYANILFYRFFDDFLTFPNIKQSGNVG-NMGD 126

Query: 122 SALNLLRIWDLVYVFDFIILIFLFATKKIHLDDRPFNKRASFSITALSGL-LFSINLFLA 180
              +++   D+ Y  D IILI +   +    L +    KR + S+  LSG+ LF INL  A
Sbjct: 127 GIFSIMAGHDIFYFLDIIILIAVLIWRP-ELKEYKMKKRFA-SLVILSGIALFFINLHYA 184

Query: 181 EIDRPELLSRGFSNTYIVKALGLPSFSIYSGNQTYQAQKERNGATAQELATAKKYVAEHY 240
```

```
            E DRP+LL+R F   YIVK LGL +++IY G QT Q + +R  A++ +L + + Y   HY
Sbjct: 185 EKDRPQLLTRTFDRNYIVKYLGLYNYTIYDGVQTAQTETQRAYASSDDLTSVENYTTSHY 244

Query: 241 AKPNPEYYGIGKGRNVIMIHLESFQQFLIDYKLNIDGKEHVVTPFINSLYHSKETVS-FS 299
           AKPN EY+G  KG+N+I IHLESFQ FLIDYKLN  G+E  VTPF+N L H  E V+ F
Sbjct: 245 AKPNAEYFGSAKGKNIIKIHLESFQSFLIDYKLN--GEE--VTPFLNKLAHGGEDVTYFD 300

Query: 300 NFFHQVKAGKTSDAETLMENSLFGLSSGSFMVNYGGENTQFAAPHILAQNGGYSSAVFHG 359
           NFFHQ   GKTSDAE  M+NS+FGL  GS  V   GENT  + P IL Q  GY+SAV HG
Sbjct: 301 NFFHQTGQGKTSDAELTMDNSIFGLPEGSAFVT-KGENTYQSLPAILDQKEGYTSAVLHG 359

Query: 360 NVGTFWNRNNAYKQWGYDYFFDSSYFSKQTKDNSFQYGLNDKYMFADSIKYLEHMQQPFY 419
           +  +FWNR+  YK  GYD FFD+S +   + +N    GL DK  F +SI  LE ++QPFY
Sbjct: 360 DYKSFWNRDQIYKHIGYDKFFDASTYD-MSDENVINMGLKDKPFFTESIPKLESLKQPFY 418

Query: 420 TKFITVSNHYPYTSLKGESDEEGFPLAKTNDETINGYFATANYLDTALKSFFEYLKAAGV 479
              IT++NHYP+     +  +     A T D T++ YF TA YLD AL+ FF+ LK AG+
Sbjct: 419 AHLITLTNHYPFNL---DEKDASLKKATTGDNTVDSYFQTARYLDEALEQFFKELKEAGL 475

Query: 480 YDNSIIVMYGDHYGISNTRNPSLAELLGKDPETWSEYDNAMLQRVPYMIHIPGYSKGFIS 539
           YDNS+I++YGDH GIS   N ++ E+LGK+    ++Y NA  QRVP MI +PG  KG ++
Sbjct: 476 YDNSVIMIYGDHNGISENHNRAMKEILGKE---ITDYQNAQNQRVPLMIRVPG-KKGGVN 531

Query: 540 NTYGGEVDNLPTLLHILGIDTSKYTQLGQDLLSKDNKQMVAMRTTGQYITPKYTNYSGHL 599
           +TYGGE+D +PTLLH+ GID+ KY   G DL SKD+   VA R  G ++TPKYT+    +
Sbjct: 532 HTYGGEIDVMPTLLHLEGIDSQKYINFGTDLFSKDHDDTVAFR-NGDFVTPKYTSVDNII 590

Query: 600 YYTDSGQEITNPDETTKAEIKAIRDATNKQLSTSDSIQTGDLLRFDENNGLKTVEVEKFN 659
           Y T +G+++   +ET     K ++   N+QLS SDS+   DLLRF + N  K V+   ++
Sbjct: 591 YDTKTGEKLKANEET-----KNLKTRVNQQLSLSDSVLYKDLLRFHKLNDFKAVDPSDYH 645

Query: 660 YTHSLKALKAKERKLK 675
           Y         KE+++K
Sbjct: 646 Y--------GKEKEIK 653
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1901> which encodes the amino acid sequence <SEQ ID 1902>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.85   Transmembrane   90 - 106 (  88 - 112)
   INTEGRAL    Likelihood = -5.68   Transmembrane  146 - 162 ( 139 - 165)
   INTEGRAL    Likelihood = -4.99   Transmembrane   63 -  79 (  60 -  84)
   INTEGRAL    Likelihood = -3.98   Transmembrane  178 - 194 ( 176 - 197)
   INTEGRAL    Likelihood = -0.59   Transmembrane   31 -  47 (  31 -  47)

----- Final Results -----
               bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 533/713 (74%), Positives = 603/713 (83%)

Query: 1   MKKIKDFASRAINTRLGFILLLVVIYWLKTIWAYHTDFNLGLENSYQLFLTIINPIPLGL 60
           +KK K   +  INTRLGFI+ L+  YW+KT+WAYHTDF+L L N YQ+FLTIINPIPL
Sbjct: 16  VKKFKTLITGFINTRLGFIITLLFCYWIKTLWAYHTDFSLDLGNIYQVFLTIINPIPLAF 75

Query: 61  LIIGLALYVVKRTKAFYITAFITYAIVNILLLIANAIYYREFSDFITVSAVLASSKTSAGLG 120
           L++G+ALYVK T+AFYI +++ Y I+NILLI+N+IYYREFSDFITVSA+LASSK SAGLG
Sbjct: 76  LLLGVALYVKNTRAFYICSWVVYIILNILLISNSIYYREFSDFITVSAMLASSKVSAGLG 135

Query: 121 DSALNLLRIWDLVYVFDFIILIFLFATKKIHLDDRPFNKRASFSITALSGLLFSINLFLA 180
           DSALNLLRIWD++Y+ DFIILI L   KKI  D RPFNKRA+F+ITALS LL SINLFLA
Sbjct: 136 DSALNLLRIWDIIYILDFIILISLSIAKKIKNDQRPFNKRAAFAITALSSLLLSINLFLA 195
```

```
Query: 181 EIDRPELLSRGFSNTYIVKALGLPSFSIYSGNQTYQAQKERNGATAQELATAKKYVAEHY 240
           EIDRPELL+RGFSNTYIV+ALGLP+F++YSGNQTYQAQKERNGATA+EL   K YV  HY
Sbjct: 196 EIDRPELLTRGFSNTYIVRALGLPAFTLYSGNQTYQAQKERNGATAEELIDVKTYVKGHY 255

Query: 241 AKPNPEYYGIGKGRNVIMIHLESFQQFLIDYKLNIDGKEHVVTPFINSLYHSKETVSFSN 300
           A P+P+Y+GIGKG+N+I++HLESFQQFLIDYKL    KE+ VTPFINSLYHS  T++F N
Sbjct: 256 AAPDPQYFGIGKGKNIIVLHLESFQQFLIDYKLKEGDKEYEVTPFINSLYHSNATLAFPN 315

Query: 301 FFHQVKAGKTSDAETLMENSLFGLSSGSFMVNYGGENTQFAAPHILAQNGGYSSAVFHGN 360
           FFHQVKAGKTSDAET+MENSLFGL+SGSFMVNYGGENTQFA P ILAQ GGY+SAVFHGN
Sbjct: 316 FFHQVKAGKTSDAETMMENSLFGLNSGSFMVNYGGENTQFATPSILAQKGGYTSAVFHGN 375

Query: 361 VGTFWNRNNAYKQWGYDYFFDSSYFSKQTKDNSFQYGLNDKYMFADSIKYLEHMQQPFYT 420
           VGTFWNRNNAYKQWGY+YFFDSSYFSKQ   NSFQYGLNDKYMF DSIKYLE MQQPFYT
Sbjct: 376 VGTFWNRNNAYKQWGYNYFFDSSYFSKQNSKNSFQYGLNDKYMFKDSIKYLEQMQQPFYT 435

Query: 421 KFITVSNHYPYTSLKGESDEEGFPLAKTNDETINGYFATANYLDTALKSFFEYLKAAGVY 480
           KFITVSNHYPYTSLKGES EEGFPLAKT+DETINGYFATANYLD ALKSFF+YLKA G+Y
Sbjct: 436 KFITVSNHYPYTSLKGESSEEGFPLAKTDDETINGYFATANYLDAALKSFFDYLKATGLY 495

Query: 481 DNSIIVMYGDHYGISNTRNPSLAELLGKDPETWSEYDNAMLQRVPYMIHIPGYSKGFISN 540
           DNSI V+YGDHYGISN+RN SLA LLGKD ETWSEYDNAMLQRVPYMIHIPGY+ G I
Sbjct: 496 DNSIFVLYGDHYGISNSRNSSLAPLLGKDSETWSEYDNAMLQRVPYMIHIPGYTNGSIKE 555

Query: 541 TYGGEVDNLPTLLHILGIDTSKYTQLGQDLLSKDNKQMVAMRTTGQYITPKYTNYSGHLY 600
           T+GGE+D LPTLLHILGIDTS++ QLGQDLLS  N Q+VA RT+G Y+TP+YTNYSG LY
Sbjct: 556 TFGGEIDALPTLLHILGIDTSQFVQLGQDLLSPQNSQIVAQRTSGTYMTPEYTNYSGRLY 615

Query: 601 YTDSGQEITNPDETTKAEIKAIRDATNKQLSTSDSIQTGDLLRFDENNGLKTVEVEKFNY 660
            T +G EITNPDE T A+ K IR A  +QL+ SD+IQTGDLLRFD  NGLK ++  +F Y
Sbjct: 616 NTQTGLEITNPDEMTIAKTKEIRSAVAQQLAASDAIQTGDLLRFDTQNGLKAIDPNQFIY 675

Query: 661 THSLKALKAKERKLKDRSTSIYSKHNNKSTVDLFHAPSYLELQDPNKTHKTSK 713
           T  LK LK    KL   STS+YSK+ +KST  LF APSYLEL        TS+
Sbjct: 676 TKQLKQLKDISAKLGSESTSLYSKNGHKSTQKLFKAPSYLELNPVEADAATSE 728
```

A related GBS gene <SEQ ID 8619> and protein <SEQ ID 8620> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 9
McG: Discrim Score:     12.63
GvH: Signal Score (-7.5): -2.99
     Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 5 value:  -6.85 threshold:  0.0
   INTEGRAL    Likelihood = -6.85   Transmembrane  129 - 145 ( 126 - 152)
   INTEGRAL    Likelihood = -4.88   Transmembrane   48 -  64 (  46 -  69)
   INTEGRAL    Likelihood = -4.83   Transmembrane   75 -  91 (  74 -  97)
   INTEGRAL    Likelihood = -4.62   Transmembrane   16 -  32 (  15 -  34)
   INTEGRAL    Likelihood = -2.28   Transmembrane  163 - 179 ( 163 - 182)
   PERIPHERAL  Likelihood =  3.76     103
modified ALOM score:   1.87

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3739(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:

```
                    45.2/63.1% over 643aa
                                                          Bacillus subtilis
EGAD|107893| hypothetical protein Insert characterized
GP|2116767|dbj|BAA20118.1||D86418 YfnI Insert characterized
```

GP|2633039|emb|CAB12545.1||Z99107 alternate gene name: yetP~similar to hypothetical proteins Insert characterized
PIR|D69815|D69815 conserved hypothetical protein yfnI - Insert characterized

ORF00125(286 - 2280 of 2742)
EGAD|1078||| S0726(3 - 646 of 653) hypothetical protein { acillus subtilis} GP|2116767|dbj|AA20118.1||D86418 YfnI { acillus subtilis} GP|2633039|emb|CA 12545.1||Z99107 alternate gene name: yetP~similar to hypothetical proteins { acillus subtilis} PIR|D69815|D69815 conserved hypothetical protein yfnI - acillus subtilis
%Match = 28.5
%Identity = 45.1 %Similarity = 63.1
Matches = 297 Mismatches = 227 Conservative Sub.s = 118

```
36        66        96        126       156       186       216       246
FVVKDRPSLRIDLTVKKVEPTG*LNWYQNLFFPVTEHLI*FFFQRQNSL*VYS*TVL*QIFIFFHTEFDLSLPYVTKFYV

276       306       336       366       396       426       456       486
II*SEILSLGKKLKEVPTVKKIKDFASRAINTRLGFILLLVVIYWLKTIWAYHTDFNLGLENSYQLFLTIINPIPLGLLI
             :|:   ||::    :  : :| | :| |:::| ||  :| |:||||:: : |  | | ||    ::
           MNEELKVFKKVEVAMKKLFSYKLSFFVLAVILFWAKTYLSYKTEFNLGVKGTTQEILLIFNPFSSAVFF
                    10        20        30        40        50        60

516       546       576       606       636       666       696       726
IGLALYVKRTKAFYITAFITYAIVNILLIANAIYYREFSDFITVSAVLASSKTSAGLGDSALNLLRIWDLVYVFDFIILI
:|||| | |: | | : ::  :| || ::|| | ||:|  :  |    :||  ::::   |: | :| ||||
LGLALLAKGRKSAIIMLIIDF-LMTFVLYANILFYRFFDDFLTFPNI-KQSGNVGNMGDGIFSIMAGHDIFYFLDIIILI
          80        90        100       110       120       130       140

756       786                 843       873       903       933       963
FLFATKKIHLDDRPFNKRASFSITALSGL-LFSINLFLAEIDRPELLSRGFSNTYIVKALGLPSFSIYSGNQTYQAQKER
      :  | :    ||   |: |||: || |||  || |||:||:| |   |||| ||| :::|| | || | : :|
-AVLIWRPELKEYKMKKR-FASLVILSGIALFFINLHYAEKDRPQLLTRTFDRNYIVKYLGLYNYTIYDGVQTAQTETQR
            160       170       180       190       200       210       220

993      1023      1053      1083      1113      1143      1173      1200
NGATAQELATAKKYVAEHYAKPNPEYYGIGKGRNVIMIHLESFQQFLIDYKLNIDGKEHVVTPFINSLYHSKETVS-FSN
  |:: :| : : |   |||||| ||:|  ||:|:| |||||||| ||||||||       ||||:| | |  | |: | |
AYASSDDLTSVENYTTSHYAKPNAEYFGSAKGKNIIKIHLESFQSFLIDYKLN----GEEVTPFLNKLAHGGEDVTYFDN
             240       250       260       270           280       290       300

1230      1260      1290      1320      1350      1380      1410      1440
FFHQVKAGKTSDAETLMENSLFGLSSGSFMVNYGGENTQFAAPHILAQNGGYSSAVFHGNVGTFWNRNNAYKQWGYDYFF
||||  |||||||  |:||:|||  ||  |   |||| : | || |  ||:|||:||:  :||||:  ||: ||| ||
FFHQTGQGKTSDAELTMDNSIFGLPEGSAFVT-KGENTYQSLPAILDQKEGYTSAVLHGDYKSFWNRDQIYKHIGYDKFF
                320       330       340       350       360       370       380

1470      1500      1530      1560      1590      1620      1650      1680
DSSYFSKQTKDNSFQYGLNDKYMFADSIKYLEHMQQPFYTKFITVSNHYPYTSLKGESDEEGFPLAKTNDETINGYFATA
|:|    : :|   || ||  | :||  || ::|||| :||::||||: :|  | |   :  | | | |:: || ||
DAS-TYDMSDENVINMGLKDKPFFTESIPKLESLKQPFYAHLITLTNHYPF-NLD-EKDAS-LKKATTGDNTVDSYFQTA
       390       400       410       420       430           440       450

1710      1740      1770      1800      1830      1860      1890      1920
NYLDTALKSFFEYLKAAGVYDNSIIVMYGDHYGISNTRNPSLAELLGKDPETWSEYDNAMLQRVPYMIHIPGYSKGFISN
 ||| ||: ||: || ||:||||:|::|||| |||   | :: |:|||     ::| ||  |||| || :||  || :::
RYLDEALEQFFKELKEAGLYDNSVIMIYGDHNGISENHNRAMKEILGK---EITDYQNAQNQRVPLMIRVPG-KKGGVNH
            470       480       490       500           510       520       530

1950      1980      2010      2040      2070      2100      2130      2160
TYGGEVDNLPTLLHILGIDTSKYTQLGQDLLSKDNKQMVAMRTTGQYITPKYTNYSGHLYYTDSGQEITNPDETTKAEIK
|||||:| :|||||: |||: ||  :| ||:|||:  || |  | ::|||||:    :| | :|:::        | |
TYGGEIDVMPTLLHLEGIDSQKYINFGTDLFSKDHDDTVAFR-NGDFVTPKYTSVDNIIYDTKTGEKL-----KANEETK
                550       560       570       580       590           600

2190      2220      2250      2280      2310      2340      2370      2400
AIRDATNKQLSTSDSIQTGDLLRFDENNGLKTVEVEKFNYTHSLKALKAKERKLKDRSTSIYSKHNNKSTVDLFHAPSYL
 ::   |:||| |||:   ||||| : | :| |:   ::|
NLKTRVNQQLSLSDSVLYKDLLRFHKLNDFKAVDPSDYHYGKEKEIK
```

```
620        630        640        650
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 613

A DNA sequence (GBSx0653) was identified in *S.agalactiae* <SEQ ID 1903> which encodes the amino acid sequence <SEQ ID 1904>. This protein is predicted to be 50S ribosomal protein L20 (rplT). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3392(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9387> which encodes amino acid sequence <SEQ ID 9388> was also identified.

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

```
>GP:CAB14845 GB:Z99118 ribosomal protein L20 [Bacillus subtilis]
 Identities = 70/89 (78%), Positives = 78/89 (86%)

Query: 1   MFRTAKEQVMNSYYYAYRDRRQKKRDFRKLWITRINAAARMNGLSYSQLMHGLKLAEIEV 60
           +++ A +QVM S  YA+RDRRQKKRDFRKLWITRINAAARMNGLSYS+LMHGLKL+ IEV
Sbjct: 31  LYKVANQQVMKSGNYAFRDRRQKKRDFRKLWITRINAAARMNGLSYSRLMHGLKLSGIEV 90

Query: 61  NRKMLADLAVNDAAAFTALADAAKAKLGK 89
           NRKMLADLAVND  AF  LADAAKA+L K
Sbjct: 91  NRKMLADLAVNDLTAFNQLADAAKAQLNK 119
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1905> which encodes the amino acid sequence <SEQ ID 1906>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.06   Transmembrane   94 - 110 (  94 - 110)

----- 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>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 87/89 (97%), Positives = 88/89 (98%)

Query: 1   MFRTAKEQVMNSYYYAYRDRRQKKRDFRKLWITRINAAARMNGLSYSQLMHGLKLAEIEV 60
           +FRTAKEQVMNSYYYAYRDRRQKKRDFRKLWITRINAAARMNGLSYSQLMHGLKLAEIEV
Sbjct: 31  LFRTAKEQVMNSYYYAYRDRRQKKRDFRKLWITRINAAARMNGLSYSQLMHGLKLAEIEV 90

Query: 61  NRKMLADLAVNDAAAFTALADAAKAKLGK 89
           NRKMLADLAV DAAAFTALADAAKAKLGK
Sbjct: 91  NRKMLADLAVADAAAFTALADAAKAKLGK 119
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 614

A DNA sequence (GBSx0654) was identified in *S.agalactiae* <SEQ ID 1907> which encodes the amino acid sequence <SEQ ID 1908>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.64   Transmembrane   32 -  48 (  32 -  48)
   INTEGRAL    Likelihood = -0.32   Transmembrane    3 -  19 (   3 -  19)

----- Final Results -----
               bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 615

A DNA sequence (GBSx0655) was identified in *S.agalactiae* <SEQ ID 1909> which encodes the amino acid sequence <SEQ ID 1910>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.63   Transmembrane  747 - 763 ( 743 - 772)
   INTEGRAL    Likelihood =-12.52   Transmembrane  840 - 856 ( 835 - 856)
   INTEGRAL    Likelihood =-11.20   Transmembrane  447 - 463 ( 440 - 466)
   INTEGRAL    Likelihood = -5.79   Transmembrane  351 - 367 ( 346 - 372)
   INTEGRAL    Likelihood = -4.25   Transmembrane  517 - 533 ( 516 - 537)
   INTEGRAL    Likelihood = -1.49   Transmembrane  397 - 413 ( 396 - 413)
   INTEGRAL    Likelihood = -0.96   Transmembrane  799 - 815 ( 799 - 817)

----- Final Results -----
               bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9349> which encodes amino acid sequence <SEQ ID 9350> was also identified.

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

```
>GP:AAB89436 GB:AE000977 A. fulgidus predicted coding region AF1820
          [Archaeoglobus fulgidus]
 Identities = 100/483 (20%), Positives = 210/483 (42%), Gaps = 61/483 (12%)

Query: 351 LFPIILYLVAALVTLTTMTRFVEEERTNAGILKALGYSDRQVIFKFIIYGFIAGTLGTTL 410
           LFP    LV+  +T   ++R    +  N  +++ALG++  +++  ++ Y  + G   +T
Sbjct: 276 LFPAFFILVSIFMTYALLSRIFRLQLGNIAVMRALGFTRNEIMLHYLQYPLLMGFFASTA 335

Query: 411 GIIGGHYLLPRIISDIISKDLTIPNTQYHLFLNYSLLAFVFSLLSIVLPVFVI------- 463
           G++ G +    + S  I+  L +P       L    L+ +   L+  +  F++
Sbjct: 336 GLVAGFFASQLLTSQYIT-FLNLPYYVSKPHLEVYSLSLMAGTLTPTISGFLVAYQASRV 394

Query: 464 ----TRRELKEKAAFLLLPKPPAKGSKIALEYINWIWKKLSFTQKVTARNIFRYKQRMIM 519
                R   E AA   + +  A  S+I      W   ++     ++ RNIFR K+R  +
Sbjct: 395 DIVKALRGYAEVAAVSFIARIDALFSRI------W---RMRLIFRLALRNIFRSKRRTAI 445

Query: 520 TIFGVAGSVALLFSGLGIQSSLKQTVNEHFGRIMPYDILLTYNTNASPPKILELLSKDSK 579
```

```
           +IF +    +L+ + +    S    +   FG++  YDI ++           E+L K  K
Sbjct: 446 SIFSIVACTSLILNSMVFVDSFDYVMQLQFGKVYAYDIKVSLEGYDGK----EVLEKVRK 501

Query: 580 IDKY--------QPIHLENLDESIPGQINKQSISLFITDKKQLLPFIYLQEATTNKSLHL 631
           +D           PI++E   E++P        +L I    Q L  +Y  E        +
Sbjct: 502 MDGVLFAEPAVEMPIYVEKGGEAVP--------TLLIASNFQTLYNVYNAEG----EKLI 549

Query: 632 NNKGIIISKKLAQFYHVNTGDFIHL------SHSQTLPSRKLKITGVVNANVGHYIFMTK 685
            ++GII SK   +   G+ + +         ++      + +  V  A++
Sbjct: 550 PSEGIIFSKTAMKNLSLVEGEKVSVYTEFGKLEAEVEDVEMIPLLSVATASL-------- 601

Query: 686 QYYRTIFKKEAKDNAFLVKLTKHKIANNLAEKLLEINGVESLTQNALQLASVEAVVRSLD 745
            Y+  I   +   N  +V   + +IA  +AEK+ +++GV+ ++       S+E ++
Sbjct: 602 DYFSRISGVDG-FNRIVVDADEGRIA-EIAEKIRQMDGVKKVSTVIEAQESIEELMGFFY 659

Query: 746 GSMTILVVVSLLLAIVILYNLTNINLAERKRELSTIKVLGFYNEEVTLYIYRETIILSTI 805
             +   +  L    ++N T+I++ ER REL+T+++LG+ + E+ + +  E + ++ +
Sbjct: 660 AFIAFSLFFGVSLGFAAVFNTTSISVIERSRELATLRMLGYTSREIIISLILENLFVAIL 719

Query: 806 GVI 808
           G++
Sbjct: 720 GLV 722
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1911> which encodes the amino acid sequence <SEQ ID 1912>. Analysis of this protein sequence reveals the following:

```
     Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.33   Transmembrane  749 - 765 ( 739 - 775)
   INTEGRAL    Likelihood =-10.88   Transmembrane  845 - 861 ( 834 - 865)
   INTEGRAL    Likelihood = -6.64   Transmembrane  350 - 366 ( 344 - 369)
   INTEGRAL    Likelihood = -6.53   Transmembrane   22 -  38 (  19 -  42)
   INTEGRAL    Likelihood = -6.32   Transmembrane  520 - 536 ( 515 - 537)
   INTEGRAL    Likelihood = -4.99   Transmembrane  446 - 462 ( 445 - 465)
   INTEGRAL    Likelihood = -2.92   Transmembrane  396 - 412 ( 395 - 413)
   INTEGRAL    Likelihood = -0.80   Transmembrane  800 - 816 ( 800 - 819)

----- Final Results -----
              bacterial membrane --- Certainty=0.6731(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:

```
>GP:AAB89436 GB:AE000977 A. fulgidus predicted coding region AF1820
           [Archaeoglobus fulgidus]
 Identities = 101/542 (18%), Positives = 237/542 (43%), Gaps = 42/542 (7%)

Query: 350 IFPVVLYLVAALVAFTTMTRYVDEERTSSGLLKAIGYSNKDISLKFLIYGLLASFLGTTL 409
           +FP    LV+  + +  ++R    +  +  +++A+G++  +I L +L Y LL  F  +T
Sbjct: 276 LFPAFFILVSIFMTYALLSRIFRLQLGNIAVMRALGFTRNEIMLHYLQYPLLMGFFASTA 335

Query: 410 GIIGGTYLLSTLISEILTGA---LTIGKTHLYSYWFYNGIAYLLAMLSAVLPAYLIVKKE 466
           G++ G +   L S+ +T       + K HL  Y       L   +S  L AY   + +
Sbjct: 336 GLVAGFFASQLLTSQYITFLNLPYYVSKPHLEVYSLSLMAGTLTPTISGFLVAYQASRVD 395

Query: 467 LFLN-------AAQLLLPKPPSKGAKIWLEHLTFVWKALSFTHKVTIRNIFRYKQRMLMT 519
           +          AA   + +  ++IW   L F          ++ +RNIFR K+R  ++
Sbjct: 396 IVKALRGYAEVAAVSFIARIDALFSRIWRMRLIF---------RLALRNIFRSKRRTAIS 446

Query: 520 IVGVAGSVALLFAGLGIQSSLAKVVEHQFGDLTTYDILAVGSAKATATEQTDLASYLKQE 579
           I  +   +L+   +    S   V++ QFG +  YDI             +  L  Y  +E
Sbjct: 447 IFSIVACTSLILNSMVFVDSFDYVMQLQFGKVYAYDI-----------KVSLEGYDGKE 494

Query: 580 PITGYQKVSYASLTLPVKGLP---DKQSISILSSS-ATSLSPYFNLLDSQEQKKVPIPTS 635
            +   +K+      P   +P   +K   ++ +   A++    +N+ +++ +K   IP+
Sbjct: 495 VLEKVRKMDGVLFAEPAVEMPIYVEKGGEAVPTLLIASNFQTLYNVYNAEGEKLIPSEG 552
```

```
Query: 636 GVLISEKLASYYKVKPGDQLVLTDRKGQSYKVTIKQVIDMTVGHYLIMSDTYFKNHFKGL 695
           G++ S+       +  G+++ +    G+     ++  ++      L+   T   ++F  +
Sbjct: 553 GIIFSKTAMKNLSLVEGEKVSVYTEFGK-----LEAEVEDVEMIPLLSVATASLDYFSRI 607

Query: 696 EAAPAYLIKVKDKDSKHIKETASDLLTLKAIRAVSQNVNHIKSVQLVVTSLNQVMTLLVF 755
                +  V D D   I E A  +  +  ++ VS  +   +S++ ++       +   +F
Sbjct: 608 SGVDGFNRIVVDADEGRIAEIAEKIRQMDGVKKVSTVIEAQESIEELMGFFYAFIAFSLF 667

Query: 756 LSILLAIVILYNLTTINIAERIRELSTIKVLGFYDQEVTLYIYRETISLSLVGILLGIYL 815
             + L    ++N T+I++ ER REL+T+++LG+  +E+ + +  E + ++++G++  + +
Sbjct: 668 FGVSLGFAAVFNTTSISVIERSRELATLRMLGYTSREIIISLILENLFVAILGLVFALPI 727

Query: 816 GKGLHTYIMTMISTGDIQFGVKVDAYVYLVPILVILSLLAVLGIWVNRHLKKVDMLEALK 875
                +  +      + +    +  +L + +++ +  +   R + ++D+ +  K
Sbjct: 728 AYSTAYFFFSSFESELYYMPMVIYPRTFAATVLAVFAIILLALLPSARRVSEMDIAKVTK 787

Query: 876 SI 877
           I
Sbjct: 788 EI 789
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 377/857 (43%), Positives = 543/857 (62%), Gaps = 7/857 (0%)

Query: 3   KTFWKDIYRSITTSKGRFSSILLLMMLGSFAFIGLKVSAPNMQRTAQNYLAHHHVMDITV 62
           KT WKDI R+I  SKGRF S+  LM LGSFA +GLKV+ P+M+RTA  YL  H VMD+TV
Sbjct: 4   KTLWKDILRAIKNSKGRFISLFFLMALGSFALVGLKVTGPDMERTASRYLERHQVMDLTV 63

Query: 63  FNSWGLDKHDQTVLESLKGSQVEFSYFVDTTPQQNSKSYRLYSNTKTISTFDLVKGRLPL 122
             S    + D+  L++LKG+ +E+ + +D +   N KS RLYS  K +S   LVKG  P
Sbjct: 64  LASHQFSQADKQELDTLKGAHLEYGHLLDVSLTSNQKSLRLYSVPKKVSKPVLVKGSWPK 123

Query: 123 NKSEIALSFQERKKYAIGDKINFKQDKNKLFSNTGPLTIVGFVNSTEIWSKTNLGSSQTG 182
            ++++ LS    K Y IGD++        L + T    +VGF NS+E+WSK+NLGSS TG
Sbjct: 124 RETDLVLSSSLAKNYQIGDELAVTSPMEGLLTTTH-FQVVGFANSSEVWSKSNLGSSSTG 182

Query: 183 DGDLDSYGVLDKTAFHSPVYTMARVTFKDLRLINPFSISYKEKVAKYQEKVSRKLNIHNK 242
           DG L +Y  ++   F S  + + R+ F  LRL N FS  Y+++V + Q  +   L  + +
Sbjct: 183 DGSLYAYAFVNPNVFKS-AFNLLRIRFSHLRLTNAFSKDYQKRVTQNQAHLDNLLKDNGQ 241

Query: 243 IRYTKTKKESLRKIDEEEKSLLKAQKQINRLDNDSLAMPLSQRQAIQMKIKQDRLSLLKR 302
            RY   + +    +L K   ++  + +   S  Q  + +I+Q + +L K
Sbjct: 242 KRYDDLQNQYDLALKNGRAALAKETVKLAASEENLTFLEGSALQEAKHQIEQGKQALAKE 301

Query: 303 TKELLKLRHNTQIMESPQIIVYNRTTFPGGQGYNTFDSSTNSTSKISNLFPIILYLVAAL 362
            K+L +++     +E P  + YNR+T PGG+GY+T+ +ST S S + N+FP++LYLVAAL
Sbjct: 302 EKQLEQVQATKDKLEKPSYLTYNRSTLPGGEGYHTYATSTTSISNVGNIFPVVLYLVAAL 361

Query: 363 VTLTTMTRFVEEERTNAGILKALGYSDRQVIFKFIIYGFIAGTLGTTLGIIGGHYLLPRI 422
           V  TTMTR+V+EERT++G+LKA+GYS++ +  KF+IYG +A  LGTTLGIIGG YLL  +
Sbjct: 362 VAFTTMTRYVDEERTSSGLLKAIGYSNKDISLKFLIYGLLASFLGTTLGIIGGTYLLSTL 421

Query: 423 ISDIISKDLTIPNTQYHLFLNYSLLAFVFSLLSIVLPVFVITRRELKEKAAFLLLPKPPA 482
           IS+I++  LTI  T  + +  Y+ +A++ ++LS VLP ++I  ++EL   AA LLLPKPP+
Sbjct: 422 ISEILTGALTIGKTHLYSYWFYNGIAYLLAMLSAVLPAYLIVKKELFLNAAQLLLPKPPS 481

Query: 483 KGSKIALEYINWIWKKLSFTQKVTARNIFRYKQRMIMTIFGVAGSVALLFSGLGIQSSLK 542
           KG+KI LE++ ++WK LSFT KVT RNIFRYKQRM+MTI GVAGSVALLF+GLGIQSSL
Sbjct: 482 KGAKIWLEHLTFVWKALSFTHKVTIRNIFRYKQRMLMTIVGVAGSVALLFAGLGIQSSLA 541

Query: 543 QTVNEHFGRIMPYDILLTYNTNASPPKILELLS--KDSKIDKYQPIHLENLDESIPGQIN 600
           + V   FG +  YDIL   + A+  + +L S  K   I  YQ +   +L   + G  +
Sbjct: 542 KVVEHQFGDLTTYDILAVGSAKATATEQTDLASYLKQEPITGYQKVSYASLTLPVKGLPD 601

Query: 601 KQSISLFITDKKQLLPFIYLQEATTNKSLHLNNKGIIISKKLAQFYHVNTGDFIHLSHSQ 660
           KQSIS+  +    L P+  L ++   K + +  G++IS+KLA +Y V  GD + L+  +
Sbjct: 602 KQSISILSSSATSLSPYFNLLDSQEQKKVPIPTSGVLISEKLASYYKVKPGDQLVLTDRK 661
```

```
Query: 661 TLPSRKLKITGVVNANVGHYIFMTKQYYRTIFKKEAKDNAFLVKL--TKHKIANNLAEKL 718
           S K+ I  V++  VGHY+ M+  Y++  FK      A+L+K+     K     A  L
Sbjct: 662 G-QSYKVTIKQVIDMTVGHYLIMSDTYFKNHFKGLEAAPAYLIKVKDKDSKHIKETASDL 720

Query: 719 LEINGVESLTQNALQLASVEAVVRSLDGSMTILVVVSLLLAIVILYNLTNINLAERKREL 778
           L +  + +++QN    + SV+ VV SL+  MT+LV +S+LLAIVILYNLT IN+AER REL
Sbjct: 721 LTLKAIRAVSQNVNHIKSVQLVVTSLNQVMTLLVFLSILLAIVILYNLTTINIAERIREL 780

Query: 779 STIKVLGFYNEEVTLYIYRETIILSTIGVILGTISGTYLHRQMMLLIGSDQILFGEKVSP 838
           STIKVLGFY++EVTLYIYRETI LS +G++LG   G  LH  +M +I +  I FG KV
Sbjct: 781 STIKVLGFYDQEVTLYIYRETISLSLVGILLGIYLGKGLHTYIMTMISTGDIQFGVKVDA 840

Query: 839 TTFIIPISVVVIILXXL 855
             +++PI V++ +L  L
Sbjct: 841 YVYLVPILVILSLLAVL 857
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 616

A DNA sequence (GBSx0656) was identified in *S.agalactiae* <SEQ ID 1913> which encodes the amino acid sequence <SEQ ID 1914>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2757(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 GENPEPT database:

```
>GP:AAB89431 GB:AE000977 ABC transporter, ATP-binding protein
          [Archaeoglobus fulgidus]
 Identities = 112/230 (48%), Positives = 167/230 (71%)

Query: 4   IEMKHSYKRYQTGETEIVANNDISFSIERGELVVILGASGAGKSTVLNILGGMDSNSEGE 63
           + ++  +K YQ G+ E+ A    I+  IERGE +V+LG SG GK+T+LNI+GG+D  + G
Sbjct: 2   LRLEDVWKVYQMGKVEVSALRGINLEIERGEFMVVLGPSGCGKTTMLNIIGGIDRPTRGR 61

Query: 64  VLIDGKNIANYTIRELTRYRRYDVGFVFQFYNLVPNLTALENVELASEIVPKALDAQQAL 123
           V+ DGK+I NY     LT +RR +VGF+FQF+NL+P LTA ENVE+A+++V     D  + L
Sbjct: 62  VIFDGKDITNYNEDRLTMHRRNNVGFIFQFFNLIPTLTARENVEIAADLVESPRDVDEVL 121

Query: 124 ENVGLGHRINHFPAQLSGGEQQRVAIARAIAKKPKLLLCDEPTGALDYQTGKQVLAILQK 183
           + VGL  R  HFPA+LSGGEQQRVAIARA+ K P ++L DEPTG+LD++TGK VL ++++
Sbjct: 122 KMVGLADRAEHFPAELSGGEQQRVAIARALVKNPPIILADEPTGSLDFETGKAVLKVMRE 181

Query: 184 MAQSKETTVIIVTHNTALAPIANRVIHMHDSKISDIVINENPSDIQNIEY 233
           + + +  T ++VTHN+A+A IA+RV+++ D K+  + N +P+D   I++
Sbjct: 182 INRKEGITFVLVTHNSAIAAIADRVVYLRDGKVERVERNLHPADPDEIQW 231
```

There is also homology to SEQ ID 1354.

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

Example 617

A DNA sequence (GBSx0657) was identified in *S.agalactiae* <SEQ ID 1915> which encodes the amino acid sequence <SEQ ID 1916>. This protein is predicted to be DNA topoisomerase I (topA). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4716(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9821> which encodes amino acid sequence <SEQ ID 9822> was also identified.

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

```
>GP:CAB13485 GB:Z99112 DNA topoisomerase I [Bacillus subtilis]
 Identities = 442/690 (64%), Positives = 535/690 (77%), Gaps = 10/690 (1%)

Query: 27  LVIVESPAKAKTIEKYLGRNYKVVASVGHIRDLKKSSMSIDFENNYEPQYINIRGKGPLI 86
           LVIVESPAKAKTIE+YLG+ YKV AS+GH+RDL KS M +D E N+EP+YI IRGKGP++
Sbjct: 5   LVIVESPAKAKTIERYLGKKYKVKASMGHVRDLPKSQMGVDIEQNFEPKYITIRGKGPVL 64

Query: 87  NDLKKEAKKAKKVYLASDPDREGEAISWHLAHILDLDKEDRNRVVFNEITKDAVKNAFVE 146
            +LK  AKKAKKVYLA+DPDREGEAI+WHLAH LDLD      RVVFNEITKDA+K +F
Sbjct: 65  KELKTAAKKAKKVYLAADPDREGEAIAWHLAHSLDLDLNSDCRVVFNEITKDAIKESFKH 124

Query: 147 PRQINMDLVDAQQARRVLDRIVGYSISPILWKKVKKGLSAGRVQSVALKLIIDRENEIKA 206
           PR INMDLVDAQQARR+LDR+VGY ISPILWKKVKKGLSAGRVQSVAL+LIIDRE EI
Sbjct: 125 PRMINMDLVDAQQARRILDRLVGYKISPILWKKVKKGLSAGRVQSVALRLIIDREKEIND 184

Query: 207 FQPEEYWTIDGSFKKGTRKFNATFYGLDGKKFKLSNNEDVKTVLKRIKTDEFLVEKVEKK 266
           F+PEEYWTIDG+F KG   F A+F+G +GKK  L++  DVK +L ++K +++ VEKV KK
Sbjct: 185 FKPEEYWTIDGTFLKGQETFEASFFGKNGKKLPLNSEADVKEILSQLKGNQYTVEKVTKK 244

Query: 267 ERRRNAPLPYTTSSLQQDAANKINFRTRKTMMIAQQLYEGLSLGTAGHQGLITYMRTDST 326
           ER+RN  LP+TTS+LQQ+AA K+NFR +KTMMIAQQLYEG+ LG  G  GLITYMRTDST
Sbjct: 245 ERKRNPALPFTTSTLQQEAARKLNFRAKKTMMIAQQLYEGIDLGREGTVGLITYMRTDST 304

Query: 327 RISPLAQNEATEFITNRFGANYSKHGNK-VKNASGAQDAHEAIRPSSVNHTPESIAKYLD 385
           RIS  A +EA  FI   +G  +     K  K    AQDAHEAIRP+SV   P  +   L
Sbjct: 305 RISNTAVDEAAAFIDQTYGKEFLGGKRKPAKKNENAQDAHEAIRPTSVLRKPSELKAVLG 364

Query: 386 KDQLKLYTLIWNRFIASQMTAAVFDTMKVNLTQNGVTFIANGSQVKFDGYMAVYND---- 441
           +DQ++LY LIW RF+ASQM  AV DTM V+LT NG+TF ANGS+VKF G+M VY +
Sbjct: 365 RDQMRLYKLIWERFVASQMAPAVLDTMSVDLTNNGLTFRANGSKVKFSGFMKVYVEGKDD 424

Query: 442 --TDKNKMLPDMEEGESVKKVNTNPEQHFTQPPARFSEASLIKTLEENGVGRPSTYAPTL 499
              +K++MLPD++EG++V   +  PEQHFTQPP R++EA L+KTLEE G+GRPSTYAPTL
Sbjct: 425 QMEEKDRMLPDLQEGDTVLSKDIEPEQHFTQPPPRYTEARLVKTLEERGIGRPSTYAPTL 484

Query: 500 ETIQKRYYVKLAAKRFEPTELGEIVNSLIVEFFPDIVDVTFTAEMEGKLDEVEIGKEQWQ 559
           +TIQ+R YV L  KRF PTELG+IV  LI+EFFP+I++V FTA+ME  LD VE G  +W
Sbjct: 485 DTIQRRGYVALDNKRFVPTELGQIVLDLIMEFFPEIINVEFTAKMERDLDHVEEGNTEWV 544

Query: 560 KIIDEFYKPFEKELAKAETEMEKIQIKDEPAGFDCELCGSPMVIKLGRYGKFYACSNFPE 619
           KIID FY  FEK + KAE+EM++++I+ E AG DCELC SPMV K+GRYGKF ACSNFP+
Sbjct: 545 KIIDNFYTDFEKRVKKAESEMKEVEIEPEYAGEDCELCSSPMVYKMGRYGKFLACSNFPD 604

Query: 620 CHNTKAITKEIGVICPICQKGQVIERKTKRNRIFYGCDRYPECEFTSWDKPIGRTCPKSN 679
           C NTK I K+IGV CP C +G ++ERK+K+ R+FYGCDRYP+CEF SWDKPI R CPK
Sbjct: 605 CRNTKPIVKQIGVKCPSCGEGNIVERKSKKRVFYGCDRYPDCEFVSWDKPIERKCPKCG 664
```

```
Query: 680 DFLVEKKVRGGGKQVVCSNEKCDYQEEKIK 709
           LVEKK++  G QV C   +CDY+EE  K
Sbjct: 665 KMLVEKKLK-KGIQVQC--VECDYKEEPQK 691
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1917> which encodes the amino acid sequence <SEQ ID 1918>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5445(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 595/704 (84%), Positives = 656/704 (92%), Gaps = 1/704 (0%)

Query: 6   TTTKTSTKKTSKKKSATAKKNLVIVESPAKAKTIEKYLGRNYKVVASVGHIRDLKKSSMS 65
           T  KT TKK++ KK +TAKKNLVIVESPAKAKTIEKYLGR+YKVVASVGHIRDLKKSSMS
Sbjct: 7   TKPKTGTKKSTTKKKSTAKKNLVIVESPAKAKTIEKYLGRSYKVVASVGHIRDLKKSSMS 66

Query: 66  IDFENNYEPQYINIRGKGPLINDLKKEAKKAKKVYLASDPDREGEAISWHLAHILDLDKE 125
           IDF+NNYEPQYINIRGKGPLIN LKKEAK AKKVYLASDPDREGEAISWHL+HIL LD +
Sbjct: 67  IDFDNNYEPQYINIRGKGPLINSLKKEAKAAKKVYLASDPDREGEAISWHLSHILGLDPQ 126

Query: 126 DRNRVVFNEITKDAVKNAFVEPRQINMDLVDAQQARRVLDRIVGYSISPILWKKVKKGLS 185
           D NRVVFNEITKDAVK+AFVEPRQI+MDLVD+QQARRVLDRIVGYSISPILWKKVKKGLS
Sbjct: 127 DNNRVVFNEITKDAVKHAFVEPRQIDMDLVDSQQARRVLDRIVGYSISPILWKKVKKGLS 186

Query: 186 AGRVQSVALKLIIDRENEIKAFQPEEYWTIDGSFKKGTRKFNATFYGLDGKKFKLSNNED 245
           AGRVQSVALKLIIDREN+IKAF P+EYW+IDG FKKGT+KF ATFYG++GKK KL NN D
Sbjct: 187 AGRVQSVALKLIIDRENDIKAFVPKEYWSIDGLFKKGTKKFQATFYGINGKKTKLDNNND 246

Query: 246 VKTVLKRIKTDEFLVEKVEKKERRRNAPLPYTTSSLQQDAANKINFRTRKTMMIAQQLYE 305
           VK VL ++  ++FLV KV+KKERRRNAPLPYTTSSLQQDAANKINFRTRKTMM+AQQLYE
Sbjct: 247 VKEVLAKLTNEDFLVSKVDKKERRRNAPLPYTTSSLQQDAANKINFRTRKTMMVAQQLYE 306

Query: 306 GLSLGTAGHQGLITYMRTDSTRISPLAQNEATEFITNRFGANYSKHGNKVKNASGAQDAH 365
           G+ LG  G QGLITYMRTDSTRISP+AQN+A +FI NRFGANYSKHGN+VKN SG QDAH
Sbjct: 307 GIHLGENGTQGLITYMRTDSTRISPVAQNDAAQFIINRFGANYSKHGNRVKNTSGVQDAH 366

Query: 366 EAIRPSSVNHTPESIAKYLDKDQLKLYTLIWNRFIASQMTAAVFDTMKVNLTQNGVTFIA 425
           EAIRPSSVNHTP+SIAKYL+KDQLKLYTLIWNRF+ASQMTAAVFDT+KVNL QNGV F+A
Sbjct: 367 EAIRPSSVNHTPDSIAKYLNKDQLKLYTLIWNRFVASQMTAAVFDTVKVNLEQNGVIFVA 426

Query: 426 NGSQVKFDGYMAVYNDTDKNKMLPDMEEGESVKKVNTNPEQHFTQPPARFSEASLIKTLE 485
           NGSQ+KFDGYMAVYND+DKNKMLP+M EGE+VKK++T+PEQHFTQPPAR+SEA+LIKTLE
Sbjct: 427 NGSQMKFDGYMAVYNDSDKNKMLPEMAEGETVKKISTSPEQHFTQPPARYSEATLIKTLE 486

Query: 486 ENGVGRPSTYAPTLETIQKRYYVKLAAKRFEPTELGEIVNSLIVEFFPDIVDVTFTAEME 545
           ENGVGRPSTYAPTLE IQ+RYYVKL+AKRFEPTELGEIVN LIVEFFPDIVDV FTAEME
Sbjct: 487 ENGVGRPSTYAPTLEVIQRRYYVKLSAKRFEPTELGEIVNKLIVEFFPDIVDVAFTAEME 546

Query: 546 GKLDEVEIGKEQWQKIIDEFYKPFEKELAKAETEMEKIQIKDEPAGFDCELCGSPMVIKL 605
           GKLD+VEIG+EQWQ +ID+FY+PF KEL KAE+E+EKIQIKDEPAGFDC++CG PMVIKL
Sbjct: 547 GKLDQVEIGEEQWQHVIDQFYQPFVKELNKAESEIEKIQIKDEPAGFDCDVCGHPMVIKL 606

Query: 606 GRYGKFYACSNFPECHNTKAITKEIGVICPICQKGQVIERKTKRNRIFYGCDRYPECEFT 665
           GR+GKFYACSNFPEC NTKAITKEIGV CP+C KGQVIERKTK+NRIFYGCD+YP+CEF
Sbjct: 607 GRFGKFYACSNFPECRNTKAITKEIGVTCPVCHKGQVIERKTKKNRIFYGCDQYPDCEFI 666

Query: 666 SWDKPIGRTCPKSNDFLVEKKVRGGGKQVVCSNEKCDYQEEKIK 709
           SWD PIGR CPKS D+L+EKKVR GGKQV+CSNE CDY+EEKIK
Sbjct: 667 SWDLPIGRACPKSGDYLIEKKVR-GGKQVMCSNETCDYKEEKIK 709
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 618

A DNA sequence (GBSx0658) was identified in *S.agalactiae* <SEQ ID 1919> which encodes the amino acid sequence <SEQ ID 1920>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2578(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 GENPEPT database:

```
>GP:AAD35341 GB:AE001708 DNA processing chain A [Thermotoga maritima]
 Identities = 97/231 (41%), Positives = 149/231 (63%), Gaps = 2/231 (0%)

Query: 51  FIENYKQLDLKKLRQEFKKFPV--LSILDSNYPLELKEIYNPPVLLFYQGNIELLSKPKL 108
           F+E   + +L++ ++  +K  V  +S  + +YP  L+EI  PP +LF +G+ ELL +  +
Sbjct: 41  FLEKCGKEELERQKELIRKHNVKLVSFWEDDYPQHLREIRYPPAVLFVRGDAELLKEKCV 100

Query: 109 AVVGARQASQIGCQSVKKIIKETNNQFVIVSGLARGIDTAAHVSALKNGGSSIAVIGSGL 168
            VVG R+ +  G     K+ +K  +  FVIVSG+A GID+ AH  AL +GG ++AV+G+G+
Sbjct: 101 GVVGTRRPTSYGVNVTKRFVKLLSEYFVIVSGMAFGIDSVAHKEALSSGGKTVAVLGTGV 160

Query: 169 DVYYPTENKKLQEYMSYNHLVLSEYFTGEQPLKFHFPERNRIIAGLCQGIVVAEAKMRSG 228
           DV YP  N++L   +  N  V+SEY  G +  K HFP RNRIIAGL   I+V EA ++SG
Sbjct: 161 DVVYPRSNERLFHEIVKNGCVVSEYPMGTRARKHHFPARNRIIAGLSDAIIVTEAPIKSG 220

Query: 229 SLITCERALEEGREVFAIPGNIIDGKSDGCHHLIQEGAKCIISGKDILSEY 279
           +LIT + ALE GR+VFA+PG+I     S+G ++LI+ GA  +    +D+ + +
Sbjct: 221 ALITVKFALESGRDVFAVPGDIDRKTSEGTNYLIKSGAYPLTDEEDLETHF 271
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1921> which encodes the amino acid sequence <SEQ ID 1922>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2856(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 185/279 (66%), Positives = 238/279 (84%), Gaps = 1/279 (0%)

Query: 1   MNHFELFKLKKAGLTNLNIHNIINYLKKNSLTSLSVRNMAVVSKCKNPTFFIENYKQLDL 60
           +NHFEL+KLKKAGLTN NI NI++Y +K+   SLS+R+MAVVS CK+P+ FIE YKQLD+
Sbjct: 1   VNHFELYKLKKAGLTNKNILNILDY-QKHQEKSLSLRDMAVVSGCKHPSHFIEAYKQLDI 59

Query: 61  KKLRQEFKKFPVLSILDSNYPLELKEIYNPPVLLFYQGNIELLSKPKLAVVGARQASQIG 120
           + L+ EFK+FP +SILD +YP+ LKEIYNPPVLLF+QGN++LL KPKLA+VG+R++S  G
Sbjct: 60  QNLKMEFKQFPSISILDKHYPMALKEIYNPPVLLFFQGNLDLLEKPKLAIVGSRRSSDTG 119

Query: 121 CQSVKKIIKETNNQFVIVSGLARGIDTAAHVSALKNGGSSIAVIGSGLDVYYPTENKKLQ 180
            +SV+KI+KE  N+FVIVSGLARGIDT+AH++ LKNGG +IA+IG+GLD +YP EN++LQ
Sbjct: 120 VKSVRKILKELGNRFVIVSGLARGIDTSAHLACLKNGGQTIAIIGTGLDRFYPKENRELQ 179

Query: 181 EYMSYNHLVLSEYFTGEQPLKFHFPERNRIIAGLCQGIVVAEAKMRSGSLITCERALEEG 240
```

```
            ++  NHLVL+EY  GE+ L +HFPERNRIIAGL +GI+V EAK RSGSLITC+  +EEG
Sbjct: 180 TFLGKNHLVLTEYGPGEEALSYHFPERNRIIAGLSRGILVVEAKNRSGSLITCQIGIEEG 239

Query: 241 REVFAIPGNIIDGKSDGCHHLIQEGAKCIISGKDILSEY 279
           R++FA+PGNI+DGKS+GC  LI+EGA C+ SG DILSEY
Sbjct: 240 RDIFAVPGNILDGKSEGCLQLIKEGATCVTSGMDILSEY 278
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 619

A DNA sequence (GBSx0659) was identified in *S.agalactiae* <SEQ ID 1923> which encodes the amino acid sequence <SEQ ID 1924>. This protein is predicted to be lipoprotein (ceuE). 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>
```

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

```
>GP:CAA06500 GB:AJ005352 lipoprotein [Staphylococcus aureus]
 Identities = 122/348 (35%), Positives = 201/348 (57%), Gaps = 16/348 (4%)

Query: 1   MTKKLIIAILALCTILTTSQAVLAKEKSQ--------TVTIKNNYSVYIKKEKRDKPDNK 52
           M K ++  +LA+  +L         KE+S+         TV I+NNY +  + EK+D  D K
Sbjct: 1   MKKTVLYLVLAVMFLLAACGNNSDKEQSKSETKGSKDTVKIENNYKM--RGEKKDGSDAK 58

Query: 53  KQISETLKVPLKPKKVVVFDMGALDTITALGAEKSVIGIPKAKNALSLLPNNVKSVYKAK 112
           K + ET++VP  P+  VV D GALD +  +G    V  +PK +   SL PN ++S +K
Sbjct: 59  K-VKETVEVPKNPENAVVLDYGALDVMKEMGLSDKVKALPKGEGGKSL-PNFLES-FKDD 115

Query: 113 RYQDVGSLFEPNFEAIARMQPDVVFLGARMASVDNIEKLKEAAPKAALVYAGVDSKKVFD 172
           +Y +VG+L E NF+ IA  +P+V+F+  R A+  N+++ K+AAPKA +VY G D K +
Sbjct: 116 KYTNVGNLKEVNFDKIAATKPEVIFISGRTANQKNLDEFKKAAPKAKIVYVGADEKNLIG 175

Query: 173 KGVAERVTMLGKIFDQNKKAKTFNKDIAQAVLKLQKTIEKKGKPTALFVMANSGELLTQS 232
             + +    +GKI+D+  KAK  NKD+   +  ++   K T ++++ N GEL T
Sbjct: 176 S-MKQNTENIGKIYDKEVKAKELNKDLDNKIASMKDKTKNFNK-TVMYLLVNEGELSTFG 233

Query: 233 PSGRFGW-IFSVGGFKAVNENEKLSSHGTPVSYEYIAEKNPNYLFVLDRGATIGQGASSK 291
           P GRFG  ++   GF AV++    S+HG  VS EY+ ++NP+ +  +DRG  +   +++K
Sbjct: 234 PKGRFGGLVYDTLGFNAVDKKVSNSNHGQNVSNEYVNKENPDVILAMDRGQAVSGKSTAK 293

Query: 292 ELFNNDVIKATDAVKNKRVHEVDGKDWYINSGGSRVTLRMIKDVQNFV 339
           +  NN V+K   A+K  +V+ +D K WY  +G +  T++ I+++   V
Sbjct: 294 QALNNPVLKNVKAIKEDKVYNLDPKLWYFAAGSTTTTIKQIEELDKVV 341
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1925> which encodes the amino acid sequence <SEQ ID 1926>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> 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>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 57/255 (22%), Positives = 104/255 (40%), Gaps = 30/255 (11%)

Query: 66  KKVVVFDMGALDTITALGAEKSVIGIPKAKNALSLLPNNVKSVYKAKRYQDVGSLFEPNF 125
           +++V   +  +D    L  +  ++G+  +K  L  LP    +V +      VG    P+
Sbjct: 45  QRIVATSVAVVDICDRLNLD--LVGVCDSK--LYTLPKRYDAVKR------VGLPMNPDI 94

Query: 126 EAIARMQPDVVFLGARMASVDNIEKLKEAAPKAALVYAGVDSKKVFDKGVAERVTMLGKI 185
           E IA ++P  +     +      E L+    K    Y  ++ + V  +G+ + +  LG +
Sbjct: 95  ELIASLKPTWILSPNSLQ-----EDLEPKYQKLDTEYGFLNLRSV--EGMYQSIDDLGNL 147

Query: 186 FDQNKKAKTFNKDIAQAVLKLQKTIEKKGKPTALFVMANSGELLTQSPSGRFGWIFSVGG 245
           F + ++AK   +          Q   + K KP  L +M   G  L  +      G +  G
Sbjct: 148 FQRQQEAKELRQQYQDYYRAFQAKRKGKKKPKVLILMGLPGSYLVATNQSYVGNLLDLAG 207

Query: 246 FKAV---NENEKLSSHGTPVSYEYIAEKNPNYLFVLDRGATIGQGAS---SKELFNNDVI 299
            + V   +E E LS++      E +  K P+   +L     I          KE   ND+
Sbjct: 208 GENVYQSDEKEFLSANP-----EDMLAKEPD--LILRTAHAIPDKVKVMFDKEFAENDIW 260

Query: 300 KATDAVKNKRVHEVD 314
           K   AVK  +V+++D
Sbjct: 261 KHFTAVKEGKVYDLD 275
```

SEQ ID 1924 (GBS181) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 39 (lane 5; MW 38.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 3; MW 64kDa).

The GBS181-GST fusion product was purified (Figure 204, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 299), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 620

A DNA sequence (GBSx0660) was identified in *S.agalactiae* <SEQ ID 1927> which encodes the amino acid sequence <SEQ ID 1928>. This protein is predicted to be iron(III) ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3231(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 GENPEPT database:

```
>GP:CAB12190 GB:Z99106 similar to ferrichrome ABC transporter
          (ATP-binding protein) [Bacillus subtilis]
 Identities = 125/247 (50%), Positives = 187/247 (75%)

Query: 1   MIQINNLHKFYGQKEILKDINISIPKGKVTAILGPNGSGKSTLLSCISRLEPYDNGEIFL 60
           M+++ N+ K YG K +L++ +++I KGK+T+ +GPNG+GKSTLLS +SRL   D+GEI++
Sbjct: 1   MVEVRNVSKQYGGKVVLEETSVTIQKGKITSFIGPNGAGKSTLLSIMSRLIKKDSGEIYI 60

Query: 61  DKVPLAHYSSNDLAKTLAILRQSNHLTLKIKVRDLIGFGRFPYSKGRLSQKDKAVIESVI 120
           D   +    S +LAK ++IL+Q+N + +++ ++DL+ FGRFPYS+GRL+++D   I   +
Sbjct: 61  DGQEIGACDSKELAKKMSILKQANQINIRLTIKDLVSFGRFPYSQGRLTEEDWVHINQAL 120
```

```
Query: 121 SYMDLNDIADEFINNLSGGQIQRAFIAMTMAQDTQYICLDEPLNNLDMKYAVQMMDLIKR 180
           SYM L DI D++++ LSGGQ QRAFIAM +AQDT YI LDEPLNNLDMK++V++M L+KR
Sbjct: 121 SYMKLEDIQDKYLDQLSGGQCQRAFIAMVIAQDTDYIFLDEPLNNLDMKHSVEIMKLLKR 180

Query: 181 YAYEFNKTIVIIIHDINFATHYADNVVALKEGQVVTCGTVEDVMQEKILSHLFDMPIRIE 240
              E  KTIVI+IHDINFA+ Y+D +VALK G++V  G  E++++  +L  ++DM I I+
Sbjct: 181 LVEELGKTIVIVIHDINFASVYSDYIVALKNGRIVKEGPPEEMIETSVLEEIYDMTIPIQ 240

Query: 241 TVDGKPI 247
           T+D + I
Sbjct: 241 TIDNQRI 247
```

There is also homology to SEQ ID 1930.

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

Example 621

A DNA sequence (GBSx0661) was identified in *S.agalactiae* <SEQ ID 1931> which encodes the amino acid sequence <SEQ ID 1932>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.74   Transmembrane  271 - 287 ( 266 - 295)
   INTEGRAL    Likelihood = -8.55   Transmembrane   49 -  65 (  47 -  69)
   INTEGRAL    Likelihood = -8.07   Transmembrane  185 - 201 ( 178 - 207)
   INTEGRAL    Likelihood = -7.70   Transmembrane  112 - 128 ( 105 - 132)
   INTEGRAL    Likelihood = -7.38   Transmembrane  231 - 247 ( 227 - 261)
   INTEGRAL    Likelihood = -2.50   Transmembrane  139 - 155 ( 135 - 156)
   INTEGRAL    Likelihood = -1.97   Transmembrane  302 - 318 ( 301 - 319)

----- 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>
```

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

```
>GP:CAB12189 GB:Z99106 similar to ferrichrome ABC transporter
          (permease) [Bacillus subtilis]
 Identities = 138/315 (43%), Positives = 222/315 (69%), Gaps = 6/315 (1%)

Query: 9   KLLILLILLIAAIILFLIYGIPTDANEFLIIYILKTRYQKLIALILVGICIGSSSLIFQT 68
           K+ +L+ L I  I LFL Y +          Y L  R +K+ A++L G  I  S++IFQT
Sbjct: 6   KIALLVGLAIVCIGLFLFYDLGNWD------YTLPRRIKKVAAIVLTGGAIAFSTMIFQT 59

Query: 69  LTNNRLLTPSIIGLDSLYILIQTGLMYLIGAQRVIKFSSFSSFLLSLLLMVGFAYLLFTI 128
           +TNNR+LTPSI+GLDSLY+LIQTG+++L G+  ++  +   +F++S+LLM+ F+ +L+ I
Sbjct: 60  ITNNRILTPSILGLDSLYMLIQTGIIFLFGSANMVIMNKNINFIISVLLMILFSLVLYQI 119

Query: 129 LFRNKKQSLYFVLLAGLIFNTLFSSISSFIQAIMDPNDFMILQNQLFASFNAINTKILWI 188
           +F+ + ++++F+LL G++F TLFSS+SSF+Q ++DPN+F ++Q+++FASFN INT +LW+
Sbjct: 120 MFKGEGRNIFFLLLIGIVFGTLFSSLSSFMQMLIDPNEFQVVQDKMFASFNNINTDLLWL 179

Query: 189 SFIIIVVSFVINWPFIKELDVLLLGKENAISLGISYQKLITRFFLWLALMVAIATALVGP 248
           +FII +++ V  W F K  DVL LG+E+A++LGI Y K+  +  + +A++V+++TALVGP
Sbjct: 180 AFIIFLLLTGVYVWRFTKFFDVLSLGREHAVNLGIDYDKVVKQMLIVVAILVSVSTALVGP 239

Query: 249 ITFLGLLVAHITYHSFHTFRHQILVPIAIVICIFTLVLGQHLVQNLLHLTVQLSVLLNLI 308
           I FLGLLV ++       T++H  L+  ++ I I  LV GQ +V+ +   +  LSV++N
Sbjct: 240 IMFLGLLVVNLAREFLKTYKHSYLIAGSVFISIIALVGGQFVVEKVFTFSTTLSVIINFA 299

Query: 309 GGSYFIFTLIKGRKN 323
           GG YFI+ L+K  K+
```

```
Sbjct: 300 GGIYFIYLLLKENKS 314
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1933> which encodes the amino acid sequence <SEQ ID 1934>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.64   Transmembrane   33 -  49 (  26 -  61)
   INTEGRAL    Likelihood = -8.97   Transmembrane  259 - 275 ( 246 - 286)
   INTEGRAL    Likelihood = -8.65   Transmembrane  296 - 312 ( 294 - 316)
   INTEGRAL    Likelihood = -8.39   Transmembrane   83 -  99 (  78 - 104)
   INTEGRAL    Likelihood = -6.26   Transmembrane  212 - 228 ( 210 - 231)
   INTEGRAL    Likelihood = -4.04   Transmembrane  113 - 129 ( 110 - 132)
   INTEGRAL    Likelihood = -3.61   Transmembrane  140 - 156 ( 134 - 157)
   INTEGRAL    Likelihood = -2.71   Transmembrane  165 - 181 ( 165 - 181)
   INTEGRAL    Likelihood = -1.06   Transmembrane  327 - 343 ( 327 - 343)
   INTEGRAL    Likelihood = -0.22   Transmembrane   50 -  66 (  50 -  66)

----- Final Results -----
              bacterial membrane --- Certainty=0.6456(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 9175> which encodes the amino acid sequence <SEQ ID 9176>. Analysis of this protein sequence reveals the following:

```
     Possible site: 49
>>> Seems to have no N-terminal signal sequence
     INTEGRAL    Likelihood =-13.64   Transmembrane   24 -  40 (  17 -  52)
     INTEGRAL    Likelihood = -8.97   Transmembrane  250 - 266 ( 237 - 277)
     INTEGRAL    Likelihood = -8.65   Transmembrane  287 - 303 ( 285 - 307)
     INTEGRAL    Likelihood = -8.39   Transmembrane   74 -  90 (  69 -  95)
     INTEGRAL    Likelihood = -6.26   Transmembrane  203 - 219 ( 201 - 222)
     INTEGRAL    Likelihood = -4.04   Transmembrane  104 - 120 ( 101 - 123)
     INTEGRAL    Likelihood = -3.61   Transmembrane  131 - 147 ( 125 - 148)
     INTEGRAL    Likelihood = -2.71   Transmembrane  156 - 172 ( 156 - 172)
     INTEGRAL    Likelihood = -1.06   Transmembrane  318 - 334 ( 318 - 334)
     INTEGRAL    Likelihood = -0.22   Transmembrane   41 -  57 (  41 -  57)

----- Final Results -----
              bacterial membrane --- Certainty=0.646(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 80/326 (24%), Positives = 157/326 (47%), Gaps = 34/326 (10%)

Query: 10  LLILLILLIAAIILFLIYGIPTDANEFL----------IIYILKTRYQKLIALILVGICI 59
           +L++L LL  A+I   + G+       +          +  I   R+ +++  +L G  I
Sbjct: 34  VLLILSLLFLAVIALSLGGLAVSYGAIVKGLFVAYDPQVALIYDLRFPRIVIALLAGAGI 93

Query: 60  GSSSLIFQTLTNNRLLTPSIIGL---DSLYILIQTGLMYLIGAQRVIKFSSFSSFL---L 113
             S ++FQ +  N +  P+IIG+    S  +L+ + L+       +++ +    SFL   +
Sbjct: 94  AVSGVLFQAVLKNPISDPAIIGICSGASFMVLVSSLLL-----PQLLLYGPIVSFLGGGV 148

Query: 114 SLLLMVGFAYLLFTILFRNKKQSLYFVLLAGLIFNTLFSSISSFIQAIMDPNDFMILQNQ 173
           S LL+ G A+         K  +    ++L G+  N LF  +S+ + +       M+  N
Sbjct: 149 SFLLIYGLAW--------KKGLNPIRLILTGIAINALFMGLSTALTSFFTSASPMV--NA 198

Query: 174 LFASFNAINTKI-LWISFIIIVVSFVINWPFIKELDVLLLGKENAISLGISYQKLTTRFF 232
           L A   +  T   + + F    +  ++      K  ++LLL  +    LGI    L
Sbjct: 199 LLAGHISQKTWADVGVLFPYTFIGLLLALLLSKTCNLLLLDDQVIRHLGIDATALRLGIS 258

Query: 233 LWLALMVAIATALVGPITFLGLLVAHITYHSFHTFRHQILVPIAIVICIFTLVLGQHLVQ 292
           L   L+ ++AT++VG ++FLGL+V H++     + +HQIL+P + ++  F  +L   L +
```

```
Sbjct: 259 LVAVLLASVATSIVGVVSFLGLIVPHMSRLLVGS-KHQILIPFSALLGAFVFLLADTLGR 317

Query: 293 NLLH-LTVQLSVLLNLIGGSYFIFTL 317
           +L + L +  +++++++GG YFI+ L
Sbjct: 318 SLAYPLEISPAIIMSIVGGPYFIYLL 343
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2491> which encodes amino acid sequence <SEQ ID 2492>. An alignment of the GAS and GBS sequences follows:

```
 Score = 51.9 bits (122), Expect = 5e-08
 Identities = 73/327 (22%), Positives = 137/327 (41%), Gaps = 38/327 (11%)

Query: 494 IISSLGTAISTVAQGIGTGLAIAFRGLGAAIAMVPPTTWLALGTAILMVGAAFALAGTQA 553
           +I  L T    +  G    L IA   +GA + +V      A+    L++ A
Sbjct: 573 VILGLVTTAVMMLLGAIAPLVIAIGAIGAPVGIVVAAIVGAIAVITLIIQAIMNWGA--- 629

Query: 554 DGISQILRTIGDXXXXXXXXXTDSLATLLTIIANAIGSMLPIVAGAISQIVG-------A 606
             I++ L++  D           ++ T  T    A  +    ++G  S +V        +
Sbjct: 630 --ITEWLQSTWDSCAAWXSELWTNIVTTAT---TAWSNFTAWLSGLWSSVVSTGQSLWSS 684

Query: 607 VAGGLSQLIIAVSTGVSLVIGAFTGLLGGI-SGVINSISAVIQSLTGVITAVFNGIATVI 665
              LS +  ++ TG   +  +FT  L  + SG++++ S +  +L+  I+ +FNGI +
Sbjct: 685 FTSSLSNIFSSLITGAQSLWSSFTSTLSNLWSGLVSTGSNLFNNLSSTISGIFNGILSTA 744

Query: 666 SSVGSTIKDVLTGLGTAFEGFGNGVKSALEGVGAVIESFGSAVR--------NVLDGVAN 717
           S++ ++IK  ++    A +G  N V +   GV A+   F   ++           + G AN
Sbjct: 745 SNIWNSIKSTIS---NAIDGAKNAVSN---GVNAIKNLFNFQIKWPHIPLPHFRVSGSAN 798

Query: 718 ILDSM--GTAALNAGRGVKEMAKGIKMLVDLSLGDLVATLAAVASGLGKMASSAGEMTTL 775
            LD +  G  ++    G+   AKG  ++   +L  +   A V   G A      +TL
Sbjct: 799 PLDWLKGGLPSI----GIDWYAKG-GIMTKPTLFGMNGNRAMVGGEAGAEAILPLNKSTL 853

Query: 776 GSAMSKVANGMTRLATSATIAITGLTV 802
           G+    +AN M   + +  +  +G+T+
Sbjct: 854 GAIGQSIANTM-NTSNNINVNFSGVTI 879
 Score = 33.2 bits (74), Expect = 0.019
 Identities = 83/477 (17%), Positives = 175/477 (36%), Gaps = 103/477 (21%)

Query: 420 GSFLDKISTKFGLFGKKAKEGTD--------------QAANGSRKSGGIISQIFNGLGNI 465
           G  + +++T+FGL G+K K  ++              +A  ++++          LG +
Sbjct: 313 GDAVGELNTQFGLTGEKLKSASELLIKYAEINETDISSSAISAKQAIEAYGLTAEDLGMV 372

Query: 466 VKSAGTAISTAAKGIGTGIKTALSGAPPIISSLGTAISTVA--------QGIGTGLAIA- 516
           + +   A     + + T ++ A+ GAP  I  LG +    A          G+ +  A++
Sbjct: 373 LDNVTKAAQDTGQSVDTIVQKAIDGAPQ-IKGLGLSFEEGAALIGKFEKSGVDSSAALSS 431

Query: 517 ---------------FRGLGAAIAMVPPTT--WLALGTAILMVGAAFALAGTQA------ 553
                         GL   ++ +  +T    AL  A  + G+  A     A
Sbjct: 432 LSKAAVIYAKDGKTLTDGLNETVSAIQNSTSETEALSIASEIFGSKAAPRMVDAIQRGAF 491

Query: 554 --DGISQILRTIGDXXXXXXXXXTDSLATLLTI--------IANAIGSMLPIVAGAISQIV 604
             D +++  ++           D +  L           +A   G +L  V  A   ++
Sbjct: 492 SFDDLAEAAKSSSGTVSTTFDETLDPIDKLTQYSNQAKEGMAELGGKLLETVIPALEPLM 551

Query: 605 GAVAGGLS----------QLII---AVSTGVSLVIGAFTGL---LGGISGVINSISAVIQ 648
           G +   ++          Q I+    V+T V +++GA   L   +G I   + + A I
Sbjct: 552 GMLESSVNWFTSLNETDQQTIVILGLVTTAVMMLLGAIAPLVIAIGAIGAPVGIVVAAIV 611

Query: 649 SLTGVITAVFNGI----------------ATVISSVGSTIKDVLTGLGTAFEGFGNGVK 691
              VIT +  I                  A   S + + I    T   + F  + +G+
Sbjct: 612 GAIAVITLIIQAIMNWGAITEWLQSTWDSCAAWXSELWTNIVTTATTAWSNFTAWLSGLW 671

Query: 692 SALEGVG-AVIESFGSAVRNV----LDGVANILDSMGTAALNAGRGVKEMAKGIKMLVDL 746
           S++   G ++  SF S++ N+     + G  ++  S   A  N   G+       +
Sbjct: 672 SSVVSTGQSLWSSFTSSLSNIFSSLITGAQSLWSSFTSTLSNLWSGLVSTGSNL------ 725

Query: 747 SLGDLVATLAAVASGLGKMASSAGEMTTLGSAMSKVANGMTRLATSATIAITGLTVF 803
```

```
              +L +T++ + +G+  +++++    ++ S +S   +G     ++   AI  L  F
Sbjct: 726 -FNNLSSTISGIFNGI--LSTASNIWNSIKSTISNAIDGAKNAVSNGVNAIKNLFNF 779
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 622

A DNA sequence (GBSx0662) was identified in *S.agalactiae* <SEQ ID 1935> which encodes the amino acid sequence <SEQ ID 1936>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2277(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 623

A DNA sequence (GBSx0663) was identified in *S.agalactiae* <SEQ ID 1937> which encodes the amino acid sequence <SEQ ID 1938>. This protein is predicted to be membrane protein (ceuB). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.30   Transmembrane  241 - 257 ( 237 - 274)
   INTEGRAL    Likelihood = -6.42   Transmembrane  127 - 143 ( 118 - 149)
   INTEGRAL    Likelihood = -5.79   Transmembrane  152 - 168 ( 150 - 174)
   INTEGRAL    Likelihood = -5.47   Transmembrane  312 - 328 ( 309 - 330)
   INTEGRAL    Likelihood = -4.83   Transmembrane  289 - 305 ( 287 - 308)
   INTEGRAL    Likelihood = -4.67   Transmembrane   24 -  40 (  22 -  46)
   INTEGRAL    Likelihood = -4.35   Transmembrane   69 -  85 (  68 -  86)
   INTEGRAL    Likelihood = -4.19   Transmembrane  200 - 216 ( 198 - 216)
   INTEGRAL    Likelihood = -2.76   Transmembrane  107 - 123 ( 107 - 123)
   INTEGRAL    Likelihood = -0.85   Transmembrane  258 - 274 ( 258 - 274)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 8621> which encodes amino acid sequence <SEQ ID 8622> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 2
SRCFLG: 0
McG: Length of UR:   23
     Peak Value of UR:   2.64
     Net Charge of CR: 2
McG: Discrim Score:      8.59
GvH: Signal Score (-7.5): -4.6
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
```

```
Amino Acid Composition: calculated from 1
ALOM program   count: 9 value: -11.30 threshold:  0.0
   INTEGRAL    Likelihood =-11.30   Transmembrane  226 - 242 ( 222 - 259)
   INTEGRAL    Likelihood = -6.42   Transmembrane  112 - 128 ( 103 - 134)
   INTEGRAL    Likelihood = -5.79   Transmembrane  137 - 153 ( 135 - 159)
   INTEGRAL    Likelihood = -4.67   Transmembrane    9 -  25 (   7 -  31)
   INTEGRAL    Likelihood = -4.35   Transmembrane   54 -  70 (  53 -  71)
   INTEGRAL    Likelihood = -4.19   Transmembrane  185 - 201 ( 183 - 201)
   INTEGRAL    Likelihood = -3.08   Transmembrane  268 - 284 ( 265 - 284)
   INTEGRAL    Likelihood = -2.76   Transmembrane   92 - 108 (  92 - 108)
   INTEGRAL    Likelihood = -0.85   Transmembrane  243 - 259 ( 243 - 259)
   PERIPHERAL  Likelihood =  5.73     203
 modified ALOM score:   2.76
icm1 HYPID: 7  CFP: 0.552

*** Reasoning Step: 3

----- 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>
```

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

```
>GP:CAB12188 GB:Z99106 similar to ferrichrome ABC transporter
          (permease) [Bacillus subtilis]
 Identities = 149/304 (49%), Positives = 234/304 (76%)

Query: 29  LVILSLTSLFVGVKSIPLEQITHLDQSQVDIFLTSRLPRTISILISGASLSVCGLLMQQL 88
           L+IL++TS+F+GV+ +    +  L + +       SRLPR ISI+I+G S+S+CGL+MQQ+
Sbjct: 10  LIILAVTSVFIGVEDLSPLDLFDLSKQEASTLFASRLPRLISIVIAGLSMSICGLIMQQI 69

Query: 89  TQNKFVSPTTSGTMDWAKLGVVVTLIFFKNTSIFIQLCIASGFAILGSLLFVTILKMITF 148
           ++NKFVSPTT+GTMDWA+LG++++L+ F + S  I++ +A  FA+ G+ LF+ IL+ I F
Sbjct: 70  SRNKFVSPTTAGTMDWARLGILISLLLFTSASPLIKMLVAFVFALAGNFLFMKILERIKF 129

Query: 149 KDNIFIPLIGLMLGQIVAAATVFLGTHFQVLQSVNSWLQGNFSIMTSHRYEILYLALPCL 208
            D IFIPL+GLMLG IV++   F+   + ++Q+V+SWLQG+FS++   RYE+LYL++P +
Sbjct: 130 NDTIFIPLVGLMLGNIVSSIATFIAYKYDLIQNVSSWLQGDFSLVVKGRYELLYLSIPLV 189

Query: 209 FLVYFFAHQFTIVGLGESFAKNLGVAYEKMIYFGLVLVSIMTSLVIIIVGALPFLGLIVP 268
            + Y +A +FT+ G+GESF+ NLG+ Y++++  GL++VS++TSLVI+ VG LPFLGLI+P
Sbjct: 190 IIAYVYADKFTLAGMGESFSVNLGLKYKRVVNIGLIIVSLITSLVILTVGMLPFLGLIIP 249

Query: 269 NLISITKGDHMSSTILETSLLGACIVMICDLFGRLVIFPYEVSIGVTLGVLGSAFFLISI 328
           N++SI +GD++ S++  T LLGA  V+ CD+ GR++IFPYE+SIG+ +G+ GS  FL  +
Sbjct: 250 NIVSIYRGDNLKSSLPHTVLLGAVFVLFCDILGRIIIFPYEISIGLMVGIIGSGIFLFML 309

Query: 329 IRNE 332
           +R +
Sbjct: 310 LRRK 313
```

There is also homology to SEQ ID 1940.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 624

A DNA sequence (GBSx0664) was identified in *S.agalactiae* <SEQ ID 1941> which encodes the amino acid sequence <SEQ ID 1942>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.90   Transmembrane  140 - 156 ( 140 - 156)
```

```
----- 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>
```

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

```
>GP:BAB06720 GB:AP001517 maltose transacetylase (maltose
          O-acetyltransferase) [Bacillus halodurans]
 Identities = 93/182 (51%), Positives = 125/182 (68%), Gaps = 2/182 (1%)

Query: 2   TEKEKMLAGQYYRPSAPELRKDREVALKNMQAFNN--EDNSSKRNVILQKWFGATGKSIH 59
           TEKEKMLAG+ Y+   PEL KDRE A +  + FN   E    +R  ++++ FG+ G+S++
Sbjct: 3   TEKEKMLAGERYKAWDPELVKDRERARRLTRLFNQTTETEEKQRTELIKELFGSMGESVN 62

Query: 60  MEQRFVCDYGCNIYVGENFYANFNQTFLDVCEIRIGDNCMFGPNCQLLTPLHPLDPIERN 119
           +E  F CDYG NI+VG NF+ANF+   LDVCE+RIG NCM  P   + T  HP+ P+ER
Sbjct: 63  IEPTFRCDYGYNIHVGNNFFANFDCVILDVCEVRIGANCMLAPGVHIYTATHPIHPLERV 122

Query: 120 SGLEYGAPIQIGNNVWLGGGVTILPGVVLGDNVVVGAGSVVTKSFENNVVIAGNPAKIIKKL 182
            G EYG P+ I NNVW+GG   + PGV +G+N V+ +GSVVTK    NVV+AGNPAK+I+ +
Sbjct: 123 EGPEYGKPVTIRNNVWIGGRAIVNPGVTIGNNAVIASGSVVTKDVPENVVVAGNPAKVIQTI 184
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1943> which encodes the amino acid sequence <SEQ ID 1944>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4052(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 68/188 (36%), Positives = 101/188 (53%), Gaps = 13/188 (6%)

Query: 2   TEKEKMLAGQYYRPSAPELRKDREVALKNMQAFN--------NEDNSSKRNVILQKWFGA 53
           TE +KM  G++Y     +   D E+  K M A +              +R+ +L + FG
Sbjct: 3   TEFDKMTRGEWY-----DANFDSELIQKRMMAQDLCFDLNQLKPSREEERSAVLNQLFGQ 57

Query: 54  TGKSIHMEQRFVCDYGCNIYVGENFYANFNQTFLDVCEIRIGDNCMFGPNCQLLTPLHPL 113
           + + + +   F+CDYG NI  G+N + N N  F+D  +I +GDN  GP+    T  HPL
Sbjct: 58  SFEGLVLLSPFICDYGKNITFGKNCFINSNCYFMDGAKIALGDNVFVGPSTGFYTANHPL 117

Query: 114 DPIERNSGLEYGAPIQIGNNVWLGGGVTILPGVVLGDNVVVGAGSVVTKSFENNVVIAGN 173
           D   RN GLE   PI IG+NVW G  V ++PGV +G   V+ +GSVVT     N + AG
Sbjct: 118 DYKRRNEGLEKALPITIGDNVWFGANVNVMPGVTIGSGCVIASGSVVTHDIPVNSLAAGV 177

Query: 174 PAKIIKKL 181
           P ++++K+
Sbjct: 178 PCQVVRKI 185
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 625

A DNA sequence (GBSx0665) was identified in *S.agalactiae* <SEQ ID 1945> which encodes the amino acid sequence <SEQ ID 1946>. This protein is predicted to be ribonuclease H (rnhB-2). Analysis of this protein sequence reveals the following:

```
Possible site: 32
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16   Transmembrane   79 -  95 (  79 -  95)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9823> which encodes amino acid sequence <SEQ ID 9824> was also identified.

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

```
>GP:CAB13479 GB:Z99112 ribonuclease H [Bacillus subtilis]
 Identities = 128/249 (51%), Positives = 168/249 (67%)

Query: 4   TIKEIKAILETIVDLKDKRWQEYQTDSRAGVQKAILQRKKNIQSDLDEEARLEQMLVYEK 63
           T+K+IK  L+ + D +D    + + D R  VQ  + Q  K    +   + +   M  YE+
Sbjct: 5   TVKDIKDRLQEVKDAQDPFIAQCENDPRKSVQTLVEQWLKKQAKEKALKEQWVNMTSYER 64

Query: 64  KLYIEHINLIAGIDEVGRGPLAGPVVAAAVILPPNCKIKHLNDSKKIPKKKHQEIYQNIL 123
               +   LIAG+DEVGRGPLAGPVVA+AVILP  C+I  L DSKK+ +KK +E Y+ I+
Sbjct: 65  LARNKGFRLIAGVDEVGRGPLAGPVVASAVILPEECEILGLTDSKKLSEKKREEYYELIM 124

Query: 124 DQALAVGIGIQDSQCIDDINIYEATKHAMIDAVSHLSVAPEHLLIDAMVLDLSIPQTKII 183
            +ALAVGIGI ++  ID+INIYEA+K AM+ A+  LS  P++LL+DAM L L   Q  II
Sbjct: 125 KEALAVGIGIVEATVIDEINIYEASKMAMVKAIQDLSDTPDYLLVDAMTLPLDTAQASII 184

Query: 184 KGDANSLSIAAASIVAKVTRDKIMSDYDSTYPGYAFSKNAGYGTKEHLEGLQKYGITPIH 243
           KGDA S+SIAA + +AKVTRD++MS Y  TYP Y F KN GYGTKEHLE L  YG T +H
Sbjct: 185 KGDAKSVSIAAGACIAKVTRDRMMSAYAETYPMYGFEKNKGYGTKEHLEALAAYGPTELH 244

Query: 244 RKSFEPIKS 252
           RK+F P++S
Sbjct: 245 RKTFAPVQS 253
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1947> which encodes the amino acid sequence <SEQ ID 1948>. Analysis of this protein sequence reveals the following:

```
      Possible site: 50

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.53   Transmembrane   79 -  95 (  79 -  95)

----- Final Results -----
               bacterial membrane --- Certainty=0.1213(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:

```
>GP:CAB13479 GB:Z99112 ribonuclease H [Bacillus subtilis]
 Identities = 130/252 (51%), Positives = 176/252 (69%), Gaps = 3/252 (1%)

Query: 4   SIKAIKESLEAVTSLLDPLFQELATDTRSGVQKALKSRQKVIQAELAEEERLEAMLSYEK 63
           ++K IK+ L+ V    DP   +   D R  VQ  ++   K    E A +E+   M SYE+
Sbjct: 5   TVKDIKDRLQEVKDAQDPFIAQCENDPRKSVQTLVEQWLKKQAKEKALKEQWVNMTSYER 64

Query: 64  ALYKKGYKAIAGIDEVGRGPLAGPVVAACVILPKYCKIKGLNDSKKIPKAKHETIYQAVK 123
               KG++ IAG+DEVGRGPLAGPVVA+ VILP+ C+I GL DSKK+ + K E  Y+ +
Sbjct: 65  LARNKGFRLIAGVDEVGRGPLAGPVVASAVILPEECEILGLTDSKKLSEKKREEYYELIM 124

Query: 124 EKALAIGIGIIDNQLIDEVNIYEATKLAMLEAIKQLEGQLTQPDYLLIDAMTLDIAISQQ 183
           ++ALA+GIGI++  +IDE+NIYEA+K+AM++AI+ L      PDYLL+DAMTL +  +Q
Sbjct: 125 KEALAVGIGIVEATVIDEINIYEASKMAMVKAIQDLS---DTPDYLLVDAMTLPLDTAQA 181
```

```
Query: 184 SILKGDANSLSIAAASIVAKVTRDQMMANYDRIFPGYDFAKNAGYGTKEHLQGLKAYGIT 243
           SI+KGDA S+SIAA + +AKVTRD+MM+ Y   +P Y F KN GYGTKEHL+ L AYG T
Sbjct: 182 SIIKGDAKSVSIAAGACIAKVTRDRMMSAYAETYPMYGFEKNKGYGTKEHLEALAAYGPT 241

Query: 244 PIHRKSFEPVKS 255
            +HRK+F PV+S
Sbjct: 242 ELHRKTFAPVQS 253
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 168/256 (65%), Positives = 203/256 (78%), Gaps = 3/256 (1%)

Query: 1   MMATIKEIKAILETIVDLKDKRWQEYQTDSRAGVQKAILQRKKNIQSDLDEEARLEQMLV 60
           M  +IK IK  LE +  L D  +QE  TD+R+GVQKA+  R+K IQ++L EE RLE ML
Sbjct: 1   MPTSIKAIKESLEAVTSLLDPLFQELATDTRSGVQKALKSRQKVIQAELAEEERLEAMLS 60

Query: 61  YEKKLYIEHINLIAGIDEVGRGPLAGPVVAAAVILPPNCKIKHLNDSKKIPKKKHQEIYQ 120
           YEK LY +    IAGIDEVGRGPLAGPVVAA VILP  CKIK LNDSKKIPK KH+ IYQ
Sbjct: 61  YEKALYKKGYKAIAGIDEVGRGPLAGPVVAACVILPKYCKIKGLNDSKKIPKAKHETIYQ 120

Query: 121 NILDQALAVGIGIQDSQCIDDINIYEATKHAMIDAVSHLS---VAPEHLLIDAMVLDLSI 177
            + ++ALA+GIGI D+Q ID++NIYEATK AM++A+  L       P++LLIDAM LD++I
Sbjct: 121 AVKEKALAIGIGIIDNQLIDEVNIYEATKLAMLEAIKQLEGQLTQPDYLLIDAMTLDIAI 180

Query: 178 PQTKIIKGDANSLSIAAASIVAKVTRDKIMSDYDSTYPGYAFSKNAGYGTKEHLEGLQKY 237
            Q  I+KGDANSLSIAAASIVAKVTRD++M++YD  +PGY F+KNAGYGTKEHL+GL+ Y
Sbjct: 181 SQQSILKGDANSLSIAAASIVAKVTRDQMMANYDRIFPGYDFAKNAGYGTKEHLQGLKAY 240

Query: 238 GITPIHRKSFEPIKSM 253
           GITPIHRKSFEP+KSM
Sbjct: 241 GITPIHRKSFEPVKSM 256
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 626

A DNA sequence (GBSx0666) was identified in *S.agalactiae* <SEQ ID 1949> which encodes the amino acid sequence <SEQ ID 1950>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

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

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

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

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

Example 627

A DNA sequence (GBSx0667) was identified in *S.agalactiae* <SEQ ID 1951> which encodes the amino acid sequence <SEQ ID 1952>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3034(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 GENPEPT database:

```
>GP:BAB06195 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 140/281 (49%), Positives = 195/281 (68%), Gaps = 5/281 (1%)

Query: 3   TIQWFPGHMSKARRQVQENIKHVDFVTILVDARLPLSSQNPMLTKIVGDKPKLMILNKAD 62
            TIQWFPGHM+KARR+V E +K +D V  L+DAR+PLSS+NPM+ +IV  KP+L++LNK D
Sbjct: 2   TIQWFPGHMAKARREVTEKLKLIDVVIELLDARVPLSSRNPMMDEIVAHKPRLVLLNKDD 61

Query: 63  LADPIRTKEWRDFYESQGLKTLAINSKEQSTVKKVTDIAKILMSDKIANLRGRGIQKETL 122
            LADP +TKEW  F+E  G   L IN++    V +++   + L    I   R +G++   +
Sbjct: 62  LADPSKTKEWTRFFEEGGATVLPINAQTGQGVSRISPACQTLAQALIEKQRAKGMKPRAI 121

Query: 123 RTMIIGIPNAGKSTLMNRLAGKKIAVVGNKPGVTKGQQWLKSNKELEILDTPGILWPKFE 182
            R MI+GIPN GKSTL+NRLA K+IA VG++PG+TK QQW+K  KELE+LDTPGILWPKF+
Sbjct: 122 RAMILGIPNVGKSTLINRLASKRIAKVGDRPGITKQQQWIKVGKELELLDTPGILWPKFD 181

Query: 183 DELVGLKLALTGAIKDQLLPMDEVTIFGLNYFKTYYPDRLKERFKSINLEDEAPEIIMAL 242
            D+  G +LA TGAIKD+LL   +V +F L Y +  YPDRL +R+K   L ++   + A+
Sbjct: 182 DQATGFRLAATGAIKDELLDFQDVALFVLRYMREMYPDRLMDRYKLNELPEDGVTLFDAI 241

Query: 243 TQKLGY-----RDDYDRFYNLFVKEVRDGKLGRYTLDIVGE 278
             +K G+       DYD+   + ++E+R G LGR TL++ G+
Sbjct: 242 GKKRGHLLSGGYIDYDKTAEMILRELRAGTLGRITLEVPGK 282
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1953> which encodes the amino acid sequence <SEQ ID 1954>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2688(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 247/282 (87%), Positives = 265/282 (93%)

Query: 1   MATIQWFPGHMSKARRQVQENIKHVDFVTILVDARLPLSSQNPMLTKIVGDKPKLMILNK 60
            MA IQWFPGHMSKARRQVQEN+KHVDFVTILVDARLPLSSQNPMLTKIVGDKPKLMILNK
Sbjct: 1   MAMIQWFPGHMSKARRQVQENVKHVDFVTILVDARLPLSSQNPMLTKIVGDKPKLMILNK 60

Query: 61  ADLADPIRTKEWRDFYESQGLKTLAINSKEQSTVKKVTDIAKILMSDKIANLRGRGIQKE 120
            ADLAD  RTKEW+ +YESQG+KTLAINSKEQSTVKKVT+ AK LM+DKI  LR RGIQKE
Sbjct: 61  ADLADATRTKEWKAYYESQGIKTLAINSKEQSTVKKVTEAAKELMADKIQRLRERGIQKE 120

Query: 121 TLRTMIIGIPNAGKSTLMNRLAGKKIAVVGNKPGVTKGQQWLKSNKELEILDTPGILWPK 180
            TLRTMIIGIPNAGKSTLMNRLAGKKIAVVGNKPGVTKGQQWLKSNKELEILDTPGILWPK
Sbjct: 121 TLRTMIIGIPNAGKSTLMNRLAGKKIAVVGNKPGVTKGQQWLKSNKELEILDTPGILWPK 180

Query: 181 FEDELVGLKLALTGAIKDQLLPMDEVTIFGLNYFKTYYPDRLKERFKSINLEDEAPEIIM 240
            FEDELVGLKLALTGAIKDQLLPMDEVTIFGLNYF+ YYP+RL +RFK+I LE+EAPEIIM
Sbjct: 181 FEDELVGLKLALTGAIKDQLLPMDEVTIFGLNYFREYYPNRLTKRFKNIPLEEEAPEIIM 240

Query: 241 ALTQKLGYRDDYDRFYNLFVKEVRDGKLGRYTLDIVGEHDGN 282
             LT++LG++DDYDRFY LFVKEVRDGKLG+YTLD VG+ D +
Sbjct: 241 TLTRQLGFKDDYDRFYTLFVKEVRDGKLGQYTLDQVGDMDAD 282
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 628

A DNA sequence (GBSx0668) was identified in *S.agalactiae* <SEQ ID 1955> which encodes the amino acid sequence <SEQ ID 1956>. 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>
```

A related GBS nucleic acid sequence <SEQ ID 9825> which encodes amino acid sequence <SEQ ID 9826> was also identified.

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

```
>GP:CAB12129 GB:Z99105 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 69/173 (39%), Positives = 102/173 (58%), Gaps = 13/173 (7%)

Query: 29  DKAKEKASV-----IKQASQTSQTSKKEVLQKKT----YPNLNKYSNLEIHVSSTRQTMT 79
           D A+E AS+     ++  +   +T+K  +   K      YP++ K  ++ I V+   Q
Sbjct: 22  DHAEEHASINTKKTVENITDVRKTAKTSIDWTKPSGGEYPDI-KQKHVWIDVNVKEQKAY 80

Query: 80  ITSNDKVIFKTIVSTG---AKESPTPKGTFVIEPERGDFFYNASSKEGAYYWVSFKEHGI 136
           I      I+  ++S+G     K+  TPKGTF +EPERG++F++    +EGA YWVS+K HG
Sbjct: 81  IKEGSNTIYTMMISSGLDQTKDDATPKGTFYVEPERGEWFFSEGYQEGAEYWVSWKNHGE 140

Query: 137 YLFHSVPTDQQGNEIPEEAKQLGKAASHGCVRMSRADAKWFYENIPQGTTVTI 189
           +LFHSVP  +    I  EA++LG   SHGC+R++  DAKW YENIP+ T V I
Sbjct: 141 FLFHSVPMTKDQKVIKTEAEKLGTKVSHGCIRLTIPDAKWVYENIPEHTKVVI 193
```

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

SEQ ID 1956 (GBS644) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 2 & 3; MW 49.6kDa) and in Figure 186 (lane 3; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 5-7; MW 24.6kDa) and in Figure 177 (lane 3; MW 25kDa).

GBS644-GST was purified as shown in Figure 236, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 629

A DNA sequence (GBSx0669) was identified in *S.agalactiae* <SEQ ID 1957> which encodes the amino acid sequence <SEQ ID 1958>. This protein is predicted to be carbon starvation protein A. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.25   Transmembrane  129 - 145 ( 122 - 157)
   INTEGRAL    Likelihood = -9.92   Transmembrane  316 - 332 ( 305 - 342)
   INTEGRAL    Likelihood = -6.42   Transmembrane  164 - 180 ( 157 - 181)
   INTEGRAL    Likelihood = -5.73   Transmembrane  443 - 459 ( 441 - 466)
```

```
    INTEGRAL    Likelihood = -5.57   Transmembrane  416 - 432 ( 414 - 435)
    INTEGRAL    Likelihood = -4.88   Transmembrane  190 - 206 ( 183 - 209)
    INTEGRAL    Likelihood = -4.83   Transmembrane   78 -  94 (  70 -  95)
    INTEGRAL    Likelihood = -3.13   Transmembrane  362 - 378 ( 359 - 379)
    INTEGRAL    Likelihood = -2.34   Transmembrane  228 - 244 ( 227 - 245)
    INTEGRAL    Likelihood = -2.02   Transmembrane    2 -  18 (   1 -  18)
    INTEGRAL    Likelihood = -1.28   Transmembrane  393 - 409 ( 393 - 410)

----- 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>
```

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

```
>GP:AAF93852 GB:AE004154 carbon starvation protein A, putative
          [Vibrio cholerae]
 Identities = 220/470 (46%), Positives = 311/470 (65%), Gaps = 16/470 (3%)

Query: 1   MVTFLGGVALLIVGYFTYGRYIEKNFQIDENRQTPAEALRDGYDFVPMPKWKNGMIELLN 60
           M+ FL  VA L+ GYF YG ++EK F I+E RQTPA    DG D+VPM   K  +++LLN
Sbjct: 1   MLWFLTCVAALVGGYFIYGAFVEKVFGINEKRQTPAHTKTDGVDYVPMSTPKVYLVQLLN 60

Query: 61  IAGTGPIFGPILGALYGFVAYIWIVLGCIFAGAVHDYMIGMISLRNNGAYLPELASRYLG 120
           IAG GPIFGPI+GALYGP A +WIV+GCIFAGAVHDY  GM+S+RN GA +P +  RYLG
Sbjct: 61  IAGVGPIFGPIMGALYGPAAMLWIVVGCIFAGAVHDYFSGMLSIRNGGASVPSITGRYLG 120

Query: 121 KSMKHVINIFSMLLLILVATVFVVTPANLILSILPAG---TLSLPWIIGLIFVYYLISTV 177
              KH +NIF+++LL+LV  VFV  PA +I +++       T+S+  ++ +IF YY+++T+
Sbjct: 121 NGAKHFMNIFAIVLLLLVGVVFVSAPAGMITNLINQQTDFTVSMTTMVVIIFAYYILATI 180

Query: 178 LPIDKALGKVYPVF-------CVILMVSTAAVGFRLLTGGFDMPNLTFETFKNMHPAGLG 230
           +P+DK +G+ YP+F        V LM + A       + GGF++ ++      KN++P  +
Sbjct: 181 VPVDKIIGRFYPLFGALLIFMSVGLMTAIAFSSEHQVLGGFEISDMV----KNLNPNDMP 236

Query: 231 IFPALFFTISCGAISGFHATQAPMVSRTTVNEREGRFTFYGMMIAEGVIAMIWAGASMSL 290
           ++PALF TI+CGAISGFHATQ+P+++R   NE+ GRF FYG MI EG+IA+IW   ++S
Sbjct: 237 LWPALFITIACGAISGFHATQSPLMARCMENEKNGRFVFYGAMIGEGIIALIWCTVALSF 296

Query: 291 FKG-QNLYEMIAAGTPSAVVNQVMLMLLGSVIGTIAIIGVIVLPVSSGLSAFRSLRTIVA 349
           F   + L E +  G P  VV      LLG   G IA +GV++LP++SG +AFRS R I+A
Sbjct: 297 FGSLEALSEAVKNGGPGNVVYGASFGLLGVFGGVIAFLGVVILPITSGDTAFRSSRLILA 356

Query: 350 DYIHVKQDTLPKIFAVTIPLYVISFVLTHVDFNLLWRYFNWANQVTAVIGLLVATRYLIL 409
           +Y +++Q TL      + +PL+VI  VLT VDF ++WRYF +ANQ TAV+ L  AT YL+
Sbjct: 357 EYFNMEQKTLRNRLLMAVPLFVIGAVLTQVDFGIIWRYFGFANQATAVMMLWTATAYLMR 416

Query: 410 KRRNYWVTFVPAMFMLYAVVVYIL-SQPIGFNMGLGILTYSLALVLTGIL 458
             + +W+   VPA+FM     + +IL S  +GF + + I T +   L   G L
Sbjct: 417 HNKLHWICTVPALFMTTVCISFILNSSTLGFGLPMQISTIAGVLASLGAL 466
```

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

A related GBS gene <SEQ ID 8623> and protein <SEQ ID 8624> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
McG: Discrim Score:      6.07
GvH: Signal Score (-7.5): -3.54
     Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 11 value: -11.25 threshold:  0.0
   INTEGRAL    Likelihood =-11.25   Transmembrane  129 - 145 ( 122 - 157)
   INTEGRAL    Likelihood = -9.92   Transmembrane  316 - 332 ( 305 - 342)
   INTEGRAL    Likelihood = -6.42   Transmembrane  164 - 180 ( 157 - 181)
   INTEGRAL    Likelihood = -5.57   Transmembrane  416 - 432 ( 414 - 435)
   INTEGRAL    Likelihood = -4.88   Transmembrane  190 - 206 ( 183 - 209)
```

```
    INTEGRAL    Likelihood = -4.83   Transmembrane   78 -  94 (  70 -  95)
    INTEGRAL    Likelihood = -4.67   Transmembrane  445 - 461 ( 441 - 463)
    INTEGRAL    Likelihood = -3.13   Transmembrane  362 - 378 ( 359 - 379)
    INTEGRAL    Likelihood = -2.34   Transmembrane  228 - 244 ( 227 - 245)
    INTEGRAL    Likelihood = -2.02   Transmembrane    2 -  18 (   1 -  18)
    INTEGRAL    Likelihood = -1.28   Transmembrane  393 - 409 ( 393 - 410)
    PERIPHERAL  Likelihood =  0.21     272
 modified ALOM score:   2.75

*** Reasoning Step: 3

----- 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>
```

The protein has homology with the following sequences in the databases:

```
ORF01729(301 - 1668 of 2082)
GP|9655126|gb|AAF93852.1||AE004154(1 - 464 of 494) carbon starvation protein A, putative
(Vibrio cholerae)
%Match = 29.9
%Identity = 47.6  %Similarity = 68.6
Matches = 218  Mismatches = 138  Conservative Sub.s = 96

174       204       234       264       294       324       354       384
TNEKLFIIKLRLFISKKQPFILKIGNFNFSMLY*SHENA**N*AKKFLGGSDMVTFLGGVALLIVGYFTYGRYIEKNFQI
                                                  |: ||  || |: ||| || ::|| | |
                                                 MLWFLTCVAALVGGYFIYGAFVEKVFGI
                                                           10        20

414       444       474       504       534       564       594       624
DENRQTPAEALRDGYDFVPMPKWKNGMIELLNIAGTGPIFGPILGALYGPVAYIWIVLGCIFAGAVHDYMIGMISLRNNG
:| |||||    || |:|||   | :::|||||| |||||||:|||||| | :|||:|||||||||||  ||:|:|| |
NEKRQTPAHTKTDGVDYVPMSTPKVYLVQLLNIAGVGPIFGPIMGALYGPAAMLWIVVGCIFAGAVHDYFSGMLSIRNGG
          40        50        60        70        80        90        100

654       684       714       744       774       795       825       855
AYLPELASRYLGKSMKHVINIFSMLLLILVATVFVVTPANLILSILPAGT---LSLPWIIGLIFVYYLISTVLPIDKALG
| :| :  ||||   || :|||:::||:||  |||  || :| :::   |   :|:  :: :|| ||:::|::|:|| :|
ASVPSITGRYLGNGAKHFMNIFAIVLLLLVGVVFVSAPAGMITNLINQQTDFTVSMTTMVVIIFAYYILATIVPVDKIIG
          120       130       140       150       160       170       180

894       924       954       984       1014      1044      1074
KVYP------VFCVI-LMVSTAAVGFRLLTGGFDMPNLTFETFKNMHPAGLGIFPALFFTISCGAISGFHATQAPMVSRT
: ||      :|  : || : |      : |||:: ::     ||::|  : ::|||| ||:|||||||||||:|:::|
RFYPLFGALLIFMSVGLMTAIAFSSEHQVLGGFEISDM----VKNLNPNDMPLWPALFITIACGAISGFHATQSPLMARC
          200       210       220           230       240       250       260

1104      1134      1164      1191      1221      1251      1281      1311
TVNEREGRFTFYGMMIAEGVIAMIWAGASMSLFKG-QNLYEMIAAGTPSAVVNQVMLMLLGSVIGTIAIIGVIVLPVSSG
  ||: ||| ||| || ||:||:||   ::|:|   : | | :  | |  ||    : |||   | || :||::||::||
MENEKNGRFVFYGAMIGEGIIALIWCTVALSFFGSLEALSEAVKNGGPGNVVYGASFGLLGVFGGVIAFLGVVILPITSG
              280       290       300       310       320       330       340

1341      1371      1401      1431      1461      1491      1521      1551
LSAFRSLRTIVADYIHVKQDTLPKIFAVTIPLYVISFVLTHVDFNLLWRYFNWANQVTAVIGLLVATRYLILKRRNYWVT
 :|||| | |:|:| :::| ||   : : :||:||  |||:||| ::|||| :||| |||: |  || ||:   : :|:
DTAFRSSRLILAEYFNMEQKTLRNRLLMAVPLFVIGAVLTQVDFGIIWRYFGFANQATAVMMLWTATAYLMRHNKLHWIC
              360       370       380       390       400       410       420

1581      1608      1638      1668      1698      1728      1758      1788
FVPAMFMLYAVVVYIL-SQPIGFNMGLGILTYSLALVLTGIXVGLFWKSGQKQLKTVHPEAFLFNDHRPINYYSSLDS*Y
 |||:||    : :|| |  :|| : : | | :  |   |
TVPALFMTTVCISFILNSSTLGFGLPMQISTIAGVLASLGALAYVAKVSKGKGETDLADEEKPQGVTKTA
              440       450       460       470       480       490
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 630

A DNA sequence (GBSx0670) was identified in *S.agalactiae* <SEQ ID 1959> which encodes the amino acid sequence <SEQ ID 1960>. This protein is predicted to be lytR (lytT). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.80   Transmembrane   27 -  43 (  27 -  43)

----- 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>
```

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

```
>GP:AAB48183 GB:L42945 lytR [Staphylococcus aureus]
 Identities = 93/245 (37%), Positives = 150/245 (60%), Gaps = 3/245 (1%)

Query: 1   MKVLVVDDEPVARNELIYLLNKYDSNLVIAEAHDMATALAILLRETFDVALLDIHLRDDS 60
           MK L++DDEP+ARNEL YLLN+       I EA ++   L  LL   +D+  LD++L D++
Sbjct: 1   MKALIIDDEPLARNELTYLLNEIGGFEEINEAENVKETLEALLINQYDIIFLDVNLMDEN 60

Query: 61  GLQLAEYINKMPKPPLLIFATAYDQYAIQAFEHDARDYLLKPYDFDRLKQAMDRVKGALS 120
           G++L   I KM +PP +IFATA+DQYA+QAFE +A DY+LKP+   R++QA+++V+   +
Sbjct: 61  GIELGAKIQKMKEPPAIIFATAHDQYAVQAFELNATDYILKPFGQKRIEQAVNKVRATKA 120

Query: 121 TSTIIESVTSGPL---FKQQYPLTVEDRIYLVSADDILLIEAMQGKLIIQTPDKNYEIDG 177
                S  +  +   F Q  P+ ++D+I+++    +I+ I    G   I T +  YE
Sbjct: 121 KDDNNASAIANDMSANFDQSLPVEIDDKIHMLKQQNIIGIGTHNGITTIHTTNHKYETTE 180

Query: 178 SLQQWQDKLPSSQFVRVHRSYIVNINAIKTIEPWFNQTLQLHLCNKITVPVSRANVKPLK 237
            L +++ +L  + F+R+HRSYI+N   IK ++ WFN T  + L N + + V R+ +K  K
Sbjct: 181 PLNRYEKRLNPTYFIRIHRSYIINTKHIKEVQQWFNYTYMVILTNGVKMQVGRSFMKDFK 240

Query: 238 QMLGI 242
             +G+
Sbjct: 241 ASIGL 245
```

There is also homology to SEQ ID 460.

SEQ ID 1960 (GBS399) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 7; MW 30.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 2; MW 55kDa). Purified GBS399-GST is shown in Figure 217, lane 9; purified GBS399d-GST is shown in Figure 236, lane 3.

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

Example 631

A DNA sequence (GBSx0671) was identified in *S.agalactiae* <SEQ ID 1961> which encodes the amino acid sequence <SEQ ID 1962>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.59   Transmembrane   95 - 111 (  86 - 116)
```

```
   INTEGRAL     Likelihood = -5.95   Transmembrane  155 - 171 ( 152 - 176)
   INTEGRAL     Likelihood = -2.28   Transmembrane  189 - 205 ( 187 - 206)
   INTEGRAL     Likelihood = -1.49   Transmembrane  122 - 138 ( 121 - 138)

----- 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>
```

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

```
>GP:AAB48182 GB:L42945 lytS [Staphylococcus aureus]
 Identities = 264/570 (46%), Positives = 389/570 (67%), Gaps = 2/570 (0%)

Query: 1   MTLFLIMMERAGLIILLAYAFVHIPFIKQTLKQPELKKHQYILLILFSLFAIISNFTGVE 60
           ++L ++++ER GLII+LAY  ++IP+ K  + +     K ++ L I+FSLFA++SN TG+
Sbjct: 2   LSLTMLLLERVGLIIILAYVLMNIPYFKNLMNRRRTWKARWQLCIIFSLFALMSNLTGIV 61

Query: 61  IQSDLSIIPQTLNHIADQSSVANTRVLTIGVSGLIGGPIVGIIVGLLSVFVRYLQGGLAP 120
           I    S+       + D  S+ANTRVLTIGV+GL+GGP VG+ VG++S   R   GG
Sbjct: 62  IDHQHSLSGSVYFRLDDDVSLANTRVLTIGVAGLVGGPFVGLFVGVISGIFRVYMGGADA 121

Query: 121 HIYVISSLLIGLCSGLSGNYLRKNYNKIRVLDAMVVGFGMEILQMICILIFSVDFNQALR 180
            +Y+ISS+ IG+ +G  G   ++      +  + ++G  ME++QM+ IL FS D   A+
Sbjct: 122 QVYLISSIFIGIIAGYFGLQAQRRKRYPSIAKSAMIGIVMEMIQMLSILTFSHDKAYAVD 181

Query: 181 LVSFISMPMILSNTLGLGIFISIISSTQKLEEHAKAFQTHQVLELANLTLPYLRKGLTTE 240
           L+S I++PMI+ N++G  IF+SII  T K E+  K  QTH VL+L N T PY ++GL  E
Sbjct: 182 LISLIALPMIIVNSVGPAIFMSIIIPTLKQEDQMKPVQTHDVLQLMNQTFPYFKEGLNRE 241

Query: 241 SCQPVAEIIHKHMDVSAVSLTSQSAILAYVGDGADHHLPNTQILTKLAKRAIDTGKVSVA 300
           S Q +A II   M VS+V++TS++ IL++VG G+DHH+P  +ILT L+K  + +GK+
Sbjct: 242 SAQQIAMIIKNLMKVSSVAITSKNEILSHVGGGSDHHIPTNEILTSLSKDVLKSGKLKEV 301

Query: 301 TDKSEIECDHKNCPLSSAIVIPLHIHDVIVGTLKLYFSDAQHMTYVDRQLAEGLGNIFST 360
             K EI C H NCPL +AIVIPL +H  IVGTLK+YF++   +T+V+RQLAEGL NIFS+
Sbjct: 302 HTKEEIGCSHPNCPLRAAIVIPLEMHGSIVGTLKMYFTNPNDLTFVERQLAEGLANIFSS 361

Query: 361 QLALGQAEEATRLLQDAEMKSLQAQVNPHFLFNALNTIYGLIRMDSEKARKLVQDFSKVI 420
           Q+ LG+AE  ++LL+DAE+KSLQAQV+PHF FN++N I  L+R++SEKAR+L+ + S
Sbjct: 362 QIELGEAETQSKLLKDAEIKSLQAQVSPHFFFNSINPISALVRINSEKARELLLELSYFF 421

Query: 421 RANLQRAKQNLIPLHDELEQVNAYLALEEARFPNMVAFNLDNQTNSDDNLMIPPFTLQVL 480
           RANLQ +KQ+ I L  EL QV AYL+LE+AR+P     N++ +    D +++PPF +Q+L
Sbjct: 422 RANLQGSKQHTITLDKELSQVRAYLSLEQARYPGRFNININVEDKYRD-VLVPPFLIQIL 480

Query: 481 IENSYKHAFKHVNKNNQLKVTIARNN-DRLHIIVQDNGIGIPKEKLITLGKKTQISKQGS 539
           +EN+ KHAF +  + N + V++ +     + IIVQDNG GI K+K+  LG+ +  S+ G+
Sbjct: 481 VENAIKHAFTNRKQGNDIDVSVIKETATHVRIIVQDNGQGISKDKMHLLGETSVESESGT 540

Query: 540 GTAIENLVRRLNIIYDGQASLKFESNDSGT 569
           G+A+ENL  RL  ++   A+L+FES  SGT
Sbjct: 541 GSALENLNLRLKGLFGKSAALQFESTSSGT 570
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1963> which encodes the amino acid sequence <SEQ ID 1964>. Analysis of this protein sequence reveals the following:

```
      Possible site: 39

>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood = -6.79   Transmembrane  283 - 299 ( 276 - 307)
   INTEGRAL     Likelihood = -5.57   Transmembrane   27 -  43 (  24 -  48)

----- Final Results -----
               bacterial membrane --- Certainty=0.3718(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:

```
>GP:CAB54576 GB:AJ006396 histidine kinase [Streptococcus pneumoniae]
 Identities = 115/231 (49%), Positives = 159/231 (68%), Gaps = 7/231 (3%)

Query: 351 MLASIKAYIDEVYVLEVEQRDAQMRALQSQINPHFLYNTLEYIRMYALSCQQEELADVIY 410
           ML  ++  I ++Y LE+ Q+DA MRALQ+QINPHF+YNTLE++RMYA+   Q+ELAD+IY
Sbjct: 1   MLDRLEKNIHDIYQLELSQKDANMRALQAQINPHFMYNTLEFLRMYAVMQSQDELADIIY 60

Query: 411 AFASLLRNNISQDKMTTLKEELAFCEKYIYLYQMRYPDSFAYHVKIDESVADLAIPKFVI 470
            F+SLLRNNIS ++ T LK+EL FC KY YL  +RYP S AY  KID  + ++ IPKF +
Sbjct: 61  EFSSLLRNNISDERETLLKQELEFCRKYSYLCMVRYPKSIAYGFKIDPELENMKIPKFTL 120

Query: 471 QPLVENYFVHGIDYSRHDNALSIKALDETDHLLIQVLDNGRGISQERLADMEKRLQ---- 526
           QPLVENYF HG+D+ R DN +SIKAL +   + I V+DNGRG+S E+LA++ ++L
Sbjct: 121 QPLVENYFAHGVDHRRTDNVISIKALKQDGFVEILVVDNGRGMSAEKLANIREKLSQRYF 180

Query: 527 EHQTT---GNSSIGLQNVYLRLFHHFRDRVSWSMAKEPNGGFIIQIRIRKD 574
           EHQ +      SIG+ NV+ R   +F DR + ++      G   +I I+ +
Sbjct: 181 EHQASYSDQRQSIGIVNVHERFVLYFGDRYAITIESAEQAGVQYRITIQDE 231
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 59/180 (32%), Positives = 97/180 (53%), Gaps = 8/180 (4%)

Query: 375 QDAEMKSLQAQVNPHFLFNALNTI--YGLIRMDSEKARKLVQDFSKVIRANLQRAKQNLI 432
           +DA+M++LQ+Q+NPHFL+N L  I  Y L     E A  ++  F+ ++R N+ + K  +
Sbjct: 370 RDAQMRALQSQINPHFLYNTLEYIRMYALSCQQEELA-DVIYAFASLLRNNISQDK--MT 426

Query: 433 PLHDELEQVNAYLALEEARFPNMVAFNLDNQTNSDDNLMIPPFTLQVLIENSYKHAFKHV 492
            L +EL     Y+ L + R+P+  A+++    +  D L IP F +Q L+EN + H   +
Sbjct: 427 TLKEELAFCEKYIYLYQMRYPDSFAYHVKIDESVAD-LAIPKFVIQPLVENYFVHGIDYS 485

Query: 493 NKNNQLKVTIARNNDRLHIIVQDNGIGIPKEKLITLGKKTQISKQ--GSGTAIENLVRRL 550
            +N L +      D L I V DNG GI +E+L  + K+ Q  +    S   ++N+ RL
Sbjct: 486 RHDNALSIKALDETDHLLIQVLDNGRGISQERLADMEKRLQEHQTTGNSSIGLQNVYLRL 545
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 632

A DNA sequence (GBSx0672) was identified in *S.agalactiae* <SEQ ID 1965> which encodes the amino acid sequence <SEQ ID 1966>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> 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>
```

A related GBS nucleic acid sequence <SEQ ID 9827> which encodes amino acid sequence <SEQ ID 9828> was also identified.

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

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 633

A DNA sequence (GBSx0673) was identified in *S.agalactiae* <SEQ ID 1967> which encodes the amino acid sequence <SEQ ID 1968>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.55   Transmembrane   52 -  68 (  45 -  74)
   INTEGRAL    Likelihood = -9.18   Transmembrane   83 -  99 (  76 - 106)
   INTEGRAL    Likelihood = -8.76   Transmembrane  126 - 142 ( 118 - 146)
   INTEGRAL    Likelihood = -7.48   Transmembrane  174 - 190 ( 170 - 191)
   INTEGRAL    Likelihood = -3.66   Transmembrane  195 - 211 ( 193 - 212)
   INTEGRAL    Likelihood = -1.28   Transmembrane   24 -  40 (  24 -  40)

----- Final Results -----
               bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

A related GBS gene <SEQ ID 8625> and protein <SEQ ID 8626> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     -8.54
GvH: Signal Score (-7.5): -5.6
     Possible site: 57
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 6 value:  -9.55 threshold:  0.0
   INTEGRAL    Likelihood = -9.55   Transmembrane   52 -  68 (  45 -  74)
   INTEGRAL    Likelihood = -9.18   Transmembrane   83 -  99 (  76 - 106)
   INTEGRAL    Likelihood = -8.76   Transmembrane  126 - 142 ( 118 - 146)
   INTEGRAL    Likelihood = -7.48   Transmembrane  174 - 190 ( 170 - 191)
   INTEGRAL    Likelihood = -3.66   Transmembrane  195 - 211 ( 193 - 212)
   INTEGRAL    Likelihood = -1.28   Transmembrane   24 -  40 (  24 -  40)
   PERIPHERAL  Likelihood = 13.05     100
 modified ALOM score:   2.41

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 634

A DNA sequence (GBSx0674) was identified in *S.agalactiae* <SEQ ID 1969> which encodes the amino acid sequence <SEQ ID 1970>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.53   Transmembrane   83 -  99 (  83 -  99)

----- Final Results -----
               bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 635

A DNA sequence (GBSx0675) was identified in *S.agalactiae* <SEQ ID 1971> which encodes the amino acid sequence <SEQ ID 1972>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1902(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 636

A DNA sequence (GBSx0676) was identified in *S.agalactiae* <SEQ ID 1973> which encodes the amino acid sequence <SEQ ID 1974>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4763(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 637

A DNA sequence (GBSx0677) was identified in *S.agalactiae* <SEQ ID 1975> which encodes the amino acid sequence <SEQ ID 1976>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5089(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 638

A DNA sequence (GBSx0678) was identified in *S.agalactiae* <SEQ ID 1977> which encodes the amino acid sequence <SEQ ID 1978>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> 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>
```

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

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

SEQ ID 1978 (GBS184) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 7; MW 21kDa), in Figure 168 (lane 14-16; MW 36kDa – thioredoxin fusion) and in Figure 238 (lane 9; MW 36kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 7; MW 46.4kDa).

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

Example 639

A DNA sequence (GBSx0679) was identified in *S.agalactiae* <SEQ ID 1979> which encodes the amino acid sequence <SEQ ID 1980>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2179(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 640

A DNA sequence (GBSx0680) was identified in *S.agalactiae* <SEQ ID 1981> which encodes the amino acid sequence <SEQ ID 1982>. This protein is predicted to be immunogenic secreted protein precursor. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2166(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9351> which encodes amino acid sequence <SEQ ID 9352> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1983> which encodes the amino acid sequence <SEQ ID 1984>. Analysis of this protein sequence reveals the following:

```
     Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.77   Transmembrane    9 -  25 (   5 -  27)

----- Final Results -----
               bacterial membrane --- Certainty=0.2508(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 64/86 (74%), Positives = 76/86 (87%)

Query: 1   MGNGGDWKNKPGYQTTHEAKTGYAISFSPGQAGADRTYGHVAIVEDVKEDGSIPISESNV 60
           MGNGGDW+ KPG+ TTH+ K GY +SF+PGQAGAD TYGHVA+VE +KEDGSI ISESNV
Sbjct: 452 MGNGGDWQRKPGFVTTHKPKVGYVVSFAPGQAGADATYGHVAVVEQIKEDGSILISESNV 511

Query: 61  LGLGTISYRTFSAAEAAQLTYVVGEK 86
           +GLGTISYRTF+A +A+ LTYVVG+K
Sbjct: 512 MGLGTISYRTFTAEQASLLTYVVGDK 537
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 641

A DNA sequence (GBSx0681) was identified in *S.agalactiae* <SEQ ID 1985> which encodes the amino acid sequence <SEQ ID 1986>. This protein is predicted to be immunogenic secreted protein precursor. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2495(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 GENPEPT database:

```
AAB52379 GB:U31811 immunogenic secreted protein precursor [Streptococcus pyogenes]
 Identities = 133/259 (51%), Positives = 170/259 (65%), Gaps = 4/259 (1%)

Query: 3   PSQPQVTATPQKSEVVTPAITSGIDLPDVAIPTAMASAAYVKHWIGNDAYTHNLLSHRYG 62
           P QP + A    + V  P   S  DL  +  P++ +SAAYV+HW G+ AYTHNLLS RYG
Sbjct: 174 PIQPPLGAA---APVFAPWRESDKDLSKLK-PSSRSSAAYVRHWTGDSAYTHNLLSRRYG 229

Query: 63  ITAAQLDGFLQSTGITYDSSRIDGQKILDREKSSGLDARAIIAIAIAESSLGTQGVATAP 122
           ITA QLDGFL S GI YD  R++G+++L+ EK +GLD RAI+AIA+AESSLGTQGVA
```

```
Sbjct: 230 ITAEQLDGFLNSLGIHYDKERLNGKRLLEWEKLTGLDVRAIVAIAMAESSLGTQGVAKEK 289

Query: 123 GANMFGFGAVDNNTTNAQNFSDDKAVIKMTQETIIQNQNTSFAIQDQKAQFLSTGNLNVA 182
           G+NMFG+GA D N  NA+ +SD+ A+  M ++TII N+N +F  QD KA+  S G L+
Sbjct: 290 GSNMFGYGAFDFNPNNAKKYSDEVAIRHMVEDTIIANKNQTFERQDLKAKKWSLGQLDTL 349

Query: 183 ARGGVYFTDASGSGKRRAAIMESIDKWIDAHGGISEISKELLNTSSVAMMAVPTSYSVSR 242
             GGVYFTD SGSG+RRA IM  +D+WID HG   +I + L  TS      VP  Y  S+
Sbjct: 350 IDGGVYFTDTSGSGQRRADIMTKLDQWIDDHGNTPDIPEHLKITSGTQFSEVPVGYKRSQ 409

Query: 243 ANQAGNYVAGTYPWGQRTW 261
                Y + TY +GQ TW
Sbjct: 410 PQNVLTYKSETYSFGQCTW 428
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1987> which encodes the amino acid sequence <SEQ ID 1988>. 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>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 143/265 (53%), Positives = 184/265 (68%), Gaps = 5/265 (1%)

Query: 2   VPSQPQVTATPQKSEVVTPA-----ITSGIDLPDVAIPTAMASAAYVKHWIGNDAYTHNL 56
           V + P  + + Q  E  TP        S +DL ++ IP+    AAYV+HW G +AYTH+L
Sbjct: 135 VDTAPASSLSKQLPEARTPIQSLSPYVSDLDLSEIDIPSVNTYAAYVEHWSGKNAYTHHL 194

Query: 57  LSHRYGITAAQLDGFLQSTGITYDSSRIDGQKILDREKSSGLDARAIIAIAIAESSLGTQ 116
           LS RYGI A Q+D +L+STGI YDS+RI+G+K+L  EK SGLD RAI+AIA++ESSLGTQ
Sbjct: 195 LSRRYGIKADQIDSYLKSTGIAYDSTRINGEKLLQWEKKSGLDVRAIVAIAMSESSLGTQ 254

Query: 117 GVATAPGANMFGFGAVDNNTTNAQNFSDDKAVIKMTQETIIQNQNTSFAIQDQKAQFLST 176
           G+AT  GANMFG+ A D + T A  F+DD A++KMTQ+TII+N+N++FA+QD KA   S
Sbjct: 255 GIATLLGANMFGYAAFDLDPTQASKFNDDSAIVKMTQDTIIKNKNSNFALQDLKAAKFSR 314

Query: 177 GNLNVAARGGVYFTDASGSGKRRAAIMESIDKWIDAHGGISEISKELLNTSSVAMMAVPT 236
           G LN A+ GGVYFTD +GSGKRRA IME +DKWID HGG   I  EL   SS +  +VP
Sbjct: 315 GQLNFASDGGVYFTDTTGSGKRRAQIMEDLDKWIDDHGGTPAIPAELKVQSSASFASVPA 374

Query: 237 SYSVSRANQAGNYVAGTYPWGQRTW 261
            Y +S++     Y A +Y WGQ TW
Sbjct: 375 GYKLSKSYDVLGYQASSYAWGQCTW 399
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 642

A DNA sequence (GBSx0682) was identified in *S.agalactiae* <SEQ ID 1989> which encodes the amino acid sequence <SEQ ID 1990>. Analysis of this protein 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>
```

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

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

A related GBS gene <SEQ ID 8627> and protein <SEQ ID 8628> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:     11.56
GvH: Signal Score (-7.5): 0.870001
     Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:  11.88 threshold:  0.0
   PERIPHERAL  Likelihood = 11.88      63
 modified ALOM score:  -2.88

*** 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 8628 (GBS159) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 28 (lane 4; MW 26kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 2; MW 41kDa).

GBS159-GST was purified as shown in Figure 198, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 643

A DNA sequence (GBSx0683) was identified in *S.agalactiae* <SEQ ID 1991> which encodes the amino acid sequence <SEQ ID 1992>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2668(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 GENPEPT database:

```
>GP:BAB04699 GB:AP001510 unknown conserved protein [Bacillus halodurans]
 Identities = 32/76 (42%), Positives = 54/76 (70%)

Query: 7  LGSVIELKNDSQKVMITSRFPLYDNEGQLGYFDYSGCIFPISIVGNETYFFNLEDIDKVL 66
          +GS++ LK  + K+MI +R P+ +  G+   FDYSGC +P  +V ++ ++FN E+ID+V+
Sbjct: 4  IGSIVYLKEGTSKLMILNRGPILEANGENKMFDYSGCFYPQGLVPDKVFYFNHENIDEVV 63

Query: 67 FEGYYDENEEEMQKIF 82
          FEG+ D+ E+  QK+F
Sbjct: 64 FEGFQDDEEQRFQKLF 79
```

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

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

Example 644

A DNA sequence (GBSx0684) was identified in *S.agalactiae* <SEQ ID 1993> which encodes the amino acid sequence <SEQ ID 1994>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.81   Transmembrane   75 -  91 (  69 -  99)
   INTEGRAL    Likelihood =-14.38   Transmembrane  134 - 150 ( 129 - 179)
   INTEGRAL    Likelihood = -8.49   Transmembrane  157 - 173 ( 151 - 179)
   INTEGRAL    Likelihood = -1.17   Transmembrane   50 -  66 (  46 -  67)

----- Final Results -----
               bacterial membrane --- Certainty=0.6922(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 645

A DNA sequence (GBSx0685) was identified in *S.agalactiae* <SEQ ID 1995> which encodes the amino acid sequence <SEQ ID 1996>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.11   Transmembrane   40 -  56 (  40 -  56)

----- Final Results -----
               bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

SEQ ID 1996 (GBS204) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 13; MW 32kDa) and Figure 53 (lane 2; MW 14.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 6; MW 39.7kDa).

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

Example 646

A DNA sequence (GBSx0686) was identified in *S.agalactiae* <SEQ ID 1997> which encodes the amino acid sequence <SEQ ID 1998>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence (or signal = aa 1-26)

----- 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>
```

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

```
>GP:CAC16670 GB:AJ302698 hypothetical protein [Staphylococcus
          haemolyticus]
 Identities = 60/254 (23%), Positives = 109/254 (42%), Gaps = 14/254 (5%)

Query: 2   VKVSVSSVGTQASTVAISMFSRVSALNDAITKLSSFAEAATLQGTAYSNAKSYATGTLTP 61
           + + V    +Q+S V  ++ S  S  +     +  F  A+ LQG AY + K + +  + P
Sbjct: 3   IDMYVGKSKSQSSDVGSTVKSISSGYDSLQKGIMQFVGASELQGQAYDSGKQFFSAVIAP 62

Query: 62  MLQGMILFSETLSEKCTELQTLYVSICGDEDLDSVVLESKLASDRASLKIAEALLEHLND 121
           + + +    E   + C +    Y S    + L    L   +      +   EA+   L
Sbjct: 63  LTESIKTLGELTEQACNDFVDQYQSEVDSQSLKESELLEDIEELNKQISQLEAMNASLKH 122

Query: 122 DPEPSKSAISSTKSNIKKLKKRIKSNQKKLDNLNEFNAHSATVFADISNAQSTVNQALAA 181
               + S +S     I  L+++ K  ++KL  L +F+A S  +F ++ + Q TV Q +
Sbjct: 123 KSSKNSSLLSGNHQMISSLEQQKKELEEKLRKLRQFDAKSPNIFKEVESFQKTVQQGINQ 182

Query: 182 VSTGFSGYNSKTGAFGKPTSGQMEWTKTVKKNWKEREDAKAEELKSKKAEESKKASKIEN 241
             T    ++     F  P    MEW K  ++   E    K +++ ++KA++ KK SK +
Sbjct: 183 AKT---AWDPGKQTFNIPAGKDMEWAKVSQQKALE---VKMDKI-NQKAKDGKKLSKNDI 235

Query: 242 TT-------KKSNV 248
           T       KKSN+
Sbjct: 236 FTIIAYQQQKKSNI 249
```

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

SEQ ID 1998 (GBS270) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 2; MW 34.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 7; MW 59.2kDa).

The GBS270-GST fusion product was purified (Figure 206, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 265), which confirmed that the protein is immunoaccessible on GBS bacteria.

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

Example 647

A DNA sequence (GBSx0687) was identified in *S.agalactiae* <SEQ ID 1999> which encodes the amino acid sequence <SEQ ID 2000>. This protein is predicted to be outer surface protein F. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

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

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

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

SEQ ID 2000 (GBS316) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 3; MW 23kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 2; MW 41.8kDa).

GBS316-GST was purified as shown in Figure 206, lane 4.

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

Example 648

A DNA sequence (GBSx0688) was identified in *S.agalactiae* <SEQ ID 2001> which encodes the amino acid sequence <SEQ ID 2002>. This protein is predicted to be actin-like protein arp3 (act4). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0217(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 649

A DNA sequence (GBSx0689) was identified in *S.agalactiae* <SEQ ID 2003> which encodes the amino acid sequence <SEQ ID 2004>. This protein is predicted to be diarrheal toxin. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.65   Transmembrane   65 -  81 (  61 -  84)
   INTEGRAL    Likelihood = -3.98   Transmembrane   89 - 105 (  85 - 106)

----- 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>
```

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

```
>GP:CAB15175 GB:Z99120 alternate gene name: yueA~similar to
           hypothetical proteins [Bacillus subtilis]
 Identities = 452/1058 (42%), Positives = 664/1058 (62%), Gaps = 39/1058 (3%)

Query: 98   VTMIFSITGYFKNRKQYKQDLQERIDSYHDYLSDKSIELQKLAKEQKRGQHYHYPTIEGL 157
            +T+I S   YF+++ Q K+  ++R   Y  YL +K  ELQ LA++QK+    +H+P+ E +
Sbjct: 1    MTLITSTVQYFRDKNQRKKREEKRERVYKLYLDNKRKELQALAEKQKQVLEFHFPSFEQM 60

Query: 158  QEMADTYHHRIYEKTPLHFDFLYYRLGLGEVPTSYNIHYSQPERSGKK-DPLENEGYNLY 216
            + +      RI+EK+    D+L  RLG G VP+SY I+ S  + + +  D L  +  ++
Sbjct: 61   KYLTSEISDRIWEKSLESKDYLQLRLGTGTVPSSYEINMSGGDLANRDIDDLMEKSQHMQ 120

Query: 217  FNNRYIKNMPIVANLSHGPVGYIGPRGLVLEQLQLMVNQLAFFHSYHDVQFITIVPEEEM 276
```

```
                + I+N P+  +L+ GP+G +G   +V  ++  ++ QL+FF+SYHD++F+ I  EEE
Sbjct: 121  RVYKDIRNAPVTVDLAEGPMGLVGKSQIVKNEIHQLIGQLSFFNSYHDLRFVFIFHEEEY 180

Query: 277  DKWSWMRWLPHETLQDVNVRGFVYNQRSRDQVLNSLNQILKLRRTQREDKSAKEGTLFSP 336
              W WM+ +P   +  +  +GF+YN+++RDQ+L+SL ++++    +R+ +  KE   F P
Sbjct: 181  KDWEWMKCVPQFQMPHIYAKGFIYNEQTRDQLLSSLYELIR----ERDLEDDKEKLQFKP 236

Query: 337  HYVVIVTDEKLILDHVIMEFFTEDPTELGCSLIFVQDVMSSLSENIKTIINIKDRNTGQL 396
             H+V ++T+++LI +HVI+E+       LG S I   +  SLSENI T++   + + G +
Sbjct: 237  HFVFVITNQQLISEHVILEYLEGQHEHLGISTIVAAETKESLSENITTLVRYINEHEGDI 296

Query: 397  VIEEGELKETDFELDHFLEDYDKENISRRLAPLNHLQNLKSSIPEAVTFMEMYQAEEFED 456
             +I++ +     F LDH   + D E  SR L  LNH   + +SIPE V+F+E++ A+E ++
Sbjct: 297  LIQKKKAVRIPFRLDHHQRE-DNERFSRTLRTLNHQVGITNSIPETVSFLELFHAKEVKE 355

Query: 457  LHVQERWISHAPYKSSAVPLGLRGQDDIVYLNLHEKAHGPHGLVAGTTGSGKSEIIQSYI 516
             + +Q+RW++    KS +VP+G +G+DDIVYLNLHEKAHGPHGL+AGTTGSGKSE +Q+YI
Sbjct: 356  IGIQQRWLTSESSKSLSVPIGYKGKDDIVYLNLHEKAHGPHGLLAGTTGSGKSEFLQTYI 415

Query: 517  LSLAVNFHPHDVAFLLIDYKGGGMANLFKDLPHLLGTITNLDGAQ--SMRALVSINAELK 574
             LSLAV+FHPH+ AFLLIDYKGGGMA  F+++PHLLGTITN++G++  SMRAL SI +ELK
Sbjct: 416  LSLAVHFHPHEAAFLLIDYKGGGMAQPFRNIPHLLGTITNIEGSKNFSMRALASIKSELK 475

Query: 575  RRQRLFAKADVNHINQYQKKYKLGEVSEPMPHLFLISDEFAELKSNQPEFMKELVSTARI 634
             +RQRLF +  VNHIN Y K YK G+    MPHLFLISDEFAELKS +P+F++ELVS ARI
Sbjct: 476  KRQRLFDQYQVNHINDYTKLYKQGKAEVAMPHLFLISDEFAELKSEEPDFIRELVSAARI 535

Query: 635  GRSLGIHLILATQKPSGVVDDQIWSNSRFKLALKVADRGDSMEMLHTPDAAEITQAGRAY 694
             GRSLG+HLILATQKP G++DDQIWSNSRFK+ALKV D  DS E+L   DAA IT  GR Y
Sbjct: 536  GRSLGVHLILATQKPGGIIDDQIWSNSRFKVALKVQDATDSKEILKNSDAANITVTGRGY 595

Query: 695  LQVGNNEVYELFQSAWSGADYQPEKDDQGIEDHTIYSINDLGQYEILNDDLSGLDQAENI 754
             LQVGNNEVYELFQSAWSGA Y  E    G ED  I  + D G        LS +D  +N
Sbjct: 596  LQVGNNEVYELFQSAWSGAPYLEEV--YGTEDE-IAIVTDTGLI-----PLSEVDTEDNA 647

Query: 755  -KEVPTELDAIVENIQALTKEMGISDLPQPWLPPLSNQIAVTDLRKEGSVDLWSKAPSYK 813
              K+V TE++A+V+ I+ +  EMGI  LP PWLPPL+ +I  T         L+
Sbjct: 648  KKDVQTEIEAVVDEIERIQDEMGIEKLPSPWLPPLAERIPRT---------LFPSNEKDH 698

Query: 814  AVLGFMDIPSQQAQEVAYHDFEDDGHLSIFAGPSMGKSTALQTVTMDLARHNSPEFLNLY 873
              ++D P  Q Q    +  +DG++ IF    GKS A  T  M  A   +PE L++Y
Sbjct: 699  FHFAYVDEPDLQRQAPIAYKMMEDGNIGIFGSSGYGKSIAAATFLMSFADVYTPEELHVY 758

Query: 874  LFDFGTNGLLPLRRLPHVADFFTIDDDEKIAKFIARIKVEMSDRKKALSRYNVATAKLYR 933
             +FDFG   LLPL +LPH AD+F +D   KI KF+ RIK E+  RK+      ++  K+Y
Sbjct: 759  IFDFGNGTLLPLAKLPHTADYFLMDQSRKIEKFMIRIKEEIDRRKRLFREKEISHIKMYN 818

Query: 934  QVSGETMPQILIVIDSYEGLREAQTPTNLEACFQNISRDGSSLGISLVISAGRTAALRSS 993
              +S E +P I I ID+++ +++      LE+ F  +SRDG SLGI  +++A R  A+R S
Sbjct: 819  ALSEEELPFIFITIDNFDIVKDEM--HELESEFVQLSRDGQSLGIYFMLTATRVNAVRQS 876

Query: 994  LMANLKERIALKLTDDSESRTLVGRHQHIMEDIPGRGLIKRDDIEVLQVALSTEGTETFD 1053
             L+ NLK +I   L D SE  ++ GR +  +E IPGR +I+++++   Q+ L  +  +
Sbjct: 877  LLNNLKTKIVHYLMDQSEGYSIYGRPKFNLEPIPGRVIIQKEELYFAQMFLPVDADDDIG 936

Query: 1054 IINNIQNESDAMNSKWTG-PRPKAIPIVPEELTFDDFMATDSVQADLSANRL--PLGLEM 1110
             + N ++++   +  ++P  IP++PE L+  +     S++  L   L  P+GL
Sbjct: 937  MFNELKSDVQKLQGRFASMEQPAPIPMLPESLSTREL----SIRFKLERKPLSVPIGLHE 992

Query: 1111 VDVESYSLALNRFKHMLYMSDSDESLEAVGSHIIKVLL 1148
              V      L + KH L +  +        ++++KV+L
Sbjct: 993  ETVSPVYFDLGKHKHCLILGQTQRG----KTNVLKVML 1026
```

There is also homology to SEQ ID 24.

A related GBS gene <SEQ ID 8629> and protein <SEQ ID 8630> were also identified. Analysis of this protein sequence reveals the following:

```
Homology to a bacterial toxin
```

The protein has homology with the following sequences in the databases:

```
>OMNI|NT01BS3725 diarrheal toxin

 Score =  203 bits (511), Expect = 4e-51
 Identities = 123/377 (32%), Positives = 198/377 (51%), Gaps = 22/377 (5%)

Query: 1    MGISDLPQPWLPPLSNQIAVTDLRKEGSVDLWSKAPSYKAVLGFMDIPSQQAQEVAYHDF 60
             MGI  LP PWLPPL+ +I  T         L+           ++D P  Q Q    +
Sbjct: 704  MGIEKLPSPWLPPLAERIPRT---------LFPSNEKDHFHFAYVDEPDLQRQAPIAYKM 754

Query: 61   EDDGHLSIFAGPSMGKSTALQTVTMDLARHNSPEFLNLYLFDFGTNGLLPLRRLPHVADF 120
              +DG++ IF      GKS A  T  M  A   +PE L++Y+FDFG   LLPL +LPH AD+
Sbjct: 755  MEDGNIGIFGSSGYGKSIAAATFLMSFADVYTPEELHVYIFDFGNGTLLPLAKLPHTADY 814

Query: 121  FTIDDDEKIAKFIARIKVEMSDRKKALSRYNVATAKLYRQVSGETMPQILIVIDSYEGLR 180
             F +D   KI KF+ RIK E+  RK+      ++  K+Y  +S E +P I I ID+++ ++
Sbjct: 815  FLMDQSRKIEKFMIRIKEEIDRRKRLFREKEISHIKMYNALSEEELPFIFITIDNFDIVK 874

Query: 181  EAQTPTNLEACFQNISRDGSSLGISLVISAGRTAALRSSLMANLKERIALKLTDDSESRT 240
             +      LE+ F  +SRDG SLGI  +++A R  A+R SL+ NLK +I   L D SE  +
Sbjct: 875  DEM--HELESEFVQLSRDGQSLGIYFMLTATRVNAVRQSLLNNLKTKIVHYLMDQSEGYS 932

Query: 241  LVGRHQHIMEDIPGRGLIKRDDIEVLQVALSTEGTETFDIINNIQNESDAMNSKWTG-PR 299
             + GR +  +E IPGR +I+++++   Q+ L  +  +  + N ++++   +  ++    +
Sbjct: 933  IYGRPKFNLEPIPGRVIIQKEELYFAQMFLPVDADDDIGMFNELKSDVQKLQGRFASMEQ 992

Query: 300  PKAIPIVPEELTFDDFMATDSVQADLSANRL--PLGLEMVDVESYSLALNRFKHMLYMSD 357
             P  IP++PE L+  +     S++  L    L  P+GL    V      L + KH L +
Sbjct: 993  PAPIPMLPESLSTREL----SIRFKLERKPLSVPIGLHEETVSPVYFDLGKHKHCLILGQ 1048

Query: 358  SDESLEAVGSHIIKVLL 374
             +        ++++KV+L
Sbjct: 1049 TQRG----KTNVLKVML 1061
```

SEQ ID 8630 (GBS326) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 65 (lane 5; MW 66kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 71 (lane 5; MW 91kDa).

GBS326-GST was purified as shown in Figure 212, lane 5.

GBS326LN was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 127 (lane 2-4; MW 114kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 184 (lane 6; MW 114kDa). The purified protein is shown in Figure 236, lane 12.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 650

A DNA sequence (GBSx0690) was identified in *S.agalactiae* <SEQ ID 2005> which encodes the amino acid sequence <SEQ ID 2006>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2693(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 651

A DNA sequence (GBSx0691) was identified in *S.agalactiae* <SEQ ID 2007> which encodes the amino acid sequence <SEQ ID 2008>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3933(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 652

A DNA sequence (GBSx0692) was identified in *S.agalactiae* <SEQ ID 2009> which encodes the amino acid sequence <SEQ ID 2010>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood = -7.32   Transmembrane  225 - 241 ( 219 - 246)

----- 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>
```

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

```
>GP:BAB04693 GB:AP001510 unknown conserved protein [Bacillus halodurans]
 Identities = 83/320 (25%), Positives = 162/320 (49%), Gaps = 1/320 (0%)

Query: 103 VNFILHPSNLFLTKNATAKIAYRSLPGIMRPEKFGPEEFLYQFKCFVFALLTQHDYIELY 162
           ++ I+ P N+ ++         +  +   + P +  PE    + +  +  LL  +    Y
Sbjct: 106 LHLIVSPENVLVSDGLDVTFIHYGVKDSIPPYETDPERLFLELRATLLVLLDGNHRFHEY 165

Query: 163 NGAISVIEVSDFLKSIYHAETIQAVRDIITIDYEQQVEVETHTLAKVSRAKYKLYKYISV 222
                 +++S   KS+    T++ +R++I   + Q+ E +   L KV + K+ + K+  +
Sbjct: 166 MNYHDTLKLSPEAKSLVQQTTLEGLRELIR-HWIQEHEQQEKQLHKVPKTKWTIQKWAGI 224

Query: 223 WLGALSTILLIPLVYLVFIHNPFKEKMLAADTSFIKVDYNQVINRLEHVKVSKLPYTQKY 282
            L A     +I +VY++    P +E   A+  +++  +Y+QVI+ LE    + +P   KY
Sbjct: 225 GLIAALVPAIIYIVYVLAFLQPRQEAFTASHAAYLNENYSQVIDTLEPYSPNSMPRVVKY 284

Query: 283 ELAYSYINGMSFSEEQREVILNNVTLKTDELYLDYWINIGRGLDDDAIDAAKRLDDSDLV 342
           +LA SY+          RE + N + L+  E Y DYWI IGRG ++ AID A+ L D + +
Sbjct: 285 QLAQSYVAIEPLQAYHRENLKNVLVLQAAESYFDYWIAIGRGENEKAIDIARGLQDKEWL 344
```

```
Query: 343 IYAIVQKMDQVRKDNSLSGKDREQKLSELQTDYDKYWKDRKTALTDEESKSKNSNNHSTN 402
          +YA V++ ++V+ D +LSGK+RE  + E++ + D Y ++ +    + E+   N+   ++N
Sbjct: 345 VYANVKRREEVKSDENLSGKEREDLIKEIEAEIDDYMRELEELAEEGEAFQPNAEPAASN 404

Query: 403 SNKESSESSSTTASTSSKTK 422
             +E    +    S + + K
Sbjct: 405 ELEEDEGDTEEDDSDNQEAK 424
```

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

SEQ ID 2010 (GBS337) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 3; MW 50.3kDa).

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

Example 653

A DNA sequence (GBSx0693) was identified in *S.agalactiae* <SEQ ID 2011> which encodes the amino acid sequence <SEQ ID 2012>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-14.01   Transmembrane  131 - 147 ( 122 - 153)

----- Final Results -----
               bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8631> which encodes amino acid sequence <SEQ ID 8632> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     13.38
GvH: Signal Score (-7.5): -1.25
     Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 1 value: -14.01 threshold:  0.0
   INTEGRAL    Likelihood =-14.01   Transmembrane  127 - 143 ( 118 - 149)
   PERIPHERAL  Likelihood = 16.13     113
modified ALOM score:   3.30

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

SEQ ID 8632 (GBS140) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 32 (lane 3; 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 49 (lane 8; MW 18kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 654

A DNA sequence (GBSx0694) was identified in *S.agalactiae* <SEQ ID 2013> which encodes the amino acid sequence <SEQ ID 2014>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1486(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 655

A DNA sequence (GBSx0695) was identified in *S.agalactiae* <SEQ ID 2015> which encodes the amino acid sequence <SEQ ID 2016>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-14.59   Transmembrane  984 -1000 ( 976 -1009)
   INTEGRAL    Likelihood = -9.71   Transmembrane   19 -  35 (  15 -  42)
   INTEGRAL    Likelihood = -9.50   Transmembrane  872 - 888 ( 865 - 890)
   INTEGRAL    Likelihood = -6.37   Transmembrane  927 - 943 ( 924 - 951)
   INTEGRAL    Likelihood = -4.19   Transmembrane  831 - 847 ( 828 - 847)
   INTEGRAL    Likelihood = -2.87   Transmembrane  899 - 915 ( 899 - 916)

----- Final Results -----
            bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8633> which encodes amino acid sequence <SEQ ID 8634> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 6
SRCFLG: 0
McG: Length of UR:   20
     Peak Value of UR:   3.40
     Net Charge of CR: 3
McG: Discrim Score:     13.67
GvH: Signal Score (-7.5): -3.27
     Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 6 value: -14.59 threshold:  0.0
   INTEGRAL    Likelihood =-14.59   Transmembrane  973 - 989 ( 965 - 998)
   INTEGRAL    Likelihood = -9.71   Transmembrane    8 -  24 (   4 -  31)
   INTEGRAL    Likelihood = -9.50   Transmembrane  861 - 877 ( 854 - 879)
   INTEGRAL    Likelihood = -6.37   Transmembrane  916 - 932 ( 913 - 940)
   INTEGRAL    Likelihood = -4.19   Transmembrane  820 - 836 ( 817 - 836)
   INTEGRAL    Likelihood = -2.87   Transmembrane  888 - 904 ( 888 - 905)
  PERIPHERAL   Likelihood =  3.82     936
 modified ALOM score:   3.42
icml HYPID: 7  CFP: 0.684

*** Reasoning Step: 3
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.6838(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 GENPEPT database:

```
>GP:AAB86324 GB:AE000938 phage infection protein homolog
          [Methanothermobacter thermoautotrophicus]
 Identities = 96/454 (21%), Positives = 190/454 (41%), Gaps = 63/454 (13%)

Query: 1   MLKIKYILGRIMKR-NNFRILWYIIAVALFLVAIAGLNLKLQGDHAKENKTTQSATNTKL 59
           M K   I  + MK   N  ++ ++IAV + + A+  +   +Q      ++T+       +
Sbjct: 1   MRKALEIFWKDMKTVKNSPVVLFVIAVIICIPALYAV-FNIQATLDPYSRTSS------I 53

Query: 60  NIALVNEDQNVSNGKESYNLGASYIKSIERDNSQNWSVVSRGTAQNGLDKGDYQLMVIIP 119
            +A+VNED          N+GA ++  + ++ + +W  V R  A +GL KG Y  ++IIP
Sbjct: 54  EVAVVNEDMGADFNGTHLNVGAEFVSELRKNRNFDWQFVDRSDAMDGLRKGKYYAVLIIP 113

Query: 120 NNFSQKLLDVNKANAEQTTISYKVNAKGNLALEKKATEKEKDIVSELNSHLVNMYMASIL 179
            NFS  LL +       Q +I Y VN K N     +         + +++NS +V      +
Sbjct: 114 GNFSSDLLSIKNGTPRQASIKYMVNDKLNPVAPRITNAGADALQAKINSEVVKTIDGIVF 173

Query: 180 SNLYTAQENVQA----------MVNVQSGNISNYQKNLLDSATNF---QNIFPAL----- 221
             + A E  +A           VN  +GN+    + L  + ++    QN++ +L
Sbjct: 174 GKISEAGELARANRDDILRTKRFVNELNGNLGKIDETLSTANSDLEKGQNLWSSLKTDLP 233

Query: 222 -VNQSSSSITANESLKKS-----------LEASDNMFNDLVTTQTNTGKDLSSL----- 263
            +  +++ +    SL +S            +++ ++  ++ +T+       L+SL
Sbjct: 234 EIRDNANFVKEKYSLLESYIGKDPAKALSTVQSMESHLSEAITSMKYLRAVLASLYSATG 293

Query: 264 -------IEQRHQDSISYEAFSTSLLEMNNELLEKQLSDIITQAQKDQETLSSQLNSIMG 316
                 I+Q   +     +    L  + ++L  K  +D I + +   + + S LN +M
Sbjct: 294 DPKLKTAIDQIDTNIEKASSVLGILQTIESDLKTKGTTDRIVKLKASIDRMDSALNKLMD 353

Query: 317 D-DNNHNHKENSSAYLNVARQKIQELSEALKSQDNIAKDQSEQLDKIVREGLASYFAKNN 375
             D       +++SA L +A  +    +  A+       +D S +L+ I  + L S     +
Sbjct: 354 SRDEIDAAMQDASAKLGIANARWPTMRSAI-------QDASRKLNMISDDDLNSLVKLAD 406

Query: 376 KDNITLLELLKSHSTNEK----TLKDFKAKVADF 405
            D   + E  +S     EK      +K++ + +A F
Sbjct: 407 IDPSAVREYFRSPVRMEKEHIYPVKNYGSALAPF 440
```

SEQ ID 8634 (GBS250) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 4; MW 136kDa).

GBS250-GST was purified as shown in Figure 203, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 656

A DNA sequence (GBSx0696) was identified in *S.agalactiae* <SEQ ID 2019> which encodes the amino acid sequence <SEQ ID 2020>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.5009(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 GENPEPT database:

```
>GP:CAA46375 GB:X65276 ORFA1 [Clostridium acetobutylicum]
 Identities = 35/91 (38%), Positives = 53/91 (57%)

Query: 1  MAQIKLTPEELRSSAQKYTAGSQQVTEVLNLLTQEQAVIDENWDGSTFDSFEAQFNELSP 60
          MAQI +TPEEL+S AQ Y    +++ + +  +    + I E W G  F ++  Q+N+L
Sbjct: 1  MAQISVTPEELKSQAQVYIQSKEEIDQAIQKVNSMNSTIAEEWKGQAFQAYLEQYNQLHQ 60

Query: 61 KITEFAQLLEDINQQLLKVADIIEQTDADIA 91
           + +F  LLE +NQQL K AD + + DA  A
Sbjct: 61 TVVQFENLLESVNQQLNKYADTVAERDAQDA 91
```

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

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

Example 657

A DNA sequence (GBSx0697) was identified in *S.agalactiae* <SEQ ID 2021> which encodes the amino acid sequence <SEQ ID 2022>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3741(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 658

A repeated DNA sequence (GBSx0698) was identified in *S.agalactiae* <SEQ ID 2023> which encodes the amino acid sequence <SEQ ID 2024>. This protein is predicted to be carbamoylphosphate synthetase (carB). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -1.33   Transmembrane  807 - 823 ( 807 - 823)

----- 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>
```

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

```
>GP:CAA03928 GB:AJ000109 carbamoylphosphate synthetase [Lactococcus
             lactis]
 Identities = 771/1062 (72%), Positives = 901/1062 (84%), Gaps = 5/1062 (0%)

Query: 1    MPKRTDIRKIMVIGSGPIVIGQAAEFDYSGTQACLSLKEEGYQVVLVNSNPATIMTDKDI 60
             MPKR DI+KIM+IGSGPI+IGQAAEFDY+GT+ACL+LKEEGY+VVLVNSNPATIMTD++I
Sbjct: 1    MPKRNDIKKIMIIGSGPIIIGQAAEFDYAGTEACLALKEEGYEVVLVNSNPATIMTDREI 60
```

```
Query: 61   ADKVYIEPITLEFVTRILRKERPDALLPTLGGQTGLNMAMALSKNGILEELNVELLGTKL 120
            AD VYIEPITLEFV++ILRKERPDALLPTLGGQTGLNMAM LSK GILEELNVELLGTKL
Sbjct: 61   ADTVYIEPITLEFVSKILRKERPDALLPTLGGQTGLNMAMELSKTGILEELNVELLGTKL 120

Query: 121  SAIDKAEDRDLFKQLMEELNQPIPESEIVNSVEEAIQFAEQIGYPLIVRPAFTLGGTGGG 180
            SAID+AEDR+LFK+L E +N+P+  S+I  +VEEAI  A++IGYP+IV PAFT+GGTGGG
Sbjct: 121  SAIDQAEDRELFKELCESINEPLCASDIATTVEEAINIADKIGYPIIVGPAFTMGGTGGG 180

Query: 181  MCDNQEQLVDITTKGLKLSPVTQCLIERSIAGFKEIEYEVMRDAADNALVVCNMENFDPV 240
            +CD +E+L +I   GLKLSPVTQCLIE SIAG+KEIEYEVMRD+ADNA+VVCNMENFDPV
Sbjct: 181  ICDTEEELREIVANGLKLSPVTQCLIEESIAGYKEIEYEVMRDSADNAIVVCNMENFDPV 240

Query: 241  GIHTGDSIVFAPAQTLSDVENQLLRDASLDIIRALKIEGGCNVQLALDPNSFKYYVIEVN 300
            G+HTGDSIVFAP+QTLSD E Q+LRDASL+IIRALKIEGGCNVQLALDPNS++Y VIEVN
Sbjct: 241  GVHTGDSIVFAPSQTLSDNEYQMLRDASLNIIRALKIEGGCNVQLALDPNSYEYRVIEVN 300

Query: 301  PRVSRSSALASKATGYPIAKLAAKIAVGLTLDEVINPITKTTYAMFEPALDYVVAKMPRF 360
            PRVSRSSALASKATGYPIAK++AKIA+G+TLDE+INP+T  TYAMFEPALDYVVAK+ RF
Sbjct: 301  PRVSRSSALASKATGYPIAKMSAKIAIGMTLDEIINPVTNKTYAMFEPALDYVVAKIARF 360

Query: 361  PFDKFESGDRKLGTQMKATGEVMAIGRNIEESLLKACRSLEIGVDHIKIADLDNVSDDVL 420
            PFDKFE+GDR LGTQMKATGEVMAIGRNIEESLLKA RSLEIGV H ++ +     D+ L
Sbjct: 361  PFDKFENGDRHLGTQMKATGEVMAIGRNIEESLLKAVRSLEIGVFHNEMTEAIEADDEKL 420

Query: 421  LEKIRKAEDDRLFYLAEALRRHYSIEKLASLTSIDSFFLDKLRVIVELEDLLSKNRLDIN 480
             EK+ K +DDRLFY++EA+RR   IE++A LT ID FFLDKL  IVE+E+ L  N  +
Sbjct: 421  YEKMVKTQDDRLFYVSEAIRRGIPIEEIADLTKIDIFFLDKLLYIVEIENQLKVNIFEPE 480

Query: 481  ILKKVKNKGFSDKAIASLWQINEDQVRNMRKEAGILPVYKMVDTCASEFDSATPYFYSTY 540
            +LK  K  GFSD+ IA LW +  ++VR  R+E  I+PVYKMVDTCA+EF+S+TPYFYSTY
Sbjct: 481  LLKTAKKNGFSDREIAKLWNVTPEEVRRRRQENKIIPVYKMVDTCAAEFESSTPYFYSTY 540

Query: 541  AVENESLISDKASILVLGSGPIRIGQGVEFDYATVHSVKAIRESGFEAIIMNSNPETVST 600
              ENES  SDK  I+VLGSGPIRIGQGVEFDYATVH VKAI+  G EAI++NSNPETVST
Sbjct: 541  EWENESKRSDKEKIIVLGSGPIRIGQGVEFDYATVHCVKAIQALGKEAIVINSNPETVST 600

Query: 601  DFSISDKLYFEPLTFEDVMNVIDLEKPEGVILQFGGQTAINLAKDLNKAGVKILGTQLED 660
            DFSISDKLYFEPLTFEDVMNVIDLE+P  VI+QFGGQTAINLA+ L+KAGVKILGTQ+ED
Sbjct: 601  DFSISDKLYFEPLTFEDVMNVIDLEEPLVVIVQFGGQTAINLAEHLSKAGVKILGTQVED 660

Query: 661  LDRAENRKQFEATLQALNIPQPPGFTATTEEEAVNAAQKIGYPVLVRPSYVLGGRAMKIV 720
            LDRAE+R  FE  LQ L+IPQPPG TAT EEEAV  A KIGYPVL+RPS+VLGGRAM+I+
Sbjct: 661  LDRAEDRDLFEKALQDLDIPQPPGATATNEEEAVANANKIGYPVLIRPSFVLGGRAMEII 720

Query: 721  ENEEDLRHYMTTAVKASPDHPVLIDAYLIGKECEVDAISDGQNILIPGIMEHIERSGVHS 780
             NE+DLR YM  AVKASP+HPVL+D+YL G+ECEVDAI DG+ +L+PGIMEHIER+GVHS
Sbjct: 721  NNEKDLRDYMNRAVKASPEHPVLVDSYLQGQECEVDAICDGKEVLLPGIMEHIERAGVHS 780

Query: 781  GDSMAVYPPQTLSETIIETIVDYTKRLAIGLNCIGMMNIQFVIKDQKVYVIEVNPRASRT 840
            GDSMAVYPPQ LS+ II+TIVDYTKRLAIGLNCIGMMNIQFVI +++VYVIEVNPRASRT
Sbjct: 781  GDSMAVYPPQNLSQAIIDTIVDYTKRLAIGLNCIGMMNIQFVIYEEQVYVIEVNPRASRT 840

Query: 841  LPFLSKVTHIPMAQVATKVILGDKLCNFTYGYDLYPASDMVHIKAPVFSFTKLAKVDSLL 900
            +PFLSKVT+IPMAQ+AT++ILG+ L +  Y   L P  DMVH+KAPVFSFTKLAKVDSLL
Sbjct: 841  VPFLSKVTNIPMAQLATQMILGENLKDLGYEAGLAPTPDMVHVKAPVFSFTKLAKVDSLL 900

Query: 901  GPEMKSTGEVMGSDINLQKALYKAFEAAYLHMPDYGNIVFTVDDTDKEEALELAKVYQSI 960
            GPEMKSTG  MGSD+ L+KALYK+FEAA LHM DYG+++FTV D DKEE L LAK +  I
Sbjct: 901  GPEMKSTGLAMGSDVTLEKALYKSFEAAKLHMADYGSVLFTVADEDKEETLALAKDFAEI 960

Query: 961  GYRIYATQGTAIYFDANGLETVLVGKL--GENDRNHIPDLIKNGKIQAVINTVGQNNID- 1017
            GY + AT GTA +   NGL    V KL  GE++  + + I+ G++QAV+NT+G
Sbjct: 961  GYSLVATAGTAAFLKENGLYVREVEKLAGGEDEEGTLVEDIRQGRVQAVVNTMGNTRASL 1020

Query: 1018 --NHDALIIRRSAIEQGVPLFTSLDTAHAMFKVLESRAFTLK 1057
               D   IR+ AI +G+PLFTSLDT  A+ KV++SR+FT K
Sbjct: 1021 TTATDGFRIRQEAISRGIPLFTSLDTVAAILKVMQSRSFTTK 1062
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2025> which encodes the amino acid sequence <SEQ ID 2026>. Analysis of this protein sequence reveals the following:

```
      Possible site: 21
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.17   Transmembrane  773 - 789 ( 773 - 789)

   ----- 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>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA03928 GB:AJ000109 carbamoylphosphate synthetase [Lactococcus
             lactis]
 Identities = 753/1030 (73%), Positives = 876/1030 (84%), Gaps = 6/1030 (0%)

Query: 1    LALKEEGYKVILVNSNPATIMTDKEIADKVYIEPLTLEFVNRIIRKERPDAILPTLGGQT 60
             LALKEEGY+V+LVNSNPATIMTD+EIAD VYIEP+TLEFV++I+RKERPDA+LPTLGGQT
Sbjct: 35   LALKEEGYEVVLVNSNPATIMTDREIADTVYIEPITLEFVSKILRKERPDALLPTLGGQT 94

Query: 61   GLNMAMALSKAGILDDLEIELLGTKLSAIDQAEDRDLFKQLMQELDQPIPESTIVKTVDE 120
             GLNMAM LSK GIL++L +ELLGTKLSAIDQAEDR+LFK+L + +++P+  S I  TV+E
Sbjct: 95   GLNMAMELSKTGILEELNVELLGTKLSAIDQAEDRELFKELCESINEPLCASDIATTVEE 154

Query: 121  AVTFARDIGYPVIVRPAFTLGGTGGGICSSEEELCEITENGLKLSPVTQCLIERSIAGFK 180
             A+  A  IGYP+IV PAFT+GGTGGGIC +EEEL EI  NGLKLSPVTQCLIE SIAG+K
Sbjct: 155  AINIADKIGYPIIVGPAFTMGGTGGGICDTEEELREIVANGLKLSPVTQCLIEESIAGYK 214

Query: 181  EIEYEVMRDSADNALVVCNMENFDPVGIHTGDSIVFAPTQTLSDIENQMLRDASLKIIRA 240
             EIEYEVMRDSADNA+VVCNMENFDPVG+HTGDSIVFAP+QTLSD E QMLRDASL IIRA
Sbjct: 215  EIEYEVMRDSADNAIVVCNMENFDPVGVHTGDSIVFAPSQTLSDNEYQMLRDASLNIIRA 274

Query: 241  LKIEGGCNVQLALDPYSFKYYVIEVNPRVSRSSALASKATGYPIAKLAAKIAVGLTLDEM 300
             LKIEGGCNVQLALDP S++Y VIEVNPRVSRSSALASKATGYPIAK++AKIA+G+TLDE+
Sbjct: 275  LKIEGGCNVQLALDPNSYEYRVIEVNPRVSRSSALASKATGYPIAKMSAKIAIGMTLDEI 334

Query: 301  INPITGTTYAMFEPALDYVVAKIPRFPFDKFEHGERQLGTQMKATGEVMAIGRNLEESLL 360
             INP+T  TYAMFEPALDYVVAKI RFPFDKFE+G+R LGTQMKATGEVMAIGRN+EESLL
Sbjct: 335  INPVTNKTYAMFEPALDYVVAKIARFPFDKFENGDRHLGTQMKATGEVMAIGRNIEESLL 394

Query: 361  KACRSLEIGVCHNEMTSLSNISDEELVTKVIKAQDDRLFYLSEAIRRGYSIEELESLTKI 420
             KA RSLEIGV HNEMT       DE+L  K++K QDDRLFY+SEAIRRG  IEE+  LTKI
Sbjct: 395  KAVRSLEIGVFHNEMTEAIEADDEKLYEKMVKTQDDRLFYVSEAIRRGIPIEEIADLTKI 454

Query: 421  DLFFLDKLLHIVEIEQELQMHVDHLESLKKAKRYGFSDQKIAEIWQKDESDIRAMRHSHS 480
             D+FFLDKLL+IVEIE +L++++   E LK AK+ GFSD++IA++W     ++R  R  +
Sbjct: 455  DIFFLDKLLYIVEIENQLKVNIFEPELLKTAKKNGFSDREIAKLWNVTPEEVRRRRQENK 514

Query: 481  LYPVYKMVDTCAAEFDAKTPYFYSTYELENESVQSNKESILVLGSGPIRIGQGVEFDYAT 540
             + PVYKMVDTCAAEF++ TPYFYSTYE ENES +S+KE I+VLGSGPIRIGQGVEFDYAT
Sbjct: 515  IIPVYKMVDTCAAEFESSTPYFYSTYEWENESKRSDKEKIIVLGSGPIRIGQGVEFDYAT 574

Query: 541  VHSVKAIQKAGYEAIIMNSNPETVSTDFSVSDKLYFEPLTFEDVMNVIDLEQPKGVIVQF 600
             VH VKAIQ  G EAI++NSNPETVSTDFS+SDKLYFEPLTFEDVMNVIDLE+P  VIVQF
Sbjct: 575  VHCVKAIQALGKEAIVINSNPETVSTDFSISDKLYFEPLTFEDVMNVIDLEEPLVVIVQF 634

Query: 601  GGQTAINLAQALSEAGVTILGTQVEDLDRAEDRDLFEKALKELGIPQPQGQTATNEEEAL 660
             GGQTAINLA+ LS+AGV ILGTQVEDLDRAEDRDLFEKAL++L IPQP G TATNEEEA+
Sbjct: 635  GGQTAINLAEHLSKAGVKILGTQVEDLDRAEDRDLFEKALQDLDIPQPPGATATNEEEAV 694

Query: 661  EAAKKIGFPVLVRPSYVLGGRAMEIVENKEDLREYIRTAVKASPEHPILVDSYIFGKECE 720
              A KIG+PVL+RPS+VLGGRAMEI+ N++DLR+Y+  AVKASPEHP+LVDSY+ G+ECE
Sbjct: 695  ANANKIGYPVLIRPSFVLGGRAMEIINNEKDLRDYMNRAVKASPEHPVLVDSYLQGQECE 754

Query: 721  VDAISDGKSVLIPGIMEHIERAGVHSGDSMAVYPPQQLSKQIQETIAEYTKRLAIGLNCI 780
```

```
              VDAI DGK VL+PGIMEHIERAGVHSGDSMAVYPPQ LS+ I +TI +YTKRLAIGLNCI
Sbjct: 755   VDAICDGKEVLLPGIMEHIERAGVHSGDSMAVYPPQNLSQAIIDTIVDYTKRLAIGLNCI 814

Query: 781   GMMNVQFVIKNEQVYVIEVNPRASRTVPFLSKVTGIPMAQIATKLILGQTLKDLGYEDGL 840
              GMMN+QFVI  EQVYVIEVNPRASRTVPFLSKVT IPMAQ+AT++ILG+ LKDLGYE GL
Sbjct: 815   GMMNIQFVIYEEQVYVIEVNPRASRTVPFLSKVTNIPMAQLATQMILGENLKDLGYEAGL 874

Query: 841   YPQSPLVHIKAPVFSFTKLAQVDSLLGPEMKSTGEVMGSDTSLEKALYKAFEANNSHLSE 900
               P   +VH+KAPVFSFTKLA+VDSLLGPEMKSTG  MGSD +LEKALYK+FEA   H+++
Sbjct: 875   APTPDMVHVKAPVFSFTKLAKVDSLLGPEMKSTGLAMGSDVTLEKALYKSFEAAKLHMAD 934

Query: 901   FGQIVFTIADDSKAEALSLARRFKAIGYQIMATQGTAAYFAEQGLSACLVGKIGDAANDI 960
              +G ++FT+AD+ K E L+LA+ F  IGY ++AT GTAA+  E GL    V K+    ++
Sbjct: 935   YGSVLFTVADEDKEETLALAKDFAEIGYSLVATAGTAAFLKENGLYVREVEKLAGGEDEE 994

Query: 961   PTLV---RHGHVQAIVNTVGIKR---TADKDGQMIRSSAIEQGVPLFTALDTAKAMLTVL 1014
               TLV   R G VQA+VNT+G  R   T   DG  IR  AI +G+PLFT+LDT  A+L V+
Sbjct: 995   GTLVEDIRQGRVQAVVNTMGNTRASLTTATDGFRIRQEAISRGIPLFTSLDTVAAILKVM 1054

Query: 1015  ESRCFNIEAI 1024
              +SR F  + I
Sbjct: 1055  QSRSFTTKNI 1064
 Identities = 141/389 (36%), Positives = 222/389 (56%), Gaps = 16/389 (4%)

Query: 518   ESILVLGSGPIRIGQGVEFDYATVHSVKAIQKAGYEAIIMNSNPETVSTDFSVSDKLYFE 577
              + I+++GSGPI IGQ  EFDYA   +  A+++ GYE +++NSNP T+ TD  ++D +Y E
Sbjct: 8     KKIMIIGSGPIIIGQAAEFDYAGTEACLALKEEGYEVVLVNSNPATIMTDREIADTVYIE 67

Query: 578   PLTFEDVMNVIDLEQPKGVIVQFGGQTAINLAQALSEAG------VTILGTQVEDLDRAE 631
              P+T E V  ++  E+P  ++   GGQT +N+A  LS+ G       V +LGT++  +D+AE
Sbjct: 68    PITLEFVSKILRKERPDALLPTLGGQTGLNMAMELSKTGILEELNVELLGTKLSAIDQAE 127

Query: 632   DRDLFEKALKELGIPQPQGQTATNEEEALEAAKKIGFPVLVRPSYVLGGRAMEIVENKED 691
              DR+LF++  +  +  P       AT  EEA+  A KIG+P++V P++ +GG    I  +E+
Sbjct: 128   DRELFKELCESINEPLCASDIATTVEEAINIADKIGYPIIVGPAFTMGGTGGGICDTEEE 187

Query: 692   LREYIRTAVKASPEHPILVDSYIFG-KECEVDAISD-GKSVLIPGIMEHIERAGVHSGDS 749
              LRE +   +K SP    L++  I G KE E + + D   + ++   ME+ +  GVH+GDS
Sbjct: 188   LREIVANGLKLSPVTQCLIEESIAGYKEIEYEVMRDSADNAIVVCNMENFDPVGVHTGDS 247

Query: 750   MAVYPPQQLSKQIQETIAEYTKRLAIGLNCIGMMNVQFVI--KNEQVYVIEVNPRASRTV 807
              +   P Q LS    + + + +   L   G  NVQ +   + +  VIEVNPR SR+
Sbjct: 248   IVFAPSQTLSDNEYQMLRDASLNIIRALKIEGGCNVQLALDPNSYEYRVIEVNPRVSRSS 307

Query: 808   PFLSKVTGIPMAQIATKLILGQTLKDL--GYEDGLY----PQSPLVHIKAPVFSFTKLAQ 861
                SK TG P+A+++ K+ +G TL ++     +  Y    P   V  K   F F K
Sbjct: 308   ALASKATGYPIAKMSAKIAIGMTLDEIINPVTNKTYAMFEPALDYVVAKIARFPFDKFEN 367

Query: 862   VDSLLGPEMKSTGEVMGSDTSLEKALYKA 890
               D  LG +MK+TGEVM    ++E++L KA
Sbjct: 368   GDRHLGTQMKATGEVMAIGRNIEESLLKA 396
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 777/1025 (75%), Positives = 896/1025 (86%), Gaps = 1/1025 (0%)

Query: 35    LSLKEEGYQVVLVNSNPATIMTDKDIADKVYIEPITLEFVTRILRKERPDALLPTLGGQT 94
              L+LKEEGY+V+LVNSNPATIMTDK+IADKVYIEP+TLEFV RI+RKERPDA+LPTLGGQT
Sbjct: 1     LALKEEGYKVILVNSNPATIMTDKEIADKVYIEPLTLEFVNRIIRKERPDAILPTLGGQT 60

Query: 95    GLNMAMALSKNGILEELNVELLGTKLSAIDKAEDRDLFKQLMEELNQPIPESEIVNSVEE 154
              GLNMAMALSK GIL++L +ELLGTKLSAID+AEDRDLFKQLM+EL+QPIPES IV +V+E
Sbjct: 61    GLNMAMALSKAGILDDLEIELLGTKLSAIDQAEDRDLFKQLMQELDQPIPESTIVKTVDE 120

Query: 155   AIQFAEQIGYPLIVRPAFTLGGTGGGMCDNQEQLVDITTKGLKLSPVTQCLIERSIAGFK 214
              A+ FA  IGYP+IVRPAFTLGGTGGG+C ++E+L +IT  GLKLSPVTQCLIERSIAGFK
Sbjct: 121   AVTFARDIGYPVIVRPAFTLGGTGGGICSSEEELCEITENGLKLSPVTQCLIERSIAGFK 180
```

```
Query: 215  EIEYEVMRDAADNALVVCNMENFDPVGIHTGDSIVFAPAQTLSDVENQLLRDASLDIIRA 274
            EIEYEVMRD+ADNALVVCNMENFDPVGIHTGDSIVFAP QTLSD+ENQ+LRDASL IIRA
Sbjct: 181  EIEYEVMRDSADNALVVCNMENFDPVGIHTGDSIVFAPTQTLSDIENQMLRDASLKIIRA 240

Query: 275  LKIEGGCNVQLALDPNSFKYYVIEVNPRVSRSSALASKATGYPIAKLAAKIAVGLTLDEV 334
            LKIEGGCNVQLALDP SFKYYVIEVNPRVSRSSALASKATGYPIAKLAAKIAVGLTLDE+
Sbjct: 241  LKIEGGCNVQLALDPYSFKYYVIEVNPRVSRSSALASKATGYPIAKLAAKIAVGLTLDEM 300

Query: 335  INPITKTTYAMFEPALDYVVAKMPRFPFDKFESGDRKLGTQMKATGEVMAIGRNIEESLL 394
            INPIT TTYAMFEPALDYVVAK+PRFPFDKFE G+R+LGTQMKATGEVMAIGRN+EESLL
Sbjct: 301  INPITGTTYAMFEPALDYVVAKIPRFPFDKFEHGERQLGTQMKATGEVMAIGRNLEESLL 360

Query: 395  KACRSLEIGVDHIKIADLDNVSDDVLLEKIRKAEDDRLFYLAEALRRHYSIEKLASLTSI 454
            KACRSLEIGV H ++  L N+SD+ L+ K+ KA+DDRLFYL+EA+RR YSIE+L SLT I
Sbjct: 361  KACRSLEIGVCHNEMTSLSNISDEELVTKVIKAQDDRLFYLSEAIRRGYSIEELESLTKI 420

Query: 455  DSFFLDKLRVIVELEDLLSKNRLDINILKKVKNKGFSDKAIASLWQINEDQVRNMRKEAG 514
            D FFLDKL  IVE+E  L +   +  LKK K  GFSD+ IA +WQ +E  +R MR
Sbjct: 421  DLFFLDKLLHIVEIEQELQMHVDHLESLKKAKRYGFSDQKIAEIWQKDESDIRAMRHSHS 480

Query: 515  ILPVYKMVDTCASEFDSATPYFYSTYAVENESLISDKASILVLGSGPIRIGQGVEFDYAT 574
            + PVYKMVDTCA+EFD+ TPYFYSTY +ENES+ S+K SILVLGSGPIRIGQGVEFDYAT
Sbjct: 481  LYPVYKMVDTCAAEFDAKTPYFYSTYELENESVQSNKESILVLGSGPIRIGQGVEFDYAT 540

Query: 575  VHSVKAIRESGFEAIIMNSNPETVSTDFSISDKLYFEPLTFEDVMNVIDLEKPEGVILQF 634
            VHSVKAI+++G+EAIIMNSNPETVSTDFS+SDKLYFEPLTFEDVMNVIDLE+P+GVI+QF
Sbjct: 541  VHSVKAIQKAGYEAIIMNSNPETVSTDFSVSDKLYFEPLTFEDVMNVIDLEQPKGVIVQF 600

Query: 635  GGQTAINLAKDLNKAGVKILGTQLEDLDRAENRKQFEATLQALNIPQPPGFTATTEEEAV 694
            GGQTAINLA+ L++AGV ILGTQ+EDLDRAE+R  FE  L+ L IPQP G TAT EEEA+
Sbjct: 601  GGQTAINLAQALSEAGVTILGTQVEDLDRAEDRDLFEKALKELGIPQPQGQTATNEEEAL 660

Query: 695  NAAQKIGYPVLVRPSYVLGGRAMKIVENEEDLRHYMTTAVKASPDHPVLIDAYLIGKECE 754
             AA+KIG+PVLVRPSYVLGGRAM+IVEN+EDLR Y+ TAVKASP+HP+L+D+Y+ GKECE
Sbjct: 661  EAAKKIGFPVLVRPSYVLGGRAMEIVENKEDLREYIRTAVKASPEHPILVDSYIFGKECE 720

Query: 755  VDAISDGQNILIPGIMEHIERSGVHSGDSMAVYPPQTLSETIIETIVDYTKRLAIGLNCI 814
            VDAISDG+++LIPGIMEHIER+GVHSGDSMAVYPPQ LS+ I ETI +YTKRLAIGLNCI
Sbjct: 721  VDAISDGKSVLIPGIMEHIERAGVHSGDSMAVYPPQQLSKQIQETIAEYTKRLAIGLNCI 780

Query: 815  GMMNIQFVIKDQKVYVIEVNPRASRTLPFLSKVTHIPMAQVATKVILGDKLCNFTYGYDL 874
            GMMN+QFVIK+++VYVIEVNPRASRT+PFLSKVT IPMAQ+ATK+ILG  L +  Y  L
Sbjct: 781  GMMNVQFVIKNEQVYVIEVNPRASRTVPFLSKVTGIPMAQIATKLILGQTLKDLGYEDGL 840

Query: 875  YPASDMVHIKAPVFSFTKLAKVDSLLGPEMKSTGEVMGSDINLQKALYKAFEAAYLHMPD 934
            YP S +VHIKAPVFSFTKLA+VDSLLGPEMKSTGEVMGSD +L+KALYKAFEA   H+ +
Sbjct: 841  YPQSPLVHIKAPVFSFTKLAQVDSLLGPEMKSTGEVMGSDTSLEKALYKAFEANNSHLSE 900

Query: 935  YGNIVFTVDDTDKEEALELAKVYQSIGYRIYATQGTAIYFDANGLETVLVGKLGENDRNH 994
            +G IVFT+ D  K EAL LA+ +++IGY+I ATQGTA YF   GL  LVGK+G+   N
Sbjct: 901  FGQIVFTIADDSKAEALSLARRFKAIGYQIMATQGTAAYFAEQGLSACLVGKIGD-AAND 959

Query: 995  IPDLIKNGKIQAVINTVGQNNIDNHDALIIRRSAIEQGVPLFTSLDTAHAMFKVLESRAF 1054
            IP L+++G +QA++NTVG     + D  +IR SAIEQGVPLFT+LDTA AM  VLESR F
Sbjct: 960  IPTLVRHGHVQAIVNTVGIKRTADKDGQMIRSSAIEQGVPLFTALDTAKAMLTVLESRCF 1019

Query: 1055 TLKVL 1059
              ++ +
Sbjct: 1020 NIEAI 1024
 Identities = 145/387 (37%), Positives = 229/387 (58%), Gaps = 16/387 (4%)

Query: 10   IMVIGSGPIVIGQAAEFDYSGTQACLSLKEEGYQVVLVNSNPATIMTDKDIADKVYIEPI 69
             I+V+GSGPI IGQ  EFDY+   +  ++++ GY+ +++NSNP T+ TD  ++DK+Y EP+
Sbjct: 520  ILVLGSGPIRIGQGVEFDYATVHSVKAIQKAGYEAIIMNSNPETVSTDFSVSDKLYFEPL 579

Query: 70   TLEFVTRILRKERPDALLPTLGGQTGLNMAMALSKNGILEELNVELLGTKLSAIDKAEDR 129
             T E V  ++  E+P  ++   GGQT +N+A ALS+ G       V +LGT++  +D+AEDR
Sbjct: 580  TFEDVMNVIDLEQPKGVIVQFGGQTAINLAQALSEAG------VTILGTQVEDLDRAEDR 633
```

```
Query: 130 DLFKQLMEELNQPIPESEIVNSVEEAIQFAEQIGYPLIVRPAFTLGGTGGGMCDNQEQLV 189
           DLF++ ++EL  P P+ +   + EEA++ A++IG+P++VRP++ LGG    + +N+E L
Sbjct: 634 DLFEKALKELGIPQPQGQTATNEEEALEAAKKIGFPVLVRPSYVLGGRAMEIVENKEDLR 693

Query: 190 DITTKGLKLSPVTQCLIERSIAGFKEIEYEVMRDAADNALVVCNMENFDPVGIHTGDSIV 249
           +     +K SP    L++  I G KE E + + D   + L+   ME+ +  G+H+GDS+
Sbjct: 694 EYIRTAVKASPEHPILVDSYIFG-KECEVDAISD-GKSVLIPGIMEHIERAGVHSGDSMA 751

Query: 250 FAPAQTLSDVENQLLRDASLDIIRALKIEGGCNVQLALDPNSFKYYVIEVNPRVSRSSAL 309
             P Q LS    + + + +   L   G  NVQ +   + + YVIEVNPR SR+
Sbjct: 752 VYPPQQLSKQIQETIAEYTKRLAIGLNCIGMMNVQFVI--KNEQVYVIEVNPRASRTVPF 809

Query: 310 ASKATGYPIAKLAAKIAVGLTLDEVINPITKTTYAMFEPALDYVVAKMPRFPFDKFESGD 369
            SK TG P+A++A K+ +G TL ++         Y    P    V  K P F F K    D
Sbjct: 810 LSKVTGIPMAQIATKLILGQTLKDL--GYEDGLY----PQSPLVHIKAPVFSFTKLAQVD 863

Query: 370 RKLGTQMKATGEVMAIGRNIEESLLKA 396
             LG +MK+TGEVM     ++E++L KA
Sbjct: 864 SLLGPEMKSTGEVMGSDTSLEKALYKA 890
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 659

A DNA sequence (GBSx0699) was identified in *S.agalactiae* <SEQ ID 2027> which encodes the amino acid sequence <SEQ ID 2028>. This protein is predicted to be carbamoyl phosphate synthetase small subunit (carA). Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2401(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 GENPEPT database:

```
>GP:CAB89872 GB:AJ132624 carbamoyl phosphate synthetase small
          subunit [Lactococcus lactis]
 Identities = 242/355 (68%), Positives = 305/355 (85%)

Query: 2   KRLLLLEDGSVFEGEAFGADVETSGEIVFSTGMTGYQESITDQSYNGQIITFTYPLIGNY 61
           KRLL+LEDG++FEGEA GA+++ +GE+VF+TGMTGYQESITDQSYNGQI+TFTYP++GNY
Sbjct: 3   KRLLILEDGTIFEGEALGANLDVTGELVFNTGMTGYQESITDQSYNGQILTFTYPIVGNY 62

Query: 62  GINRDDYESIRPTCKGVVIYEWAEYPSNWRQQMTLDEFLKLKGIPGISGIDTRALTKIIR 121
           G+NRDDYESI PTCK VV++E A  PSNWR QM+ DEFLK K IPGI+G+DTRA+TKI+R
Sbjct: 63  GVNRDDYESIHPTCKAVVVHEAARRPSNWRMQMSFDEFLKSKNIPGITGVDTRAITKIVR 122

Query: 122 KHGTMKACLINEGNSIHEALENLQKSVLLNDQIEQVSTKLAYASPGVGKNIVLVDFGLKH 181
           +HGTMKA L+   + +   + LQ +VL  +Q+E  ST  AY SP  G+ +V+VDFGLKH
Sbjct: 123 EHGTMKASLVQARDEVDHQMSQLQATVLPTNQVETSSTATAYPSPNTGRKVVVVDFGLKH 182

Query: 182 SILRELSQRQCHITVVPHTTTAQEILNLNPDGVLLSNGPGNPEQLPNALQMIQEIQGKIP 241
           SILRELS+R+C++TVVP+ T+A+EIL + PDGV+L+NGPG+P  +P A++MI+E+QGKIP
Sbjct: 183 SILRELSKRECNLTVVPYNTSAKEILEMEPDGVMLTNGPGDPTDVPEAIEMIKEVQGKIP 242

Query: 242 IFGICMGHQLFAKANGAKTYKMTFGHRGFNHAVRHLQTGQVDFTSQNHGYAVSREDFPEA 301
           IFGIC+GHQLF+ ANGA TYKM FGHRGFNHAVR + TG++DFTSQNHGYAVS E+ PE
Sbjct: 243 IFGICLGHQLFSLANGATTYKMKFGHRGFNHAVREVATGRIDFTSQNHGYAVSSENLPED 302

Query: 302 LFITHEEINDKTVEGVRHKYYPAFSVQFHPDAAPGPHDTSYLFDEFINMIDDFQQ 356
```

```
           L ITH EIND +VEGVRHKY+PAFSVQFHPDAAPGPHD SYLFD+F++++D+F++
Sbjct: 303 LMITHVEINDNSVEGVRHKYFPAFSVQFHPDAAPGPHDASYLFDDFMDLMDNFKK 357
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2029> which encodes the amino acid sequence <SEQ ID 2030>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3534(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 265/354 (74%), Positives = 309/354 (86%)

Query: 2   KRLLLLEDGSVFEGEAFGADVETSGEIVFSTGMTGYQESITDQSYNGQIITFTYPLIGNY 61
           KRLL+LEDG++FEGE FGAD++ +GEIVF+TGMTGYQESITDQSYNGQI+TFTYPLIGNY
Sbjct: 3   KRLLILEDGTIFEGEPFGADIDVTGEIVFNTGMTGYQESITDQSYNGQILTFTYPLIGNY 62

Query: 62  GINRDDYESIRPTCKGVVIYEWAEYPSNWRQQMTLDEFLKLKGIPGISGIDTRALTKIIR 121
           GINRDDYESI PTCKGVV+ E +   SNWR+QMTLD FLK+KGIPGISGIDTRALTKIIR
Sbjct: 63  GINRDDYESISPTCKGVVVSEVSRLASNWRKQMTLDAFLKIKGIPGISGIDTRALTKIIR 122

Query: 122 KHGTMKACLINEGNSIHEALENLQKSVLLNDQIEQVSTKLAYASPGVGKNIVLVDFGLKH 181
           +HGTMKA + ++G+SI    + L+ +VL  + IEQVSTK AY +PG+GKNIVLVDFGLKH
Sbjct: 123 QHGTMKATMADDGDSIQHLKDQLRATVLPTNTIEQVSTKTAYPAPGIGKNIVLVDFGLKH 182

Query: 182 SILRELSQRQCHITVVPHTTTAQEILNLNPDGVLLSNGPGNPEQLPNALQMIQEIQGKIP 241
           SILRE S+RQC+ITVVP   TA+E+L LNPDG++LSNGPGNPE LP AL MI+ +QGKIP
Sbjct: 183 SILREFSKRQCNITVVPFNITAEEVLQLNPDGLMLSNGPGNPEDLPEALDMIRGVQGKIP 242

Query: 242 IFGICMGHQLFAKANGAKTYKMTFGHRGFNHAVRHLQTGQVDFTSQNHGYAVSREDFPEA 301
           IFGICMGHQLF+ ANGAKT KMTFGHRGFNHAVR + TG++DFTSQNHGYAV R   P+
Sbjct: 243 IFGICMGHQLFSLANGAKTCKMTFGHRGFNHAVREIATGRIDFTSQNHGYAVERSSLPDT 302

Query: 302 LFITHEEINDKTVEGVRHKYYPAFSVQFHPDAAPGPHDTSYLFDEFINMIDDFQ 355
           L +THE+INDKTVEGV+H+ +PAFSVQFHPDAAPGPHD SYLFDEF+ MID ++
Sbjct: 303 LMVTHEDINDKTVEGVKHRDFPAFSVQFHPDAAPGPHDASYLFDEFLEMIDSWR 356
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 660

A DNA sequence (GBSx0700) was identified in *S.agalactiae* <SEQ ID 2031> which encodes the amino acid sequence <SEQ ID 2032>. This protein is predicted to be aspartate carbamoyltransferase (pyrB). Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3260(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 GENPEPT database:

```
>GP:AAF72727 GB:AF264709 aspartate transcarbamoylase [Enterococcus
          faecalis]
 Identities = 197/303 (65%), Positives = 250/303 (82%)
```

```
Query: 5   TQTLSLEHFVSLEELSNQEVMSLIKRSIEVKENPSNIGFDKDYYVSNLFFENSTRTHKSF 64
           ++ +SL+H ++ E L+++EVM LI+R+ E K+        ++ Y+ +NLFFENSTRTHKSF
Sbjct: 5   SERISLKHLLTAEALTDREVMGLIRRAGEFKQGAKWHPEERQYFATNLFFENSTRTHKSF 64

Query: 65  EMAELKLGLKTIEFNADTSSVNKGETLYDTILTMSALGLDVCVIRHPDIDYYKELIASPN 124
           E+AE KLGL+ IEF A  SSV KGETLYDT+LTMSA+G+DV VIRH   +YY ELI S
Sbjct: 65  EVAEKKLGLEVIEFEASRSSVQKGETLYDTVLTMSAIGVDVAVIRHGKENYYDELIQSKT 124

Query: 125 IHSAIVNGGDGSGQHPSQSLLDLVTIYEEFGYFKGLKIAIVGDLTHSRVAKSNMQVLKRL 184
           I  +I+NGGDGSGQHP+Q LLDL+TIYEEFG F+GLK+AIVGD+THSRVAKSNMQ+L RL
Sbjct: 125 IQCSIINGGDGSGQHPTQCLLDLMTIYEEFGGFEGLKVAIVGDITHSRVAKSNMQLLNRL 184

Query: 185 GAEIFFSGPKEWYSSQFDEYGQYLPIDQLVDQIDVLMLLRVQHERHDGKGVFSKESYHQQ 244
           GAEI+FSGP+EWY  QFD YGQY+P+D++V+++DV+MLLRVQHERHDGK  FSKE YH +
Sbjct: 185 GAEIYFSGPEEWYDHQFDVYGQYVPLDEIVEKVDVMMLLRVQHERHDGKESFSKEGYHLE 244

Query: 245 FGLTKERYKHLRDTAIIMHPAPVNRDVEIASDLVEADKARIVKQMSNGVYARIAILEAVL 304
           +GLT ER   L+  AIIMHPAPVNRDVE+A +LVE+ ++RIV QMSNGV+ R+AILEA+L
Sbjct: 245 YGLTNERATRLQKHAIIMHPAPVNRDVELADELVESLQSRIVAQMSNGVFMRMAILEAIL 304

Query: 305 NSR 307
           + +
Sbjct: 305 HGK 307
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2033> which encodes the amino acid sequence <SEQ ID 2034>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> 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>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 208/300 (69%), Positives = 249/300 (82%)

Query: 8   LSLEHFVSLEELSNQEVMSLIKRSIEVKENPSNIGFDKDYYVSNLFFENSTRTHKSFEMA 67
           ++L + VS+E L+ +EV+ LI R  E K     I   +   V+NLFFENSTRTHKSFE+A
Sbjct: 26  VALTNLVSMEALTTEEVLGLINRGSEYKAGKVVISDHQKDLVANLFFENSTRTHKSFEVA 85

Query: 68  ELKLGLKTIEFNADTSSVNKGETLYDTILTMSALGLDVCVIRHPDIDYYKELIASPNIHS 127
           E KLGL  ++FNAD S+VNKGE+LYDT+LTMSALG D+CVIRHP+ DYYKEL+ SP I +
Sbjct: 86  EKKLGLTVLDFNADASAVNKGESLYDTVLTMSALGTDICVIRHPEDDYYKELVESPTITA 145

Query: 128 AIVNGGDGSGQHPSQSLLDLVTIYEEFGYFKGLKIAIVGDLTHSRVAKSNMQVLKRLGAE 187
           +IVNGGDGSGQHPSQ LLDL+TIYEEFG F+GLKIAI GDLTHSRVAKSNMQ+LKRLGAE
Sbjct: 146 SIVNGGDGSGQHPSQCLLDLLTIYEEFGRFEGLKIAIAGDLTHSRVAKSNMQILKRLGAE 205

Query: 188 IFFSGPKEWYSSQFDEYGQYLPIDQLVDQIDVLMLLRVQHERHDGKGVFSKESYHQQFGL 247
           ++F GP+EWYS  F+ YG Y+ IDQ++ ++DVLMLLRVQHERHDG   FSKE YHQ FGL
Sbjct: 206 LYFYGPEEWYSEAFNAYGTYIAIDQIIKELDVLMLLRVQHERHDGHQSFSKEGYHQAFGL 265

Query: 248 TKERYKHLRDTAIIMHPAPVNRDVEIASDLVEADKARIVKQMSNGVYARIAILEAVLNSR 307
           T+ERY+ L+D+AIIMHPAPVNRDVEIA  LVEA KARIV QM+NGV+ R+AI+EA+LN R
Sbjct: 266 TQERYQQLKDSAIIMHPAPVNRDVEIADSLVEAPKARIVSQMANGVFVRMAIIEAILNGR 325
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 661

A DNA sequence (GBSx0701) was identified in *S.agalactiae* <SEQ ID 2035> which encodes the amino acid sequence <SEQ ID 2036>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2392(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 GENPEPT database:

```
>GP:AAC06948 GB:AE000708 dihydroorotase [Aquifex aeolicus]
 Identities = 176/422 (41%), Positives = 255/422 (59%), Gaps = 8/422 (1%)

Query: 11  IIKNGLIIDPQSGFNQVSDMLIDQGKIKQISKEIDIKGIPIIDASNKIVAPGLVDIHVHF 70
           I+KNG +IDP         D+L++ GKIK+I K I +     IIDA   IV PG +DIHVH
Sbjct: 5   IVKNGYVIDPSQNLEGEFDILVENGKIKKIDKNILVPEAEIIDAKGLIVCPGFIDIHVHL 64

Query: 71  REPGQTHKENIHTGALSAAVGGFTTVLMMANTNPTISSPEIVKQVKESAAKEAI-KIETV 129
           R+PGQT+KE+I +G+  A  GGFTT++ M NTNP I +  +V  + + +     + ++
Sbjct: 65  RDPGQTYKEDIESGSRCAVAGGFTTIVCMPNTNPPIDNTTVVNYILQKSKSVGLCRVLPT 124

Query: 130 ATITKSLNGKDLVNFEELLEAGVAGFSDDGIPLTDTKVLQEAMNLARKHDVVLSLHEEDP 189
            TITK   GK++ +F  L EAG   F+DDG P+ D+ V+++A+ LA +  V +  H ED
Sbjct: 125 GTITKGRKGKEIADFYSLKEAGCVAFTDDGSPVMDSSVMRKALELASQLGVPIMDHCEDD 184

Query: 190 SLN-GVLGINEHIAQKIYHVCGASGLAEYSMIARDAMIAYQTQAKVHIQHLSSSESVEVV 248
            L  GV  INE     +  +   + AE   IARD ++A +T   VHIQH+S+  S+E++
Sbjct: 185 KLAYGV--INEGEVSALLGLSSRAPEAEEIQIARDGILAQRTGGHVHIQHVSTKLSLEII 242

Query: 249 DFAQKLGANLTAEVTPQHFSKTENLLLTKGANAKLNPPLRLEKDRQALIDGLKSGVISII 308
           +F ++ G  +T EV P H   TE  +L  GANA++NPPLR ++DR ALI+G+K G+I
Sbjct: 243 EFFKEKGVKITCEVNPNHLLFTEREVLNSGANARVNPPLRKKEDRLALIEGVKRGIIDCF 302

Query: 309 ASDHAPHHIMEKAADNISQAPSGMTGLETSLALGITYLVSTKELSMIDFLAKMTCNPAQL 368
           A+DHAPH   EK  + +  A  G+ GL+T+L    +  L      +S+   +  T NPA++
Sbjct: 303 ATDHAPHQTFEK--ELVEFAMPGIIGLQTALPSALE-LYRKGIISLKKLIEMFTINPARI 359

Query: 369 YGFDAGYLREGGPADIVIFDQAEERIIKAEF-ASKSSNSPFIGDKLKGVIHYTICNGEIV 427
            G D G L+ G PADI IFD  +E I+  E   SKS N+P  G  LKG + YTI +G++V
Sbjct: 360 IGVDLGTLKLGSPADITIFDPNKEWILNEETNLSKSRNTPLWGKVLKGKVIYTIKDGKMV 419

Query: 428 YQ 429
           Y+
Sbjct: 420 YK 421
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2037> which encodes the amino acid sequence <SEQ ID 2038>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have no N-terminal signal sequence
     INTEGRAL    Likelihood = -0.80   Transmembrane   76 -  92 ( 76 -  92)
     INTEGRAL    Likelihood = -0.00   Transmembrane  286 - 302 ( 286 - 302)

----- Final Results -----
             bacterial membrane --- Certainty= 0.132(Affirmative) < succ>
              bacterial outside --- Certainty= 0.000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
!GB:AE000708 dihydroorotase [Aquifex aeolicus]      316  3e-85

>GP:AAC06948 GB:AE000708 dihydroorotase [Aquifex aeolicus]

 Score =  316 bits (801), Expect = 3e-85
 Identities = 177/422 (41%), Positives = 254/422 (59%), Gaps = 8/422 (1%)

Query: 2   ILIKNGRVMDPKSQRDQVADVLIDGKQIVKIASAIECQEAQVIDASGLIVAPGLVDIHVH 61
           +++KNG V+DP    +   D+L++  +I KI   I   EA++IDA GLIV PG +DIHVH
Sbjct: 4   LIVKNGYVIDPSQNLEGEFDILVENGKIKKIDKNILVPEAEIIDAKGLIVCPGFIDIHVH 63

Query: 62  FREPGQTHKEDIHTGALAAAAGGVTTVVMMANTNPVISDVETLQEVLASAAKEKI-HIYT 120
            R+PGQT+KEDI +G+  A AGG TT+V M NTNP I +   +L  +    + +
Sbjct: 64  LRDPGQTYKEDIESGSRCAVAGGFTTIVCMPNTNPPIDNTTVVNYILQKSKSVGLCRVLP 123

Query: 121 NASVTQAFNGKDVTDFKALLEAGAVSFSDDGIPLESSKVLKEAFDLANANQTFISLHEED 180
             ++T+   GK++ DF +L EAG V+F+DDG P+  S V+++A +LA+     I  H ED
Sbjct: 124 TGTITKGRKGKEIADFYSLKEAGCVAFTDDGSPVMDSSVMRKALELASQLGVPIMDHCED 183

Query: 181 PQL-NGVLGFNEGIAEEHFHFCGATGVAEYSMIARDVMIAYDRQAHVHIQHLSKAESVQV 239
            +L  GV+  NEG              AE   IARD ++A     HVHIQH+S   S+++
Sbjct: 184 DKLAYGVI--NEGEVSALLGLSSRAPEAEEIQIARDGILAQRTGGHVHIQHVSTKLSLEI 241

Query: 240 VAFAQQLGAKVTAEVSPQHFSTTEDLLLIAGTSAKMNPPLRTQRDRLAVIEGLKSGVITV 299
           + F ++ G K+T EV+P H   TE  +L +G +A++NPPLR + DRLA+IEG+K G+I
Sbjct: 242 IEFFKEKGVKITCEVNPNHLLFTEREVLNSGANARVNPPLRKKEDRLALIEGVKRGIIDC 301

Query: 300 IATDHAPHHKDEKTVDDMTKAPSGMTGLETSLSLGLTHLVEPGHLTLMSLLEKMTLNPAL 359
            ATDHAPH   EK + +   A  G+ GL+T+L   L L   G ++L  L+E  T+NPA
Sbjct: 302 FATDHAPHQTFEKELVEF--AMPGIIGLQTALPSAL-ELYRKGIISLKKLIEMFTINPAR 358

Query: 360 LYGFDAGYLAENGPADLVIFADKQERLITENF-ASKASNSPFIGNKLKGVVKYTIADGEV 418
           + G D G L    PAD+ IF   +E ++ E    SK+ N+P  G  LKG V YTI DG++
Sbjct: 359 IIGVDLGTLKLGSPADITIFDPNKEWILNEETNLSKSRNTPLWGKVLKGKVIYTIKDGKM 418

Query: 419 VY 420
           VY
Sbjct: 419 VY 420
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 269/420 (64%), Positives = 338/420 (80%)

Query: 9   MYIIKNGLIIDPQSGFNQVSDMLIDQGKIKQISKEIDIKGIPIIDASNKIVAPGLVDIHV 68
           M +IKNG ++DP+S  +QV+D+LID  +I +I+  I+ +   +IDAS  IVAPGLVDIHV
Sbjct: 1   MILIKNGRVMDPKSQRDQVADVLIDGKQIVKIASAIECQEAQVIDASGLIVAPGLVDIHV 60

Query: 69  HFREPGQTHKENIHTGALSAAVGGFTTVLMMANTNPTISSPEIVKQVKESAAKEAIKIET 128
           HFREPGQTHKE+IHTGAL+AA GG TTV+MMANTNP IS  E +++V  SAAKE I I T
Sbjct: 61  HFREPGQTHKEDIHTGALAAAAGGVTTVVMMANTNPVISDVETLQEVLASAAKEKIHIYT 120

Query: 129 VATITKSLNGKDLVNFEELLEAGVAGFSDDGIPLTDTKVLQEAMNLARKHDVVLSLHEED 188
            A++T++ NGKD+ +F+ LLEAG   FSDDGIPL  +KVL+EA +LA  +   +SLHEED
Sbjct: 121 NASVTQAFNGKDVTDFKALLEAGAVSFSDDGIPLESSKVLKEAFDLANANQTFISLHEED 180

Query: 189 PSLNGVLGINEHIAQKIYHVCGASGLAEYSMIARDAMIAYQTQAKVHIQHLSSSESVEVV 248
           P LNGVLG NE IA++ +H CGA+G+AEYSMIARD MIAY  QA VHIQHLS +ESV+VV
Sbjct: 181 PQLNGVLGFNEGIAEEHFHFCGATGVAEYSMIARDVMIAYDRQAHVHIQHLSKAESVQVV 240

Query: 249 DFAQKLGANLTAEVTPQHFSKTENLLLTKGANAKLNPPLRLEKDRQALIDGLKSGVISII 308
            FAQ+LGA +TAEV+PQHFS TE+LLL  G +AK+NPPLR ++DR A+I+GLKSGVI++I
Sbjct: 241 AFAQQLGAKVTAEVSPQHFSTTEDLLLIAGTSAKMNPPLRTQRDRLAVIEGLKSGVITVI 300

Query: 309 ASDHAPHHIMEKAADNISQAPSGMTGLETSLALGITYLVSTKELSMIDFLAKMTCNPAQL 368
           A+DHAPHH  EK  D++++APSGMTGLETSL+LG+T+LV     L+++  L KMT NPA L
Sbjct: 301 ATDHAPHHKDEKTVDDMTKAPSGMTGLETSLSLGLTHLVEPGHLTLMSLLEKMTLNPALL 360

Query: 369 YGFDAGYLREGGPADIVIFDQAEERIIKAEFASKSSNSPFIGDKLKGVIHYTICNGEIVY 428
```

```
            YGFDAGYL E GPAD+VIF   +ER+I   FASK+SNSPFIG+KLKGV+ YTI +GE+VY
Sbjct: 361 YGFDAGYLAENGPADLVIFADKQERLITENFASKASNSPFIGNKLKGVVKYTIADGEVVY 420
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 662

A DNA sequence (GBSx0702) was identified in *S.agalactiae* <SEQ ID 2039> which encodes the amino acid sequence <SEQ ID 2040>. This protein is predicted to be orotate phosphoribosyltransferase PyrE (pyrE). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2214(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 GENPEPT database:

```
>GP:AAC95453 GB:AF068902 orotate phosphoribosyltransferase PyrE
           [Streptococcus pneumoniae]
 Identities = 152/208 (73%), Positives = 180/208 (86%)

Query: 1   MDLARQIAMELLDIQAVYLRPQQPFTWASGVKSPIYTDNRVTLSYPETRTLIENGFVKQI 60
           M LA+ IA  LL IQAVYL+P++PFTWASG+KSPIYTDNRVTL+YPETRTLIENGFV  I
Sbjct: 1   MTLAKDIASHLLKIQAVYLKPEEPFTWASGIKSPIYTDNRVTLAYPETRTLIENGFVDAI 60

Query: 61  QKHFPNVDIIAGTATAGIPHGAIIADKMNLPFAYIRSKAKDHGVGNQIEGRVYSGQKMVI 120
           ++ FP V++IAGTATAGIPHGAIIADKMNLPFAYIRSK KDHG GNQIEGRV  GQKMV+
Sbjct: 61  KEAFPEVEVIAGTATAGIPHGAIIADKMNLPFAYIRSKPKDHGAGNQIEGRVAQGQKMVV 120

Query: 121 IEDLISTGGSVLEAVTAAQSQGIEVLGVVAIFTYQLAKAEQAFREADIPLVTLTDYNQLI 180
           +EDLISTGGSVLEAV AA+ +G +VLGVVAIF+YQL KA++ F +A + LVTL++Y++LI
Sbjct: 121 VEDLISTGGSVLEAVAAAKREGADVLGVVAIFSYQLPKADKNFADAGVKLVTLSNYSELI 180

Query: 181 KVAKVNGYITADQLVLLKKFKEDQMNWQ 208
            +A+  GYIT + L LLK+FKEDQ NWQ
Sbjct: 181 HLAQEEGYITPEGLDLLKRFKEDQENWQ 208
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2041> which encodes the amino acid sequence <SEQ ID 2042>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1612(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 158/208 (75%), Positives = 179/208 (85%)

Query: 1   MDLARQIAMELLDIQAVYLRPQQPFTWASGVKSPIYTDNRVTLSYPETRTLIENGFVKQI 60
           M LA QIA +LLDI+AVYL+P+ PFTWASG+KSPIYTDNRVTLSYP+TR LIENGFV+ I
Sbjct: 1   MTLASQIATQLLDIKAVYLKPEDPFTWASGIKSPIYTDNRVTLSYPKTRDLIENGFVETI 60

Query: 61  QKHFPNVDIIAGTATAGIPHGAIIADKMNLPFAYIRSKAKDHGVGNQIEGRVYSGQKMVI 120
           + HFP V++IAGTATAGIPHGAIIADKM LPFAYIRSK KDHG GNQIEGRV  GQKMVI
Sbjct: 61  KAHFPEVEVIAGTATAGIPHGAIIADKMTLPFAYIRSKPKDHGAGNQIEGRVLKGQKMVI 120
```

```
Query: 121 IEDLISTGGSVLEAVTAAQSQGIEVLGVVAIFTYQLAKAEQAFREADIPLVTLTDYNQLI 180
           IEDLISTGGSVL+A  AA  +G +VLGVVAIFTY+L KA Q F+EA I L+TL++Y +LI
Sbjct: 121 IEDLISTGGSVLDAAAAAASREGADVLGVVAIFTYELPKASQNFKEAGIKLITLSNYTELI 180

Query: 181 KVAKVNGYITADQLVLLKKFKEDQMNWQ 208
            VAK+ GYIT D L LKKFKEDQ+NWQ
Sbjct: 181 AVAKLQGYITNDGLHLLKKFKEDQVNWQ 208
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 663

A DNA sequence (GBSx0703) was identified in *S.agalactiae* <SEQ ID 2043> which encodes the amino acid sequence <SEQ ID 2044>. This protein is predicted to be orotidine 5'-phosphate decarboxylase (pyrF). Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> 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>
```

A related GBS nucleic acid sequence <SEQ ID 9829> which encodes amino acid sequence <SEQ ID 9830> was also identified.

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

```
>GP:AAC95452 GB:AF068902 orotidine-5'-decarboxylase PyrF
          [Streptococcus pneumoniae]
 Identities = 149/231 (64%), Positives = 176/231 (75%), Gaps = 1/231 (0%)

Query: 19  MLEKCPIIALDFSDLASVTTFLEHFPKEELLFVKIGMELYYSEGPSIIRYIKSLGHRIFL 78
           M E  PIIALDF    +V  FL  FP EE L++K+GMELYY+ GP I+ Y+K LGH +FL
Sbjct: 1   MREHRPIIALDFPSFEAVKEFLALFPAEESLYLKVGMELYYAAGPEIVSYLKGLGHSVFL 60

Query: 79  DLKLHDIPNTVRSSMSVLAKLGIDMTNVHAAGGVEMMKAAREGLGKGPILLAVTQLTSTS 138
           DLKLHDIPNTV+S+M VL++LG+DMTNVHAAGGVEMMKAAREGLG     L+AVTQLTSTS
Sbjct: 61  DLKLHDIPNTVKSAMKVLSQLGVDMTNVHAAGGVEMMKAAREGLGSQAKLIAVTQLTSTS 120

Query: 139 QEQMQVDQHINLSVVDSVCHYAQKAQEAGLDGVVASAQEGMQIKKQTNEHFICLTPGIRP 198
           + QMQ  Q+I  S+ +SV HYA+K  EAGLDGVV SAQE   IK+ TN  FICLTPGIRP
Sbjct: 121 EAQMQEFQNIQTSLQESVIHYAKKTAEAGLDGVVCSAQEVQVIKQATNPDFICLTPGIRP 180

Query: 199 PQTNQLDDQKRTMTPEQARIVGADYIVVGRPITKAENPYQAYLEIKEEWNR 249
                + DQKR MTP  A  +G+DYIVVGRPIT+AE+P  AY  IK+EW +
Sbjct: 181 AGV-AVGDQKRVMTPADAYQIGSDYIVVGRPITQAEDPVAAYHAIKDEWTQ 230
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2045> which encodes the amino acid sequence <SEQ ID 2046>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1934(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 149/229 (65%), Positives = 180/229 (78%), Gaps = 1/229 (0%)

Query: 19  MLEKCPIIALDFSDLASVTTFLEHFPKEELLFVKIGMELYYSEGPSIIRYIKSLGHRIFL 78
           M E+ PIIALDFS        FL+ FP EE L+VKIGMELYY++GP I+RYIKSLGH +FL
Sbjct: 1   MKEERPIIALDFSSFEETKAFLDLFPAEEKLYVKIGMELYYAQGPDIVRYIKSLGHNVFL 60

Query: 79  DLKLHDIPNTVRSSMSVLAKLGIDMTNVHAAGGVEMMKAAREGLGKGPILLAVTQLTSTS 138
           DLKLHDIPNTVR++M+VL +L IDM  VHAAGGVEM+KAAREGLG+GP L+AVTQLTSTS
Sbjct: 61  DLKLHDIPNTVRAAMAVLKELDIDMATVHAAGGVEMLKAAREGLGQGPTLIAVTQLTSTS 120

Query: 139 QEQMQVDQHINLSVVDSVCHYAQKAQEAGLDGVVASAQEGMQIKKQTNEHFICLTPGIRP 198
           ++QM+ DQ+I  S+++SV HY++ A +A LDG V SAQE   IK  T   F CLTPGIRP
Sbjct: 121 EDQMRGDQNIQTSLLESVLHYSKGAAKAQLDGAVCSAQEVEAIKAVTPTGFTCLTPGIRP 180

Query: 199 PQTNQLDDQKRTMTPEQARIVGADYIVVGRPITKAENPYQAYLEIKEEW 247
             +N + DQKR MTP QAR +G+DYIVVGRPIT+A++P  AY  IK EW
Sbjct: 181 KGSN-IGDQKRVMTPNQARRIGSDYIVVGRPITQAKDPVAAYQAIKAEW 228
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 664

A DNA sequence (GBSx0704) was identified in *S.agalactiae* <SEQ ID 2047> which encodes the amino acid sequence <SEQ ID 2048> in others. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.70   Transmembrane  192 - 208 ( 190 - 211)
   INTEGRAL    Likelihood = -7.64   Transmembrane  226 - 242 ( 218 - 250)
   INTEGRAL    Likelihood = -3.77   Transmembrane  388 - 404 ( 378 - 404)
   INTEGRAL    Likelihood = -3.08   Transmembrane  293 - 309 ( 292 - 311)
   INTEGRAL    Likelihood = -2.87   Transmembrane  165 - 181 ( 162 - 182)
   INTEGRAL    Likelihood = -2.13   Transmembrane  267 - 283 ( 267 - 284)
   INTEGRAL    Likelihood = -0.90   Transmembrane  114 - 130 ( 114 - 130)
   INTEGRAL    Likelihood = -0.75   Transmembrane  318 - 334 ( 318 - 334)
   INTEGRAL    Likelihood = -0.53   Transmembrane  140 - 156 ( 140 - 156)

----- Final Results -----
              bacterial membrane --- Certainty=0.4482(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 GENPEPT database:

```
>GP:BAB03800 GB:AP001507 unknown conserved protein in others
          [Bacillus halodurans]
 Identities = 63/243 (25%), Positives = 120/243 (48%)

Query: 5   MSVVLRAGKLLIESGAEVYRVEDTMKHFAKALQIENFEAYVVSSSIIASGINRYGKQEAK 64
           M + + AG++++ +GAE YRVE+T++  AKA Q  N  ++V ++ I  S         +
Sbjct: 8   MDICMLAGEIMLINGAETYRVEETLERMAKAGQFRNVHSFVTTTGIFLSFEEEGAGDVMQ 67

Query: 65  VCNTDGVTANLGRLEAVNNLSRQIAKQDLVSPEEIVKQLDLIEHQKDYSLLVTLISYFCG 124
           +   D    +L ++  VN +SR+    ++ + E + K  ++ +   +YS L+   +
Sbjct: 68  MIRVDDRMQDLNKVTLVNQVSREFVNGEIDAAEALTKLQNIAKQPMNYSPLLLHTASGVA 127

Query: 125 AGSFSLALGSSLLDSFSAAVTGLILGYFLNLMESRIHTGFLLTILGSSVVALSANLLYFS 184
            G+FS   G +L D+  A + G +    + ++S +   F    + +       A LL
Sbjct: 128 GGAFSYLFGGNLFDTLPAFIAGFVASMAVVHLQSYLKVRFFAEFMAAFTGGAVAILLVLI 187

Query: 185 GLGEHRSIIILGALMVMVPGAAFVNSVREFSQNNFSTGLALIMSALLICISISAGVAITI 244
           GLGE+   +I+G LM +VPG    N+VR+    +   G+       +  +SI+ G+A+ I
Sbjct: 188 GLGENVDQVIIGTLMPLVPGIPLTNAVRDLISGDLLAGVTRGAECFVTSLSIATGIALAI 247

Query: 245 EII 247
```

```
            ++
Sbjct: 248 ALL 250
```

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

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

Example 665

A DNA sequence (GBSx0705) was identified in *S.agalactiae* <SEQ ID 2049> which encodes the amino acid sequence <SEQ ID 2050>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.5134(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9353> which encodes amino acid sequence <SEQ ID 9354> was also identified.

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

```
>GP:CAB12571 GB:Z99108 similar to ABC transporter (ATP-binding
           protein) [Bacillus subtilis]
 Identities = 193/288 (67%), Positives = 231/288 (80%)

Query: 1   MNDVINIVYHVENQDLVRYSGDYTNFESVYAMKKAQLEAAYERQQKEIADLQDFVNRNKA 60
           +N VIN++YHVENQ+L RY GDY  F  VY +KK QLEAAY++QQ+E+A+L+DFV RNKA
Sbjct: 222 LNSVINLIYHVENQELTRYVGDYHQFMEVYEVKKQQLEAAYKKQQQEVAELKDFVARNKA 281

Query: 61  RVATRNMAMSRQKKLDKMDIIELQAEKPKPSFEFKESRTPGRFIFQAKDLQIGYDRALTK 120
           RV+TRNMAMSRQKKLDKMD+IEL AEKPKP F FK +RT G+ IF+ KDL IGYD  L++
Sbjct: 282 RVSTRNMAMSRQKKLDKMDMIELAAEKPKPEFHFKPARTSGKLIFETKDLVIGYDSPLSR 341

Query: 121 PLNLTFERNQKIAIVGANGIGKTTLLKSLLGIIPPISGNVERGDFIDLGYFEQEVPGGNR 180
           PLNL  ER QKIA+ GANGIGKTTLLKSLLG I P+ G+VERG+ I  GYFEQEV   N
Sbjct: 342 PLNLRMERGQKIALYGANGIGKTTLLKSLLGEIQPLEGSVERGEHIYTGYFEQEVKETNN 401

Query: 181 QTPLEAVWDAFPALNQAEVRAALARCGLTSKHIESQIQVLSGGEQSKVRFCLLMNRENNV 240
            T +E VW  FP+  Q E+RAA A+CGLT+KHIES++ VLSGGE++KVR C L+N E N+
Sbjct: 402 NTCIEEVWSEFPSYTQYEIRAAPAKCGLTTKHIESRVSVLSGGEKAKVRLCKLINSETNL 461

Query: 241 LVLDEPTNHLDVDAKDELKRALKAYKGSILMVCHEPDFYEGWMDDVWD 288
           LVLDEPTNHLD DAK+ELKRALK YKGSIL++ HEPDFY     + W+
Sbjct: 462 LVLDEPTNHLDADAKEELKRALKEYKGSILLISHEPDFYMDIATETWN 509
 Identities = 56/219 (25%), Positives = 97/219 (43%), Gaps = 44/219 (20%)

Query: 104 IFQAKDLQIGY-DRALTKPLNLTFERNQKIAIVGANGIGKTTLLKSLLGIIPPISGNVER 162
           I   KDL  G+ DRA+   ++    + + + ++GANG GK+T +  + G + P  G VE
Sbjct: 3   ILSVKDLSHGFGDRAIFNNVSFRLLKGEHVGLIGANGEGKSTFMNIITGKLEPDEGKVEW 62

Query: 163 GDFIDLGYFEQEVPGGNRQTPLEAVWDAFPALNQAE---------------------- 198
              + +GY +Q       ++  + + DAF  L   E
Sbjct: 63  SKNVRVGYLDQHTVLEKGKSIRDVLKDAFHYLFAMEEEMNEIYNKMGEADPDELEKLLEE 122

Query: 199 ---VRAALAR----------------CGLTSKHIESQIQVLSGGEQSKVRFCLLMNRENN 239
              ++ AL                  GL+   +E  +  LSGG+++KV     L+  +
Sbjct: 123 VGVIQDALTNNDFYVIDSKVEEIARGLGLSDIGLERDVTDLSGGQRTKVLLAKLLLEKPE 182
```

```
Query: 240 VLVLDEPTNHLDVDAKDELKRALKAYKGSILMVCHEPDF 278
          +L+LDEPTN+LD     + LKR L+ Y+ + +++ H+  F
Sbjct: 183 ILLLDEPTNYLDEQHIEWLKRYLQEYENAFILISHDIPF 221
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2051> which encodes the amino acid sequence <SEQ ID 2052>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2794(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 246/294 (83%), Positives = 274/294 (92%), Gaps = 1/294 (0%)

Query: 1   MNDVINIVYHVENQDLVRYSGDYTNFESVYAMKKAQLEAAYERQQKEIADLQDFVNRNKA 60
           +NDVINIVYHVENQ LVRY+GDY  F++VY MK++QLEAAYERQQKEIA+LQDFVNRNKA
Sbjct: 233 LNDVINIVYHVENQSLVRYTGDYYQFQAVYEMKQSQLEAAYERQQKEIANLQDFVNRNKA 292

Query: 61  RVATRNMAMSRQKKLDKMDIIELQAEKPKPSFEFKESRTPGRFIFQAKDLQIGYDRALTK 120
           RVATRNMAMSRQKKLDKMDIIELQAEKPKP+FEFK++RTP RFIFQ K+L IGYD  LTK
Sbjct: 293 RVATRNMAMSRQKKLDKMDIIELQAEKPKPNFEFKQARTPSRFIFQTKNLVIGYDYPLTK 352

Query: 121 -PLNLTFERNQKIAIVGANGIGKTTLLKSLLGIIPPISGNVERGDFIDLGYFEQEVPGGN 179
            PLN+TFERNQKIAIVGANGIGK+TLLKSLLG+I P+ G++  GDF+++GYFEQEV G N
Sbjct: 353 EPLNITFERNQKIAIVGANGIGKSTLLKSLLGVIEPLEGHIVTGDFLEVGYFEQEVTGVN 412

Query: 180 RQTPLEAVWDAFPALNQAEVRAALARCGLTSKHIESQIQVLSGGEQSKVRFCLLMNRENN 239
           RQTPLE VWDAFPALNQAEVRAALARCGLTSKHIESQIQVLSGGEQ+KVRFCLLMNRENN
Sbjct: 413 RQTPLEVVWDAFPALNQAEVRAALARCGLTSKHIESQIQVLSGGEQAKVRFCLLMNRENN 472

Query: 240 VLVLDEPTNHLDVDAKDELKRALKAYKGSILMVCHEPDFYEGWMDDVWDFNQLS 293
           VL+LDEPTNHLD+DAK+ELKRALKAYKGSILMVCHEPDFY GW+ D WDF++L+
Sbjct: 473 VLILDEPTNHLDIDAKNELKRALKAYKGSILMVCHEPDFYNGWVTDTWDFSKLT 526
 Identities = 60/218 (27%), Positives = 102/218 (46%), Gaps = 43/218 (19%)

Query: 104 IFQAKDLQIGY-DRALTKPLNLTFERNQKIAIVGANGIGKTTLLKSLLGIIPPISGNVER 162
           I + K L  G+ DRA+ + ++     + + I +VGANG GK+T +  + G + P  G VE
Sbjct: 15  ILEVKQLSHGFGDRAIFENVSFRLLKGEHIGLVGANGEGKSTFMSIVTGHLQPDEGKVEW 74

Query: 163 GDFIDLGYFEQEVPGGNRQTPLEAVWDAFPALNQAEVR-----AALA------------ 204
             ++  GY +Q     + QT  + +  AF  L + E R     A++A
Sbjct: 75  SKYVTAGYLDQHTVLESGQTVRDVLRTAFDELFKTENRINEIYASMADDKADIAVLMEEV 134

Query: 205 ----------------------RCGLTSKHIESQIQVLSGGEQSKVRFCLLMNRENNV 240
                                 G+     +ES +  LSGG+++KV     L+  + ++
Sbjct: 135 GELQDRLESRDFYTLDAKIDEVARALGVMDFGMESDVTSLSGGQRTKVLLAKLLLEKPDI 194

Query: 241 LVLDEPTNHLDVDAKDELKRALKAYKGSILMVCHEPDF 278
           L+LDEPTNHLD +  + LKR L+ Y+ + +++ H+  F
Sbjct: 195 LLLDEPTNHLDAEHIEWLKRYLQHYENAFVLISHDISF 232
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 666

A DNA sequence (GBSx0706) was identified in *S.agalactiae* <SEQ ID 2053> which encodes the amino acid sequence <SEQ ID 2054>. This protein is predicted to be lipoprotein Nlp1 precursor (pstS). Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2637(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 GENPEPT database:

```
>GP:CAB14429 GB:Z99116 alternate gene name: yzmB~similar to
          phosphate ABC transporter (binding protein) [Bacillus subtilis]
 Identities = 42/62 (67%), Positives = 49/62 (78%)

Query: 15  SITSVGSTALQPLVEAAADEFGKTNLGKTINVQGGGSGTGLSQVQSGAVQIGNSDLFAEE 74
           S+T  GS+A+QPLV AAA++F + N     I VQ GGSGTGLSQV  GAVQIGNSD+FAEE
Sbjct: 45  SLTISGSSAMQPLVLAAAEKFMEENPDADIQVQAGGSGTGLSQVSEGAVQIGNSDVFAEE 104

Query: 75  KE 76
           KE
Sbjct: 105 KE 106
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1695> which encodes the amino acid sequence <SEQ ID 1696>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> 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>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 63/74 (85%), Positives = 71/74 (95%)

Query: 3   LSGCANWIDKGQSITSVGSTALQPLVEAAADEFGKTNLGKTINVQGGGSGTGLSQVQSGA 62
           LS C++WIDKG+SIT+VGSTALQPLVEA ADEFG +NLGKT+NVQGGGSGTGLSQVQSGA
Sbjct: 20  LSACSSWIDKGESITAVGSTALQPLVEAVADEFGSSNLGKTVNVQGGGSGTGLSQVQSGA 79

Query: 63  VQIGNSDLFAEEKE 76
           VQIGNSD+FAEEK+
Sbjct: 80  VQIGNSDVFAEEKD 93
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 667

A DNA sequence (GBSx0707) was identified in *S.agalactiae* <SEQ ID 2055> which encodes the amino acid sequence <SEQ ID 2056>. This protein is predicted to be lipoprotein Nlp1 precursor (pstS). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> 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>
```

A related GBS nucleic acid sequence <SEQ ID 9343> which encodes amino acid sequence <SEQ ID 9344> was also identified.

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

```
>GP:CAB14429 GB:Z99116 alternate gene name: yzmB~similar to
          phosphate ABC transporter (binding protein) [Bacillus subtilis]
 Identities = 95/184 (51%), Positives = 126/184 (67%), Gaps = 1/184 (0%)

Query: 3   DHQVAVAGLAVIVNKKVNVKNLTTHQLRDIFAGKIKNWKEVGGQDLDISIINRAASSGSR 62
           DHQVAV G+A  VN    VK+++  +L+ IF GKIKNWKE+GG+D  I+++NR  SSG+R
Sbjct: 115 DHQVAVVGMAAAVNPDAGVKDISKDELKKIFTGKIKNWKELGGKDQKITLVNRPDSSGTR 174

Query: 63  ATFDNTIMGNVAPIQSQEQDSNGMVKSIVSQTPGAISYLAFAYV-DKSVGTLKLNGFAPT 121
           ATF    +   P +   +DS+  VK I++ TPGAI YLAF+Y+ D  V  L ++G  P
Sbjct: 175 ATFVKYALDGAEPAEGITEDSSNTVKKIIADTPGAIGYLAFSYLTDDKVTALSIDGVKPE 234

Query: 122 AKNVTTDNWKLWSYEHMYTKGNETGLTKEFLDYMKSDKVQSSIVQHMGYISINDMKVVKD 181
           AKNV T  + +W+Y+H YTKG  TGL KEFLDY+KS+ +Q SIV   GYI + DMKV +D
Sbjct: 235 AKNVATGEYPIWAYQHSYTKGEATGLAKEFLDYLKSEDIQKSIVTDQGYIPVTDMKVTRD 294

Query: 182 AEGK 185
           A GK
Sbjct: 295 ANGK 298
```

There is also homology to SEQ ID 1696.

SEQ ID 9344 (GBS659) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 135 (lane 2 & 3; MW 60kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 135 (lane 5-7; MW 35kDa) and in Figure 178 (lane 11; MW 35kDa).

GBS659-His was purified as shown in Figure 228, lane 6-8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 668

A DNA sequence (GBSx0708) was identified in *S.agalactiae* <SEQ ID 2057> which encodes the amino acid sequence <SEQ ID 2058>. This protein is predicted to be phosphate transporter permease PstC (pstC-2). Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-15.50   Transmembrane   35 -  51 (  27 -  61)
   INTEGRAL    Likelihood = -7.64   Transmembrane  167 - 183 ( 154 - 186)
   INTEGRAL    Likelihood = -6.37   Transmembrane  282 - 298 ( 277 - 302)
   INTEGRAL    Likelihood = -5.52   Transmembrane   85 - 101 (  81 - 116)
   INTEGRAL    Likelihood = -3.24   Transmembrane  133 - 149 ( 131 - 155)

----- Final Results -----
              bacterial membrane --- Certainty=0.7198(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8635> which encodes amino acid sequence <SEQ ID 8636> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
SRCFLG: 0
McG: Length of UR:  5
     Peak Value of UR:  -0.12
     Net Charge of CR: 2
McG: Discrim Score:    -16.22
GvH: Signal Score (-7.5): -4.26
     Possible site: 41
>>> Seems to have no N-terminal signal sequence
Amino Acid Composition: calculated from 1
ALOM program   count: 5 value: -15.50 threshold:  0.0
   INTEGRAL    Likelihood =-15.50   Transmembrane   29 -  45 (  21 -  55)
   INTEGRAL    Likelihood = -7.64   Transmembrane  161 - 177 ( 148 - 180)
   INTEGRAL    Likelihood = -6.37   Transmembrane  276 - 292 ( 271 - 296)
   INTEGRAL    Likelihood = -5.52   Transmembrane   79 -  95 (  75 - 110)
   INTEGRAL    Likelihood = -3.24   Transmembrane  127 - 143 ( 125 - 149)
   PERIPHERAL  Likelihood =  0.69     205
 modified ALOM score:   3.60
icm1 HYPID: 7  CFP: 0.720

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.7198(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 GENPEPT database:

```
>GP:CAB14428 GB:Z99116 alternate gene name: yzmC~similar to
          phosphate ABC transporter (permease) [Bacillus subtilis]
 Identities = 145/303 (47%), Positives = 209/303 (68%), Gaps = 4/303 (1%)

Query: 8   KNQELAKKLTSPSKNSRLEKFGKGITFLSLALIVFIVAM-ILIFVAQKGLSTFFVDGVKL 66
           +N  ++++L S  +N +L++    +   + ALI+   ++ I IF+  KGL +F V+GV
Sbjct: 6   ENMSVSERLISSRQNRQLDEVRGRMIVTACALIMIAASVAITIFLGVKGLQSFLVNGVSP 65

Query: 67  TDFLFNTKWEP--SAKSFGAFPMIAGSFIVTILSAIIATPFAIGAAVFMTEISPKYGSKI 124
            +FL +  W P  S   +G  P I GSF VTILSA+IA P  I   +FMTEI+P +G K+
Sbjct: 66  IEFLTSLNWNPTDSDPKYGVLPFIFGSFAVTILSALIAAPLGIAGPIFMTEIAPNWGKKV 125

Query: 125 LQPAVELLVGIPSVVYGFIGLQIIVPFVRSI-FGGTGFGILSGVCVLFVMILPTVTFMTV 183
           LQP +ELLVGIPSVVYGFIGL ++VPF+      GTG  +L+G  VL VMILPT+T ++
Sbjct: 126 LQPVIELLVGIPSVVYGFIGLTVLVPFIAQFKSSGTGHSLLAGTIVLSVMILPTITSISA 185

Query: 184 DSLRAVPRHYKEASLAMGATRWQTIWRVILNAARPGIFTAIVFGMARAFGEALAIQMVVG 243
           D++ ++P+  +E S A+GATRWQTI +V++ AA P + TA+V GMARAFGEALA+QMV+G
Sbjct: 186 DAMASLPKSLREGSYALGATRWQTIRKVLVPAAFPTLMTAVVLGMARAFGEALAVQMVIG 245

Query: 244 NSAILPTSLTTPAATLTSVLTMGIGNTVMGTVQNNVLWSLALVLLIMSLAFNTVIKLITR 303
           N+ +LP S    A TLT+++T+ +G+T  G+V+NN LWS+ LVLL+MS  F  +I+ ++
Sbjct: 246 NTRVLPESPFDTAGTLTTIITLNMGHTTYGSVENNTLWSMGLVLLVMSFLFILLIRYLSS 305

Query: 304 EGK 306
             K
Sbjct: 306 RRK 308
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1691> which encodes the amino acid sequence <SEQ ID 1692>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-17.25   Transmembrane   29 -  45 (  21 -  55)
```

```
    INTEGRAL    Likelihood = -7.22   Transmembrane  162 - 178 ( 154 - 184)
    INTEGRAL    Likelihood = -5.57   Transmembrane  282 - 298 ( 277 - 302)
    INTEGRAL    Likelihood = -5.41   Transmembrane   96 - 112 (  81 - 116)
    INTEGRAL    Likelihood = -3.08   Transmembrane  133 - 149 ( 131 - 152)

----- Final Results -----
              bacterial membrane --- Certainty=0.7899(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 266/311 (85%), Positives = 290/311 (92%), Gaps = 6/311 (1%)

Query: 7   MKNQELAKKLTSPSKNSRLEKFGKGITFLSLALIVFIVAMILIFVAQKGLSTFFVDGVKL 66
           M+NQELAKKL SPSKNSRLE FG+ ITFL LALIVFIVAMILIFVAQKGLSTFFVD V L
Sbjct: 1   MENQELAKKLASPSKNSRLETFGRTITFLCLALIVFIVAMILIFVAQKGLSTFFVDKVNL 60

Query: 67  TDFLFNTKWEPSAKS------FGAFPMIAGSFIVTILSAIIATPFAIGAAVFMTEISPKY 120
            DFLF  +W+PS K+        GA PMI GSF+VTILSAIIATPFAIGAAVFMTEISPKY
Sbjct: 61  FDFLFGKEWQPSVKNAAGIPYLGALPMITGSFLVTILSAIIATPFAIGAAVFMTEISPKY 120

Query: 121 GSKILQPAVELLVGIPSVVYGFIGLQIIVPFVRSIFGGTGFGILSGVCVLFVMILPTVTF 180
           G+K+LQPAVELLVGIPSVVYGFIGLQ+IVPF+RSIFGGTGFGILSGVCVLFVMILPTVTF
Sbjct: 121 GAKLLQPAVELLVGIPSVVYGFIGLQVIVPFMRSIFGGTGFGILSGVCVLFVMILPTVTF 180

Query: 181 MTVDSLRAVPRHYKEASLAMGATRWQTIWRVILNAARPGIFTAIVFGMARAFGEALAIQM 240
           MT DSLRAVPRHY+EAS+AMGATRWQTIWRV+LNAARPGIFTA++FGMARAFGEALAIQM
Sbjct: 181 MTTDSLRAVPRHYREASMAMGATRWQTIWRVVLNAARPGIFTAVIFGMARAFGEALAIQM 240

Query: 241 VVGNSAILPTSLTTPAATLTSVLTMGIGNTVMGTVQNNVLWSLALVLLIMSLAFNTVIKL 300
           VVGNSA++P+SLTTPAATLTSVLTMGIGNTVMGTVQNNVLWSLALVLL+MSLAFN+++KL
Sbjct: 241 VVGNSAVMPSSLTTPAATLTSVLTMGIGNTVMGTVQNNVLWSLALVLLLMSLAFNSLVKL 300

Query: 301 ITREGKKNYER 311
           IT+E K+NYER
Sbjct: 301 ITKERKRNYER 311
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 669

A DNA sequence (GBSx0709) was identified in *S.agalactiae* <SEQ ID 2059> which encodes the amino acid sequence <SEQ ID 2060>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2469(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 670

A DNA sequence (GBSx0710) was identified in *S.agalactiae* <SEQ ID 2061> which encodes the amino acid sequence <SEQ ID 2062>. This protein is predicted to be probable abc transporter permease protein in soda-comga intergenic reg. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.24   Transmembrane   20 -  36 (  19 -  41)
   INTEGRAL    Likelihood = -8.28   Transmembrane   66 -  82 (  57 -  88)
   INTEGRAL    Likelihood = -6.90   Transmembrane  260 - 276 ( 258 - 285)
   INTEGRAL    Likelihood = -5.47   Transmembrane  109 - 125 ( 106 - 129)
   INTEGRAL    Likelihood = -2.87   Transmembrane  181 - 197 ( 178 - 198)

----- 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>
```

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

```
>GP:CAB14427 GB:Z99116 alternate gene name: yzmD~similar to
          phosphate ABC transporter (permease) [Bacillus subtilis]
 Identities = 157/294 (53%), Positives = 225/294 (76%)

Query: 1   MNAKKADKLATTILYSIAAIIVTILASLLIFILVRGLPHVSWSFLTGKSSSYEAGGGIGI 60
           MN K  DKLAT +    AAII  IL  L  +I++ G+  +S+ F+T KSS+  AGGGI
Sbjct: 1   MNRKITDKLATGMFGLCAAIIAAILVGLFSYIIINGVSQLSFQFITTKSSAIAAGGGIRD 60

Query: 61  QLYNSFFLLIVTLIISIPLSLGAGIYLSEYAKKGRLTNFVRTCIEILSSLPSVVVGLFGY 120
           QL+NSF++L +T++I+IPL +G G++++EYA   ++T+F+RTCIE+LSSLPS+V+G+FG
Sbjct: 61  QLFNSFYILFITMLITIPLGVGGGVFMAEYAPNNKVTDFIRTCIEVLSSLPSIVIGMFGL 120

Query: 121 LIFVVQFQYGFSIISGALALTVFNLPQMTRSVEDSLQNVHHTQREAGLALGISRWETVIY 180
           L+FV    +G++II GALALTVFNLP M R  ED++++V    +EA LALG+SRW TV
Sbjct: 121 LMFVNLTGWGYTIIGGALALTVFNLPVMVRVTEDAIRSVPKDLKEASLALGVSRWHTVKT 180

Query: 181 VVVPEALPSIVTGVVLASGRIFGEAAALIYTAGQSAPALDWSNWNVLSVTSPISIFRQAE 240
           V++P A+PSI+TG +LASGR+FGEAAAL++TAG + P L+++ WN  S TSP++IFR AE
Sbjct: 181 VLIPSAIPSIITGAILASGRVFGEAAALLFTAGLTTPRLNFTEWNPFSETSPLNIFRPAE 240

Query: 241 TLAVHIWKVNSEGTIPDATQVSAGSAAVLLVVILIFNLSARSIGKKLHSKLTSS 294
           TLAVHIW VN++G IPDA  ++ G + VL++ +L+FNL+AR +G  ++ KLT++
Sbjct: 241 TLAVHIWNVNTQGMIPDAEAIANGGSPVLVISVLVFNLAARWLGTMIYKKLTAN 294
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1685> which encodes the amino acid sequence <SEQ ID 1686>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.89   Transmembrane   17 -  33 (   8 -  40)
   INTEGRAL    Likelihood =-10.19   Transmembrane  260 - 276 ( 257 - 285)
   INTEGRAL    Likelihood = -5.89   Transmembrane   66 -  82 (  57 -  87)
   INTEGRAL    Likelihood = -5.47   Transmembrane  109 - 125 ( 106 - 129)
   INTEGRAL    Likelihood = -2.02   Transmembrane  181 - 197 ( 180 - 197)

----- Final Results -----
               bacterial membrane --- Certainty=0.5755(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 257/294 (87%), Positives = 278/294 (94%)

Query: 1   MNAKKADKLATTILYSIAAIIVTILASLLIFILVRGLPHVSWSFLTGKSSSYEAGGGIGI 60
```

```
            MNAKK DK+AT  LY+IA IIV ILASL+++ILVRGLPH+SWSFLTGKSSSYEAGGGIGI
Sbjct: 1    MNAKKVDKVATGTLYTIAGIIVAILASLILYILVRGLPHISWSFLTGKSSSYEAGGGIGI 60

Query: 61   QLYNSFFLLIVTLIISIPLSLGAGIYLSEYAKKGRLTNFVRTCIEILSSLPSVVVGLFGY 120
            QLYNSFFLLIVTLIISIPLS GAGIYL+EYAKKG +TNF+RTCIEILSSLPSVVVGLFGY
Sbjct: 61   QLYNSFFLLIVTLIISIPLSTGAGIYLAEYAKKGPVTNFIRTCIEILSSLPSVVVGLFGY 120

Query: 121  LIFVVQFQYGFSIISGALALTVFNLPQMTRSVEDSLQNVHHTQREAGLALGISRWETVIY 180
            LIFVVQF+YGFSIISGALALTVFNLPQMTR+VEDSL +VHHTQREAGLALG+SRWETV Y
Sbjct: 121  LIFVVQFEYGFSIISGALALTVFNLPQMTRNVEDSLLHVHHTQREAGLALGLSRWETVFY 180

Query: 181  VVVPEALPSIVTGVVLASGRIFGEAAALIYTAGQSAPALDWSNWNVLSVTSPISIFRQAE 240
            VV+PEALP +VTG+VLASGRIFGEAAALIYTAGQSAPALDWSNWN LSVTSPISIFRQ+E
Sbjct: 181  VVIPEALPGMVTGIVLASGRIFGEAAALIYTAGQSAPALDWSNWNPLSVTSPISIFRQSE 240

Query: 241  TLAVHIWKVNSEGTIPDATQVSAGSAAVLLVVILIFNLSARSIGKKLHSKLTSS 294
            TLAVHIWKVNSEGTIPDAT VSAGSAAVLL+ ILIFN SA  IGKKLHSK+T++
Sbjct: 241  TLAVHIWKVNSEGTIPDATLVSAGSAAVLLIFILIFNFSAHFIGKKLHSKMTAA 294
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 671

A DNA sequence (GBSx0711) was identified in *S.agalactiae* <SEQ ID 2063> which encodes the amino acid sequence <SEQ ID 2064>. This protein is predicted to be phosphate ABC transporter, ATP-binding protein (pstB) (pstB-2). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4506(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 GENPEPT database:

```
>GP:AAB99016 GB:U67544 phosphate specific transport complex
            component (pstB) [Methanococcus jannaschii]
 Identities = 154/247 (62%), Positives = 204/247 (82%)

Query: 21   LTTKDLHVYYGEKEAIKGIDMQFEKNKITALIGPSGCGKSTYLRSLNRMNDTIDIARVTG 80
            + TK+L+++YGEK+A+  I++   +NKITALIGPSGCGKST+LR LNR+ND I   R+ G
Sbjct: 6    METKNLNLWYGEKQALFDINLPIYENKITALIGPSGCGKSTFLRCLNRLNDLIPNVRIEG 65

Query: 81   QIMYEGIDVNAQDINVYEMRKHIGMVFQRPNPFAKSIYKNITFAYERAGVKDKKFLDEVV 140
            +++ +G ++   +D++VYE+RK +GMVFQ+PNPFA SIY N+ F     G+KDKK LD++V
Sbjct: 66   EVLLDGKNIYDKDVDVYELRKRVGMVFQKPNPFAMSIYDNVAFGPRIHGIKDKKELDKIV 125

Query: 141  ETSLKQAALWDQVKDDLHKSAFTLSGGQQQRLCIARAIAVKPEILLMDEPASALDPIATM 200
            E +LK+AALWD+VKD+LHK+A +LSGGQQQRLCIARAIAVKPE+LLMDEP SALDPI+T+
Sbjct: 126  EWALKKAALWDEVKDELHKNALSLSGGQQQRLCIARAIAVKPEVLLMDEPTSALDPISTL 185

Query: 201  QLEETMFELKKNYTIIIVTHNMQQAARASDYTAFFYLGDLIEYDKTNNIFQNAKCQSTSD 260
            ++EE M EL K+YTI++VTHNMQQA+R SDYTAFF +G LIE+ +T  IF N + + T D
Sbjct: 186  KIEELMVELAKDYTIVVVTHNMQQASRVSDYTAFFLMGKLIEFGETEQIFLNPQKKETDD 245

Query: 261  YVSGRFG 267
            Y+SGRFG
Sbjct: 246  YISGRFG 252
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1681> which encodes the amino acid sequence <SEQ ID 1682>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2796(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 242/267 (90%), Positives = 258/267 (95%)

Query: 1   MAEYNWDERHIITFPEENSALTTKDLHVYYGEKEAIKGIDMQFEKNKITALIGPSGCGKS 60
            M EYNW+ERHIITFPEE  AL TKDLHVYYG KEAIKGIDMQFEK+KITALIGPSGCGKS
Sbjct: 1   MTEYNWNERHIITFPEETLALATKDLHVYYGAKEAIKGIDMQFEKHKITALIGPSGCGKS 60

Query: 61  TYLRSLNRMNDTIDIARVTGQIMYEGIDVNAQDINVYEMRKHIGMVFQRPNPFAKSIYKN 120
            TYLRSLNRMNDTIDIARVTG+I+Y+GIDVN +D+NVYE+RKH+GMVFQRPNPFAKSIYKN
Sbjct: 61  TYLRSLNRMNDTIDIARVTGEILYQGIDVNRKDMNVYEIRKHLGMVFQRPNPFAKSIYKN 120

Query: 121 ITFAYERAGVKDKKFLDEVVETSLKQAALWDQVKDDLHKSAFTLSGGQQQRLCIARAIAV 180
            ITFA+ERAGVKDKK LDE+VETSLKQAALWDQVKDDLHKSAFTLSGGQQQRLCIARAI+V
Sbjct: 121 ITFAHERAGVKDKKVLDEIVETSLKQAALWDQVKDDLHKSAFTLSGGQQQRLCIARAISV 180

Query: 181 KPEILLMDEPASALDPIATMQLEETMFELKKNYTIIIVTHNMQQAARASDYTAFFYLGDL 240
            KP+ILLMDEPASALDPIATMQLEETMFELKKNYTIIIVTHNMQQAARASDYTAFFYLG+L
Sbjct: 181 KPDILLMDEPASALDPIATMQLEETMFELKKNYTIIIVTHNMQQAARASDYTAFFYLGNL 240

Query: 241 IEYDKTNNIFQNAKCQSTSDYVSGRFG 267
            IEYDKT NIFQNA+CQST+DYVSG FG
Sbjct: 241 IEYDKTRNIFQNAQCQSTNDYVSGHFG 267
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 672

A DNA sequence (GBSx0712) was identified in *S.agalactiae* <SEQ ID 2065> which encodes the amino acid sequence <SEQ ID 2066>. This protein is predicted to be phosphate ABC transporter, ATP-binding protein (pstB-1). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3806(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9815> which encodes amino acid sequence <SEQ ID 9816> was also identified.

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

```
>GP:CAB14426 GB:Z99116 alternate gene name: yzmE~similar to
          phosphate ABC transporter (ATP-binding protein)
          [Bacillus subtilis]
 Identities = 148/248 (59%), Positives = 189/248 (75%)

Query: 5   ILQVSDLSVYYNKKKALKEVSMDFYPNEITALIGPSGSGKSTLLRAINRMGDLNPEVTLT 64
            +L+V DLS+YY  K+A+  V+MD   N +TALIGPSG GKST LR INRM DL P
Sbjct: 22  VLEVKDLSIYYGNKQAVHHVNMDIEKNAVTALIGPSGCGKSTFLRNINRMNDLIPSARAE 81

Query: 65  GAVMYNGHNVYSPRTDTVELRKEIGMVFQQPNPFPMSVFENVVYGLRLKGIKDKATLDEA 124
```

```
            G ++Y G N+     + V LR+EIGMVFQ+PNPFP S++ N+ + L+  G ++KA LDE
Sbjct: 82   GEILYEGLNILGGNINVVSLRREIGMVFQKPNPFPKSIYANITHALKYAGERNKAVLDEI 141

Query: 125 VETSLKGASIWDEVKDRLHDSALGLSGGQQQRVCIARTLATKPKIILLDEPTSALDPISA 184
            VE SL  A++WDEVKDRLH SAL LSGGQQQR+CIARTLA KP ++LLDEP SALDPIS
Sbjct: 142 VEESLTKAALWDEVKDRLHSSALSLSGGQQQRLCIARTLAMKPAVLLLDEPASALDPISN 201

Query: 185 GKIEETLHGLKDQYTMLLVTRSMQQASRISDRTGFFLDGNLIEYGNTKEMFMNPKHKETE 244
             KIEE + GLK +Y++++VT +MQQA R+SDRT FFL+G L+EYG T+++F +PK ++TE
Sbjct: 202 AKIEELITGLKREYSIIIVTHNMQQALRVSDRTAFFLNGELVEYGQTEQIFTSPKKQKTE 261

Query: 245 DYITGKFG 252
            DYI GKFG
Sbjct: 262 DYINGKFG 269
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2067> which encodes the amino acid sequence <SEQ ID 2068>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3590(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 208/252 (82%), Positives = 235/252 (92%)

Query: 1   MTQPILQVSDLSVYYNKKKALKEVSMDFYPNEITALIGPSGSGKSTLLRAINRMGDLNPE 60
            MT+PILQ+ DLSVYYN+KK LK+VS+D YPNEITALIGPSGSGKSTLLR+INRM DLNPE
Sbjct: 2   MTEPILQIRDLSVYYNQKKTLKDVSLDLYPNEITALIGPSGSGKSTLLRSINRMNDLNPE 61

Query: 61  VTLTGAVMYNGHNVYSPRTDTVELRKEIGMVFQQPNPFPMSVFENVVYGLRLKGIKDKAT 120
            VT+TG+++YNGHN+YSPRTDTV+LRKEIGMVFQQPNPFPMS++ENVVYGLRLKGI+DK+
Sbjct: 62  VTITGSIVYNGHNIYSPRTDTVDLRKEIGMVFQQPNPFPMSIYENVVYGLRLKGIRDKSI 121

Query: 121 LDEAVETSLKGASIWDEVKDRLHDSALGLSGGQQQRVCIARTLATKPKIILLDEPTSALD 180
            LD AVE+SLKGASIW+EVKDRLHDSA+GLSGGQQQRVCIAR LAT P+IILLDEPTSALD
Sbjct: 122 LDHAVESSLKGASIWNEVKDRLHDSAVGLSGGQQQRVCIARVLATSPRIILLDEPTSALD 181

Query: 181 PISAGKIEETLHGLKDQYTMLLVTRSMQQASRISDRTGFFLDGNLIEYGNTKEMFMNPKH 240
            PISAGKIEETL  LK  YT+ +VTRSMQQASR+SDRTGFFL+G+L+E G TK MFMNPK
Sbjct: 182 PISAGKIEETLLLLKKDYTLAIVTRSMQQASRLSDRTGFFLEGDLLECGPTKAMFMNPKR 241

Query: 241 KETEDYITGKFG 252
            KETEDYI+GKFG
Sbjct: 242 KETEDYISGKFG 253
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 673

A DNA sequence (GBSx0713) was identified in *S.agalactiae* <SEQ ID 2069> which encodes the amino acid sequence <SEQ ID 2070>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1937(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 GENPEPT database:

```
>GP:AAD22042 GB:AF118229 PhoU [Streptococcus pneumoniae]
 Identities = 75/216 (34%), Positives = 126/216 (57%), Gaps = 1/216 (0%)

Query: 2   LRSKFDEELDKLHNQFYAMGIEAIGQIKKTVRAFVSHDRELAKEVIEDDVTLNNFETKLE 61
           +R++FD EL +L   F  +G   +    K + A  S D+E+A+ +I  D  +N  ++ +E
Sbjct: 1   MRNQFDLELHELEQSFLGLGQLVLETASKALLALASKDKEMAELIINKDHAINQGQSAIE 60

Query: 62  KKSLEIIALQQPVSQDLRTVITVLKATSDVERMGDHAAAVAKATIRMKGEERIPAVELEI 121
                ++ALQQP   DLR VI+++ + SD+ERMGDH A +AKA +++K E ++   E ++
Sbjct: 61  LTCARLLALQQPQVSDLRFVISIMSSCSDLERMGDHMAGIAKAVLQLK-ENQLAPDEEQL 119

Query: 122 NNMGKAVKNMLEEALTAYINGDDEKAYEVAAMDEIVDDYFRDIQKMVVETIQKHPDVAFA 181
           + MGK   +ML + L A+       KA  +A  DE +D Y+  + K ++  ++
Sbjct: 120 HQMGKLSLSMLADLLVAFPLHQASKAISIAQKDEQIDQYYYALSKEIIGLMKDQETSIPN 179

Query: 182 AKEYFQVLMHLERIGDYGKNICEWIVYLKTGKIIEL 217
             +Y  ++ HLER  DY  NICE +VYL+TG++++L
Sbjct: 180 GTQYLYIIGHLERFADYIANICERLVYLETGELVDL 215
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1677> which encodes the amino acid sequence <SEQ ID 1678>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2229(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 174/217 (80%), Positives = 194/217 (89%)

Query: 1   MLRSKFDEELDKLHNQFYAMGIEAIGQIKKTVRAFVSHDRELAKEVIEDDVTLNNFETKL 60
           MLR+KF+EELDKLHNQFY+MG+E + QI KTVRAFVSHDRELAKEVIE+D T+NNFETKL
Sbjct: 1   MLRTKFEEELDKLHNQFYSMGMEVLAQINKTVRAFVSHDRELAKEVIEEDDTINNFETKL 60

Query: 61  EKKSLEIIALQQPVSQDLRTVITVLKATSDVERMGDHAAAVAKATIRMKGEERIPAVELE 120
           EKKSLEIIALQQPVS DLR VITVLKA+SD+ERMGDHAA++AKATIRMKGEERIP VE +
Sbjct: 61  EKKSLEIIALQQPVSNDLRMVITVLKASSDIERMGDHAASIAKATIRMKGEERIPVVEEQ 120

Query: 121 INNMGKAVKNMLEEALTAYINGDDEKAYEVAAMDEIVDDYFRDIQKMVVETIQKHPDVAF 180
           IN MGKAVK M+EEAL AYIN DD KAYE+AA DEI+D YFR+IQ + VE I+K PD  F
Sbjct: 121 INLMGKAVKQMVEEALNAYINADDTKAYEIAASDEIIDQYFRNIQTLAVEEIRKSPDAVF 180

Query: 181 AAKEYFQVLMHLERIGDYGKNICEWIVYLKTGKIIEL 217
           A KEYFQVLM+LERIGDY +NICEWIVYLKTGKIIEL
Sbjct: 181 AGKEYFQVLMYLERIGDYARNICEWIVYLKTGKIIEL 217
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 674

A DNA sequence (GBSx0714) was identified in *S.agalactiae* <SEQ ID 2071> which encodes the amino acid sequence <SEQ ID 2072>. This protein is predicted to be aminopeptidase N. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2845(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 GENPEPT database:

```
>GP:CAB50785 GB:AJ007700 aminopeptidase N [Streptococcus thermophilus]
 Identities = 556/847 (65%), Positives = 673/847 (78%), Gaps = 4/847 (0%)

Query: 3   TVEHFVTKFVPENYNLFLDINRQTKTFSGNVAVSGEALDNNISFHQKGLTIKSVLLDNQP 62
           +V  F+  F+PENYNLFLDINR  KTF+GNVA++GEA+DN+IS HQK LTI SVLLDN+
Sbjct: 4   SVARFIESFIPENYNLFLDINRSEKTFTGNVAITGEAIDNHISLHQKDLTINSVLLDNES 63

Query: 63  LDFQLDEDNEAMHIQLHETGSMVLVFEFSGHITDNMTGMYPSYYTVNGIKKEVISTQFES 122
           L+FQ+D+ NEA HI+L ETG + +  EFSG ITDNMTG+YPSYYT NG KKE+ISTQFES
Sbjct: 64  LNFQMDDANEAFHIELPETGVLTIFIEFSGRITDNMTGIYPSYYTYNGEKKEIISTQFES 123

Query: 123 HFAREVFPSIDEPEAKATFDLSLKFDQKEGEIALSNMPEINAEQRQETGLWTFDTTPKMS 182
           HFARE FP +DEPEAKATFDLSLKFD +EG+ ALSNMPEIN+  R+ETG+WTF+TTP+MS
Sbjct: 124 HFAREAFPCVDEPEAKATFDLSLKFDAEEGDTALSNMPEINSHLREETGVWTFETTPRMS 183

Query: 183 SYLLAFALGELHGKTTHTKNGTLVGSYATKAHQLNELDFSLDIVVRVIEFYEDYFGVRYP 242
           +YLLAF  G LHGKT  TKNGT VG +AT A   N +DF+LDI VRVIEFYEDYF V+YP
Sbjct: 184 TYLLAFGFGALHGKTAKTKNGTEVGVFATVAQAENSVDFALDIAVRVIEFYEDYFQVKYP 243

Query: 243 IPQSLHVALPDFSAGAMENWGLVTYREVYLLVDENSSVSSRQQVALVVAHEIAHQWFGNL 302
           IP S H+ALPD SAGAMENWGLVTYREVYLLVDENSS +SRQQVALVVAHE+AHQWFGNL
Sbjct: 244 IPLSYHLALPDLSAGAMENWGLVTYREVYLLVDENSSAASRQQVALVVAHELAHQWFGNL 303

Query: 303 VTMKWWDDLWLNESFANMMEYVSIDYIEPKLNIFEDFQTG-GLPLALKRDATDGVQSVHV 361
           VTMKWWDDLWLNESFANMMEYVS++ IEP  NIFE F    G+P AL+RDATDGVQSVH+
Sbjct: 304 VTMKWWDDLWLNESFANMMEYVSVNAIEPSWNIFEGFPNKLGVPNALQRDATDGVQSVHM 363

Query: 362 EVNHPDEINTLFDPAIVYAKGSRLMHMLRRWLGDTDFAAGLKIYFEKHQYQNTIGRDLWN 421
           EVNHPDEINTLFD AIVYAKGSRLMHMLRRWLGD  FA GLK YFEKHQY NT+GRDLWN
Sbjct: 364 EVNHPDEINTLFDSAIVYAKGSRLMHMLRRWLGDEAFAKGLKAYFEKHQYNNTVGRDLWN 423

Query: 422 ALSQTSGKDVAAFMDSWLEQPGYPVMAAKIEEDELILTQKQFFIGEHEDKSRLWQIPLNS 481
           ALS+ SGKDV++FMD+WLEQPGYPV++A++ +D LIL+QKQFFIGEHEDK RLW+IPLN+
Sbjct: 424 ALSRASGKDVSSFMDTWLEQPGYPVVSAEVVDDTLILSQKQFFIGEHEDKGRLWEIPLNT 483

Query: 482 NWEGIPEILTEETVVIPNFSQLAEKNKENGALRFNTENTAHYITNYQGQLLEHIISDLPL 541
           NW G+P+ L+EE + IPN+SQLA +N  NG LR NT NTAHYIT+YQGQLL++I+ D
Sbjct: 484 NWNGLPDTLSEERIEIPNYSQLATEN--NGVLRLNTANTAHYITDYQGQLLDNILEDFAN 541

Query: 542 MDNISKLQIVQERHLLAESGMISYSSLIPLVSLLSQETSYLVNSAIKSVIDGLSLFVQED 601
           +D +SKLQI+QER LLAESG ISY+SL+ L+ L+ +E S+L++ A   ++ GL  F+ ED
Sbjct: 542 LDTVSKLQILQERRLLAESGRISYASLVGLLDLVEKEESFLISQAKSQILAGLKRFIDED 601

Query: 602 SQDEFDFKEFVNKLSAFNFNRLGFEKREGEGDDSEMVRHLSLSLALYSDNEHAIEEAHHI 661
           ++ E  +K  V++     +F RLGF+ +EGE D+ EMVR  +LS  + +D +  +  A ++
Sbjct: 602 TEAEVHYKALVSRQFQNDFERLGFDAKEGESDEDEMVRQTALSYLIEADYQPTVLAAANV 661

Query: 662 FKAHENNIAAIPAAIRLLVLTNEMKHFESKELSHLLLETYSTTTDGNFKRQLASALSHTT 721
           F+AH+ NI +IPA+IR LVL N+MK   S  L    +  Y  T D NF+RQL  ALS+
Sbjct: 662 FQAHKENIESIPASIRGLVLINQMKQENSLSLVEEYINAYVATNDSNFRRQLTQALSYLK 721

Query: 722 DSKTLKKLLSDWKNKDIVKPQDLAMSWYATFLKNSFTQESVWEWAQENWEWIKATLGGDM 781
           + + L  +L   K+K++VKPQDL + WY  FL  SF QE+VW+WA+ENWEWIKA LGGDM
Sbjct: 722 NQEGLDYVLGQLKDKNVVKPQDLYL-WYMNFLSKSFAQETVWDWAKENWEWIKAALGGDM 780

Query: 782 SFDKFVIYPSSSFKTEERLEQYKNFFEPQLSDMAISRNISMGIKEISARVLLITKQKEEV 841
           SFD FV  P+  FK +ERL+QY  FFEPQ SD A+ RNI MGIK I+ARV LI K+K  V
Sbjct: 781 SFDSFVNIPAGIFKNQERLDQYIAFFEPQTSDKALERNILMGIKTIAARVDLIEKEKAAV 840

Query: 842 INTIKKY 848
```

```
            + +K Y
Sbjct: 841 ESALKDY 847
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2073> which encodes the amino acid sequence <SEQ ID 2074>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1098(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 576/848 (67%), Positives = 692/848 (80%), Gaps = 3/848 (0%)

Query: 1   MKTVEHFVTKFVPENYNLFLDINRQTKTFSGNVAVSGEALDNNISFHQKGLTIKSVLLDN 60
            MKTVEH +  FVPENYN+FLDINRQTKTF+GNVA++GEALDN+++FHQK L IKS+LLDN
Sbjct: 21  MKTVEHLIETFVPENYNIFLDINRQTKTFTGNVAINGEALDNHVAFHQKDLDIKSILLDN 80

Query: 61  QPLDFQLDEDNEAMHIQLHETGSMVLVFEFSGHITDNMTGMYPSYYTVNGIKKEVISTQF 120
            + + +Q+D DNE + ++L ETG M LV EFSG ITDNMTG+YPSYYT NG KKEVISTQF
Sbjct: 81  EAVIYQVDNDNEVVRVELPETGMMTLVIEFSGSITDNMTGIYPSYYTKNGEKKEVISTQF 140

Query: 121 ESHFAREVFPSIDEPEAKATFDLSLKFDQKEGEIALSNMPEINAEQRQETGLWTFDTTPK 180
            ESHFARE FP IDEP+AKATFDLSL FDQ+ GEIALSNMPE+N ++R+ETGLWTFDTT +
Sbjct: 141 ESHFAREAFPCIDEPQAKATFDLSLTFDQEIGEIALSNMPEVNIDRREETGLWTFDTTLR 200

Query: 181 MSSYLLAFALGELHGKTTHTKNGTLVGSYATKAHQLNELDFSLDIVVRVIEFYEDYFGVR 240
            MSSYLLAFALGELHGKT  +K GT VG YAT AH L+ LDFSLDI VRVI FYEDYFGV
Sbjct: 201 MSSYLLAFALGELHGKTVESKKGTTVGVYATTAHPLSSLDFSLDIAVRVINFYEDYFGVH 260

Query: 241 YPIPQSLHVALPDFSAGAMENWGLVTYREVYLLVDENSSVSSRQQVALVVAHEIAHQWFG 300
            YPIPQSL++ALPDFS+GAMENWGL+TYRE+YLLVDENS+V SRQQVALV+AHEIAHQWFG
Sbjct: 261 YPIPQSLNIALPDFSSGAMENWGLITYREIYLLVDENSTVQSRQQVALVIAHEIAHQWFG 320

Query: 301 NLVTMKWWDDLWLNESFANMMEYVSIDYIEPKLNIFEDFQTGGLPLALKRDATDGVQSVH 360
            NLVTMKWWDDLWLNESFANMMEYVSI+ IEP   I EDFQTGG+PLALKRDATDGVQSVH
Sbjct: 321 NLVTMKWWDDLWLNESFANMMEYVSIEAIEPSWKIIEDFQTGGIPLALKRDATDGVQSVH 380

Query: 361 VEVNHPDEINTLFDPAIVYAKGSRLMHMLRRWLGDTDFAAGLKIYFEKHQYQNTIGRDLW 420
            VEVNHPDEINTLFDPAIVYAKGSRLMHMLRR++GD DFA GL  YFEK+QY+NT+GRDLW
Sbjct: 381 VEVNHPDEINTLFDPAIVYAKGSRLMHMLRRFIGDRDFAIGLHHYFEKYQYRNTVGRDLW 440

Query: 421 NALSQTSGKDVAAFMDSWLEQPGYPVMAAKIEEDELILTQKQFFIGEHEDKSRLWQIPLN 480
            N LS TSGKDVAAFMD+WLEQPGYPV+ A++E D+LIL+QKQFFIG+ E+K RLW IPLN
Sbjct: 441 NILSDTSGKDVAAFMDAWLEQPGYPVLTARLENDQLILSQKQFFIGKGEEKGRLWPIPLN 500

Query: 481 SNWEGIPEILTEETVVIPNFSQLAEKNKENGALRFNTENTAHYITNYQGQLLEHIISDLP 540
            +NW G+PE LTE  +VIPNFSQLA +N+  GALRFN +NTAHYIT+YQG LL+ ++++L
Sbjct: 501 TNWHGLPETLTEAEMVIPNFSQLAAENE--GALRFNIDNTAHYITDYQGSLLDALVTELA 558

Query: 541 LMDNISKLQIVQERHLLAESGMISYSSLIPLVSLLSQETSYLVNSAIKSVIDGLSLFVQE 600
             +DN S LQ++QER LLA+SG+ISY+ L+ L++ L    SY+V  A++ V+ GL  F+ E
Sbjct: 559 QLDNTSALQVIQERRLLADSGLISYAELVDLIAQLDDSKSYMVAEAVQQVVSGLKRFIDE 618

Query: 601 DSQDEFDFKEFVNKLSAFNFNRLGFEKREGEGDDSEMVRHLSLSLALYSDNEHAIEEAHH 660
             S  E  F   V  +   +FN+ GFEK+  E D+ EMVR ++L      ++N+  I+
Sbjct: 619 GSLAEKSFNRLVTTIYQEDFNQHGFEKKADESDEDEMVRQVALGRLWLAENQTIIDGLRT 678

Query: 661 IFKAHENNIAAIPAAIRLLVLTNEMKHFESKELSHLLLETYSTTTDGNFKRQLASALSHT 720
            IF+A++NNIA+IPAA+R LVL N+MK+FE+  L  + ETY  TTD N +  L A S T
Sbjct: 679 IFEAYQNNIASIPAAVRRLVLANQMKYFETDSLVDIYFETYVATTDNNLRSDLTVAFSQT 738

Query: 721 TDSKTLKKLLSDWKNKDIVKPQDLAMSWYATFLKNSFTQESVWEWAQENWEWIKATLGGD 780
```

```
              T++++L   K+KDI+KPQDL+  WY   L  SFTQ+ +WEWA+ENW+WIK+ LGGD
Sbjct: 739 KQPTTIRRILVSLKDKDIIKPQDLSY-WYNALLGQSFTQDIIWEWARENWDWIKSALGGD 797

Query: 781 MSFDKFVIYPSSSFKTEERLEQYKNFFEPQLSDMAISRNISMGIKEISARVLLITKQKEE 840
           MSFDKFVIYP+S+FKT + L +YK+FFEP+L DMAISRNI+MGI EI ARV LITK+KE
Sbjct: 798 MSFDKFVIYPASNFKTPKHLAEYKSFFEPKLDDMAISRNITMGINEIEARVALITKEKEA 857

Query: 841 VINTIKKY 848
           VI  +  Y
Sbjct: 858 VIAALSHY 865
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 675

A DNA sequence (GBSx0715) was identified in *S.agalactiae* <SEQ ID 2075> which encodes the amino acid sequence <SEQ ID 2076>. This protein is predicted to be response regulator (trcR). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2741(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 GENPEPT database:

```
>GP:CAA54465 GB:X77249 response regulator [Streptococcus pneumoniae]
 Identities = 198/224 (88%), Positives = 213/224 (94%)

Query: 1   MIKILLIEDDLSLSNSVFDFLDDFADVMQIFDGEEGLYEAESGVYDLILLDLMLPEKNGF 60
           MIKILL+EDDL LSNSVFDFLDDFADVMQ+FDGEEGLYEAESGVYDLILLDLMLPEKNGF
Sbjct: 1   MIKILLVEDDLGLSNSVFDFLDDFADVMQVFDGEEGLYEAESGVYDLILLDLMLPEKNGF 60

Query: 61  QVLKELREKGITTPVLIMTAKESIDDKGQGFDLGADDYLTKPFYLEELKMRIQALLKRSG 120
           QVLKELREKGITTPVLIMTAKES+DDKG GF+LGADDYLTKPFYLEELKMRIQALLKRSG
Sbjct: 61  QVLKELREKGITTPVLIMTAKESLDDKGHGFELGADDYLTKPFYLEELKMRIQALLKRSG 120

Query: 121 KFNDNSLIYGDIRVDMSTNSTFVNQTEVELLGKEFDLLVYFLQNQNVILPKSQIFDRIWG 180
           KFN+N+L YG+I V++STN+  V  T VELLGKEFDLLVYFLQNQNVILPK+QIFDR+WG
Sbjct: 121 KFNENTLTYGNIVVNLSTNTVKVEDTPVELLGKEFDLLVYFLQNQNVILPKTQIFDRLWG 180

Query: 181 FDSDTTISVVEVYVSKVRKKLKGTLFSENLQTLRSVGYILKHVE 224
           FDSDTTISVVEVYVSKVRKKLKGT F+ENLQTLRSVGY+LK V+
Sbjct: 181 FDSDTTISVVEVYVSKVRKKLKGTTFAENLQTLRSVGYLLKDVQ 224
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2077> which encodes the amino acid sequence <SEQ ID 2078>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2689(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 180/224 (80%), Positives = 200/224 (88%)
```

```
Query: 1   MIKILLIEDDLSLSNSVFDFLDDFADVMQIFDGEEGLYEAESGVYDLILLDLMLPEKNGF 60
           MIKILL+EDDLSLSNS+FDFLDDFADVMQ+FDG+EGLYEAESG+YDLILLDLMLPEKNGF
Sbjct: 1   MIKILLVEDDLSLSNSIFDFLDDFADVMQVFDGDEGLYEAESGIYDLILLDLMLPEKNGF 60

Query: 61  QVLKELREKGITTPVLIMTAKESIDDKGQGFDLGADDYLTKPFYLEELKMRIQALLKRSG 120
           QVLKELREK I  PVLIMTAKE +DDKG GF+LGADDYLTKPFYLEELKMRIQALLKR+G
Sbjct: 61  QVLKELREKDIKIPVLIMTAKEGLDDKGHGFELGADDYLTKPFYLEELKMRIQALLKRTG 120

Query: 121 KFNDNSLIYGDIRVDMSTNSTFVNQTEVELLGKEFDLLVYFLQNQNVILPKSQIFDRIWG 180
           KF D ++ +G++ VD++      V    VELLGKEFDLLVY LQNQNVILPK+QIFDR+WG
Sbjct: 121 KFADKNISFGNLVVDLARKEVKVEGKVVELLGKEFDLLVYLLQNQNVILPKTQIFDRLWG 180

Query: 181 FDSDTTISVVEVYVSKVRKKLKGTLFSENLQTLRSVGYILKHVE 224
           FDSDTTISVVEVY+SK+RKKLKGT F   LQTLRSVGYILK+ E
Sbjct: 181 FDSDTTISVVEVYISKIRKKLKGTCFVNRLQTLRSVGYILKNNE 224
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 676

A DNA sequence (GBSx0716) was identified in *S.agalactiae* <SEQ ID 2079> which encodes the amino acid sequence <SEQ ID 2080>. This protein is predicted to be histidine kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.18   Transmembrane   22 -  38 (  17 -  46)
   INTEGRAL    Likelihood = -4.94   Transmembrane  182 - 198 ( 178 - 201)

----- Final Results -----
               bacterial membrane --- Certainty=0.4673(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 GENPEPT database:

```
>GP:CAA54466 GB:X77249 histidine kinase [Streptococcus pneumoniae]
 Identities = 218/420 (51%), Positives = 305/420 (71%), Gaps = 4/420 (0%)

Query: 17  SHFIHFFTVFSGIFLVMTVIILQVMRYGVYSSVDSSLKYISTHPKNYINMVMSRTAAY-- 74
           S+FI  F VF+ IF  MT+IILQVM   +Y+SVD  L  +S +P+  I + ++R
Sbjct: 15  SYFIRNFGVFTLIFSTMTLIILQVMHSSLYTSVDDKLHGLSENPQAVIQLAINRATEEIK 74

Query: 75  -LDNSNIASVKLKPGGQTVANTDIILFTSEEEVINYFDAFSNYQFLKPNKKNLGGISELT 133
            L+N+   + K++       +NT++ILF  +   +  F      +K  KK LG I ++
Sbjct: 75  DLENARADASKVEIKPNVSSNTEVILFDKDFTQLLSGNRFLGLDKIKLEKKELGHIYQIQ 134

Query: 134 LTNIFGQDETYHAVTVKVN-NPAYPNVTYMTAIVNIDQLVNAKERYEKIIIFVMTTFWII 192
           + N +GQ+E Y  + ++ N +    N+ Y   ++N  QL  A +++E++I+ VM +FWI+
Sbjct: 135 VFNSYGQEEIYRVILMETNISSVSTNIKYAAVLINTSQLEQASQKHEQLIVVVMASFWIL 194

Query: 193 SIGASIYLAKWAQKPIIENYERQKAFVENASHELRTPLAVLQNRLETLFRKPNATILENS 252
           S+ AS+YLA+ + +P++E+ ++Q++FVENASHELRTPLAVLQNRLETLFRKP ATI++ S
Sbjct: 195 SLLASLYLARVSVRPLLESMQKQQSFVENASHELRTPLAVLQNRLETLFRKPEATIMDVS 254

Query: 253 ENIASSLDEVRNMRILTTNLLNLARRDDGIKPELAVIKPTLFDSIFENYDLITQENGKNF 312
           E+IASSL+EVRNMR LTT+LLNLARRDDGIKPELA +  + F++ F NY++I  EN + F
Sbjct: 255 ESIASSLEEVRNMRFLTTSLLNLARRDDGIKPELAEVPTSFFNTTFTNYEMIASENNRVF 314

Query: 313 TGHNMIQDSFKTDKTLLKQLMTILFDNAIKYTDNDGSIDFTISETDKYLFLEIADNGPGI 372
              N I  + TD+ LLKQLMTILFDNA+KYT+ DG IDF IS TD+ L+L ++DNG GI
Sbjct: 315 RFENRIHRTIVTDQLLLKQLMTILFDNAVKYTEEDGEIDFLISATDRNLYLLVSDNGIGI 374

Query: 373 SEEDKVRIFDRFYRVDKARTRQQGGFGLGLSLAQQIVNSLRGNITVIDNKPRGSIFKIKL 432
```

```
            S EDK +IFDRFYRVDKARTRQ+GGFGLGLSLA+QIV++L+G +TV DNKP+G+IF++K+
Sbjct: 375 STEDKKKIFDRFYRVDKARTRQKGGFGLGLSLAKQIVDALKGTVTVKDNKPKGTIFEVKI 434
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2081> which encodes the amino acid sequence <SEQ ID 2082>. Analysis of this protein sequence reveals the following:

```
     Possible site: 57
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.09   Transmembrane   19 -  35 ( 14 -  44)
   INTEGRAL    Likelihood =-10.24   Transmembrane  185 - 201 ( 182 - 206)

----- 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:

```
>GP:CAA54466 GB:X77249 histidine kinase [Streptococcus pneumoniae]
 Identities = 223/436 (51%), Positives = 313/436 (71%), Gaps = 5/436 (1%)

Query: 2   NKLKKEILSDNYNHFFHFFAVFTGIFVIMTIIILQIMRFGVYSSVDSSLVSVSNNASSYA 61
           +KLKK   +D++++F   F VFT IF  MT+IILQ+M   +Y+SVD  L  +S N  +
Sbjct: 3   SKLKKTWYADDFSYFIRNFGVFTLIFSTMTLIILQVMHSSLYTSVDDKLHGLSENPQAVI 62

Query: 62  NRTMARISSFYFDTENNIIKALPDSDSSKLLGTPAANTDIILFSANGTILNAFDAFSNYQ 121
              + R +    D EN    A  D+   ++    ++NT++ILF  + T L + + F
Sbjct: 63  QLAINRATEEIKDLEN----ARADASKVEIKPNVSSNTEVILFDKDFTQLLSGNRFLGLD 118

Query: 122 NFHLDKRRLGSIETTSLMNFYGQEEKYHTITVGVHIKNYPA-VAYMMAVVNVEQLDRANE 180
              L+K+ LG I    + N YGQEE Y  I +  +I +    + Y   ++N  QL++A++
Sbjct: 119 KIKLEKKELGHIYQIQVFNSYGQEEIYRVILMETNISSVSTNIKYAAVLINTSQLEQASQ 178

Query: 181 RYERIIIIVMSVFWLISILASIYLAKWSRKPILESYEKQKMFVENASHELRTPLAVLQNR 240
           ++E++I++VM+ FW++S+LAS+YLA+ S +P+LES +KQ+ FVENASHELRTPLAVLQNR
Sbjct: 179 KHEQLIVVVMASFWILSLLASLYLARVSVRPLLESMQKQQSFVENASHELRTPLAVLQNR 238

Query: 241 LESLFRKPNETILENSEHLASSLDEVRNMRILTTNLLNLARRDDGINPQWTHLDTDFFNA 300
           LE+LFRKP  TI++ SE +ASSL+EVRNMR LTT+LLNLARRDDGI P+   + T FFN
Sbjct: 239 LETLFRKPEATIMDVSESIASSLEEVRNMRFLTTSLLNLARRDDGIKPELAEVPTSFFNT 298

Query: 301 IFENYELVAKEYGKIFYFQNQVNRSLRMDKALLKQLITILFDNAIKYTDKNGIIEIIVKT 360
            F NYE++A E  ++F F+N+++R++  D+ LLKQL+TILFDNA+KYT+++G I+ ++
Sbjct: 299 TFTNYEMIASENNRVFRFENRIHRTIVTDQLLLKQLMTILFDNAVKYTEEDGEIDFLISA 358

Query: 361 TDKNLLISVIDNGPGITDEEKKKIFDRFYRVDKARTRQTGGFGLGLALAQQIVMSLKGNI 420
           TD+NL + V DNG GI+ E+KKKIFDRFYRVDKARTRQ GGFGLGL+LA+QIV +LKG +
Sbjct: 359 TDRNLYLLVSDNGIGISTEDKKKIFDRFYRVDKARTRQKGGFGLGLSLAKQIVDALKGTV 418

Query: 421 TVKDNDPKGSIFEVKL 436
           TVKDN PKG+IFEVK+
Sbjct: 419 TVKDNKPKGTIFEVKI 434
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 265/436 (60%), Positives = 334/436 (75%), Gaps = 10/436 (2%)

Query: 7   ISKFKKNV-SDS--HFIHFFTVFSGIFLVMTVIILQVMRYGVYSSVDSSLKYISTHPKNY 63
           ++K KK + SD+  HF HFF VF+GIF++MT+IILQ+MR+GVYSSVDSSL  +S +  +Y
Sbjct: 1   MNKLKKEILSDNYNHFFHFFAVFTGIFVIMTIIILQIMRFGVYSSVDSSLVSVSNNASSY 60

Query: 64  INMVMSRTAAYLDNSNIASVKLKPG-------GQTVANTDIILFTSEEEVINYFDAFSNY 116
            N  M+R +++  ++    +K  P         G   ANTDIILF++   ++N FDAFSNY
Sbjct: 61  ANRTMARISSFYFDTENNIIKALPDSDSSKLLGTPAANTDIILFSANGTILNAFDAFSNY 120

Query: 117 QFLKPNKKNLGGISELTLTNIFGQDETYHAVTVKVNNPAYPNVTYMTAIVNIDQLVNAKE 176
           Q     +K+ LG I    +L N +GQ+E YH +TV V+    YP V YM A+VN++QL  A E
```

```
Sbjct: 121 QNFHLDKRRLGSIETTSLMNFYGQEEKYHTITVGVHIKNYPAVAYMMAVVNVEQLDRANE 180

Query: 177 RYEKIIIFVMTTFWIISIGASIYLAKWAQKPIIENYERQKAFVENASHELRTPLAVLQNR 236
           RYE+III VM+ FW+ISI ASIYLAKW++KPI+E+YE+QK FVENASHELRTPLAVLQNR
Sbjct: 181 RYERIIIIVMSVFWLISILASIYLAKWSRKPILESYEKQKMFVENASHELRTPLAVLQNR 240

Query: 237 LETLFRKPNATILENSENIASSLDEVRNMRILTTNLLNLARRDDGIKPELAVIKPTLFDS 296
           LE+LFRKPN TILENSE++ASSLDEVRNMRILTTNLLNLARRDDGI P+   +    F++
Sbjct: 241 LESLFRKPNETILENSEHLASSLDEVRNMRILTTNLLNLARRDDGINPQWTHLDTDFFNA 300

Query: 297 IFENYDLITQENGKNFTGHNMIQDSFKTDKTLLKQLMTILFDNAIKYTDNDGSIDFTISE 356
           IFENY+L+ +E GK F   N +  S + DK LLKQL+TILFDNAIKYTD +G I+  +
Sbjct: 301 IFENYELVAKEYGKIFYFQNQVNRSLRMDKALLKQLITILFDNAIKYTDKNGIIEIIVKT 360

Query: 357 TDKYLFLEIADNGPGISEEDKVRIFDRFYRVDKARTRQQGGFGLGLSLAQQIVNSLRGNI 416
           TDK L + + DNGPGI++E+K +IFDRFYRVDKARTRQ GGFGLGL+LAQQIV SL+GNI
Sbjct: 361 TDKNLLISVIDNGPGITDEEKKKIFDRFYRVDKARTRQTGGFGLGLALAQQIVMSLKGNI 420

Query: 417 TVIDNKPRGSIFKIKL 432
           TV DN P+GSIF++KL
Sbjct: 421 TVKDNDPKGSIFEVKL 436
```

SEQ ID 2080 (GBS339d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 9; MW 73kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 185 (lane 5; MW 73kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 677

A DNA sequence (GBSx0717) was identified in *S.agalactiae* <SEQ ID 2083> which encodes the amino acid sequence <SEQ ID 2084>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1783(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9813> which encodes amino acid sequence <SEQ ID 9814> was also identified.

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

```
>GP:AAB48049 GB:U88582 YlxM [Streptococcus mutans]
 Identities = 95/110 (86%), Positives = 103/110 (93%)

Query: 1   MEIEKTNRMNALFEFYAALLTDKQMNYIELYYADDYSLAEIAEESGVSRQAVYDNIKRTE 60
           MEIEKTNRMNALFEFYAALLTDKQMNYIELYYADDYSLAEIAEE VSRQAVYDNIKRTE
Sbjct: 1   MEIEKTNRMNALFEFYAALLTDKQMNYIELYYADDYSLAEIAEEFDVSRQAVYDNIKRTE 60

Query: 61  KILEAYEMKLHMYSDYIVRSQIFDDILEKYTDDAFLQEKISILSSIDNRD 110
           KILE YEMKLHMYSDY+VRS+IFD I++KY +D +LQ KISIL++IDNRD
Sbjct: 61  KILEDYEMKLHMYSDYVVRSEIFDAIMKKYPNDPYLQNKISILTTIDNRD 110
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2085> which encodes the amino acid sequence <SEQ ID 2086>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1767(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 95/110 (86%), Positives = 103/110 (93%)

Query: 1   MEIEKTNRMNALFEFYAALLTDKQMNYIELYYADDYSLAEIAEESGVSRQAVYDNIKRTE 60
           MEIEKTNRMNALFEFYAALLTDKQMNYIELYYADDYSLAEIA+E GVSRQAVYDNIKRTE
Sbjct: 4   MEIEKTNRMNALFEFYAALLTDKQMNYIELYYADDYSLAEIADEPGVSRQAVYDNIKRTE 63

Query: 61  KILEAYEMKLHMYSDYIVRSQIFDDILEKYTDDAFLQEKISILSSIDNRD 110
           KILE YEMKLHMYSDY+VRS+IFDD++  Y  D +LQEKISIL+SIDNR+
Sbjct: 64  KILETYEMKLHMYSDYVVRSEIFDDMIAHYPHDEYLQEKISILTSIDNRE 113
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 678

A DNA sequence (GBSx0719) was identified in *S.agalactiae* <SEQ ID 2087> which encodes the amino acid sequence <SEQ ID 2088>. This protein is predicted to be signal recognition particle protein (ffh). Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22   Transmembrane   37 -  53 (  37 -  53)

----- Final Results -----
             bacterial membrane --- Certainty=0.1086(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 GENPEPT database:

```
>GP:AAB48050 GB:U88582 Ffh [Streptococcus mutans]
 Identities = 437/522 (83%), Positives = 484/522 (92%), Gaps = 7/522 (1%)

Query: 1   MAFESLTERLQGVFKNIRGKKKLSEKDVQEVTKEIRLALLEADVALPVVKTFIKHVRERA 60
           MAFESLTERLQGVFKN+RGK+KLSEKDVQEVTKEIRLALLEADVALPVVK FIK VR+RA
Sbjct: 1   MAFESLTERLQGVFKNLRGKRKLSEKDVQEVTKEIRLALLEADVALPVVKEFIKRVRKRA 60

Query: 61  VGHEIIDTLDPTQQIVKIVNEELTDLLGAETSEIEKSPKIPTIIMMVGLQGAGKTTFAGK 120
           VGHE+IDTLDP+QQI+KIVNEELT +LG+ET+EIEKS KIPTIIMMVGLQGAGKTTFAGK
Sbjct: 61  VGHEVIDTLDPSQQIIKIVNEELTAVLGSETAEIEKSSKIPTIIMMVGLQGAGKTTFAGK 120

Query: 121 LANKLIKEDNARPMMIAADIYRPAAIDQLKTLGSQINVPVFDMGTNHSAVEIVTKGLEQA 180
           LANKL+KE+NARP+MIAADIYRPAAIDQLK LG QINVPVFDMGT HSAVEIV++GL QA
Sbjct: 121 LANKLVKEENARPLMIAADIYRPAAIDQLKILGQQINVPVFDMGTEHSAVEIVSQGLAQA 180

Query: 181 RENRNDYVLIDTAGRLQIDATLMQELHDVKAIAQPNEILLVVDSMIGQEAANVAEEFNRQ 240
           +ENRNDYVLIDTAGRLQID  LM EL D+KA+A PNEILLVVDSMIGQEAANVA EFN+Q
Sbjct: 181 KENRNDYVLIDTAGRLQIDEKLMTELRDIKALANPNEILLVVDSMIGQEAANVAREFNQQ 240

Query: 241 LSISGVVLTKIDGDTRGGAALSVREITGKPIKFTGTGEKITDIETFHPDRMASRILGMGD 300
           L ++GV+LTKIDGDTRGGAALSVR+ITGKPIKFTGTGEKITDIETFHPDRM+SRILGMGD
Sbjct: 241 LEVTGVILTKIDGDTRGGAALSVRQITGKPIKFTGTGEKITDIETFHPDRMSSRILGMGD 300

Query: 301 LLTLIERASQEYDEKRSMELAEKMRENTFDFNDFIDQLDQVQNMGPMEDLLKMLPGMANN 360
           LLTLIE+ASQ+YDE++S ELAEKMREN+FDFNDFI+QLDQVQNMG MED+LKM+PGMANN
Sbjct: 301 LLTLIEKASQDYDEQKSAELAEKMRENSFDFNDFIEQLDQVQNMGSMEDILKMIPGMANN 360
```

```
Query: 361 PAMKNFKVDENEIARKRAIVSSMTPEERENPDLLNPSRRRRIAAGSGNTFVDVNKFIKDF 420
           PA+ N +VDE EIARKRAIVSSMTPEERENPDLL PSRRRRIA+GSGNTFV+VNKFIKDF
Sbjct: 361 PALANVEVDEGEIARKRAIVSSMTPEERENPDLLTPSRRRRIASGSGNTFVNVNKFIKDF 420

Query: 421 NQAKQMMQGVMSGDMNKMMKKMGIDPNNLPKDMPGMDGMDMSNLEGMMGQNGMPDLSSL- 479
           NQAK+MMQGVMSGDMNK+MK+MGI+PNN+P       + MD S LEGMMGQ GMPD+S L
Sbjct: 421 NQAKKMMQGVMSGDMNKVMKQMGINPNNMP------NNMDSSALEGMMGQGGMPDMSGLS 474

Query: 480 GGDMDFSQMFGGGLKGKVGAFAAKQSMKRMANKMKKAKKKRK 521
           G +MD SQMFGGGLKGKVG FA KQSMK+MA +MKKAKK++K
Sbjct: 475 GANMDVSQMFGGGLKGKVGEFAMKQSMKKMAKRMKKAKKRKK 516
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2089> which encodes the amino acid sequence <SEQ ID 2090>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22   Transmembrane   39 -  55 (  39 -  55)

----- Final Results -----
               bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 458/522 (87%), Positives = 489/522 (92%), Gaps = 4/522 (0%)

Query: 1   MAFESLTERLQGVFKNIRGKKKLSEKDVQEVTKEIRLALLEADVALPVVKTFIKHVRERA 60
           MAFESLT+RLQ VFK+IRGKKKLSE DVQEVTKEIRLALLEADVALPVVKTFIK VRERA
Sbjct: 3   MAFESLTQRLQDVFKHIRGKKKLSESDVQEVTKEIRLALLEADVALPVVKTFIKRVRERA 62

Query: 61  VGHEIIDTLDPTQQIVKIVNEELTDLLGAETSEIEKSPKIPTIIMMVGLQGAGKTTFAGK 120
           +GHEIIDTLDPTQQI+KIVNEELT +LG+ET+EI+KSPKIPTIIMMVGLQGAGKTTFAGK
Sbjct: 63  IGHEIIDTLDPTQQILKIVNEELTSILGSETAEIDKSPKIPTIIMMVGLQGAGKTTFAGK 122

Query: 121 LANKLIKEDNARPMMIAADIYRPAAIDQLKTLGSQINVPVFDMGTNHSAVEIVTKGLEQA 180
           LANKLIKE+NARP+MIAADIYRPAAIDQLKTLG QINVPVFDMGT+HSAV+IV KGLEQA
Sbjct: 123 LANKLIKEENARPLMIAADIYRPAAIDQLKTLGQQINVPVFDMGTDHSAVDIVRKGLEQA 182

Query: 181 RENRNDYVLIDTAGRLQIDATLMQELHDVKAIAQPNEILLVVDSMIGQEAANVAEEFNRQ 240
           REN NDYVLIDTAGRLQID  LM EL DVKA+AQPNEILLVVDSMIGQEAANVA EFN Q
Sbjct: 183 RENHNDYVLIDTAGRLQIDEKLMGELRDVKALAQPNEILLVVDSMIGQEAANVAYEFNHQ 242

Query: 241 LSISGVVLTKIDGDTRGGAALSVREITGKPIKFTGTGEKITDIETFHPDRMASRILGMGD 300
           LSI+GVVLTKIDGDTRGGAALSVREITGKPIKFTG GEKITDIETFHPDRM+SRILGMGD
Sbjct: 243 LSITGVVLTKIDGDTRGGAALSVREITGKPIKFTGIGEKITDIETFHPDRMSSRILGMGD 302

Query: 301 LLTLIERASQEYDEKRSMELAEKMRENTFDFNDFIDQLDQVQNMGPMEDLLKMLPGMANN 360
           LLTLIE+ASQEYDEK+S+ELAEKMRENTFDFNDFI+QLDQVQNMGPMEDLLKM+PGMA N
Sbjct: 303 LLTLIEKASQEYDEKKSLELAEKMRENTFDFNDFIEQLDQVQNMGPMEDLLKMIPGMAGN 362

Query: 361 PAMKNFKVDENEIARKRAIVSSMTPEERENPDLLNPSRRRRIAAGSGNTFVDVNKFIKDF 420
           PA+ N KVDEN+IARKRAIVSSMTP ERENPDLLNPSRRRRIAAGSGN+FVD NKFIKDF
Sbjct: 363 PALANIKVDENQIARKRAIVSSMTPAERENPDLLNPSRRRRIAAGSGNSFVD-NKFIKDF 421

Query: 421 NQAKQMMQGVMSGDMNKMMKKMGIDPNNLPKDMPGMDGM-DMSNLEGMMGQNGMPDLSSL 479
           NQAK MMQGVMSGDM+KMMK MGI+PNNLPK+MP   GM DMS+LEGMMGQ GMPDLS L
Sbjct: 422 NQAKSMMQGVMSGDMSKMMKDMGINPNNLPKNMPA--GMPDMSSLEGMMGQGGMPDLSGL 479

Query: 480 GGDMDFSQMFGGGLKGKVGAFAAKQSMKRMANKMKKAKKKRK 521
           GGDMD SQ+FG G KGK+G FA KQ+MKR ANK+KKAKKKRK
Sbjct: 480 GGDMDMSQLFGKGFKGKIGQFAMKQAMKRQANKLKKAKKKRK 521
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 679

A DNA sequence (GBSx0721) was identified in *S.agalactiae* <SEQ ID 2091> which encodes the amino acid sequence <SEQ ID 2092>. This protein is predicted to be SatD. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.28   Transmembrane    3 -  19 (   2 -  19)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9811> which encodes amino acid sequence <SEQ ID 9812> was also identified.

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

```
>GP:AAG28336 GB:U88582 SatD [Streptococcus mutans]
 Identities = 106/222 (47%), Positives = 162/222 (72%), Gaps = 2/222 (0%)

Query: 13  MYLALIGDIINSKQILERETFQQSFQQLMTELSDVYGEELISPFTITAGDEFQALLKPSK 72
           +Y+A+IGD+I+SK I  R   Q+  + L+ +++  Y E L S FTIT GDEFQALL P+
Sbjct: 2   IYIAIIGDLISSKAITNRPKSQKQLKNLLNQINKKYKELLKSAFTITTGDEFQALLVPNP 61

Query: 73  KVFQIIDHIQLALKPVNVRFGLGTGNIITSINSNESIGADGPAYWHARSAINHIHDKNDY 132
           ++FQIID I L  KP  +RFG+G+G+I+T IN  +SIG+DGPAYWHAR+AI++IHDKNDY
Sbjct: 62  QIFQIIDEIALGFKPYQIRFGVGSGSILTEINPEQSIGSDGPAYWHARAAIDYIHDKNDY 121

Query: 133 GTVQVAICLDDEDQNLELTLNSLISAGDFIKSKWTTNHFQMLEHLILQDNYQEQFQHQKL 192
           G+  +A+ L+D + + +  +N++++A +FIKSKWT    +++++ L+    Y+E+F H+K+
Sbjct: 122 GSNHLAVDLEDTETSQQ--INAILAACEFIKSKWTVTQYEVIDGLLQAGIYEEKFSHKKM 179

Query: 193 AQLENIEPSALTKRLKASGLKIYLRTRTQAADLLVKSCTQTK 234
           A+  ++ PS+  KRLK+SGLKIYLR +  A  LL+ +  + K
Sbjct: 180 AEKLDLSPSSFNKRLKSSGLKIYLRNKKVATTLLLNAIRKEK 221
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2093> which encodes the amino acid sequence <SEQ ID 2094>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3744(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 94/213 (44%), Positives = 137/213 (64%), Gaps = 3/213 (1%)

Query: 14  YLALIGDIINSKQILERETFQQSFQQLMTELSDVYGEELISPFTITAGDEFQALLKPSKK 73
           Y+ALIGDII SKQ+ +R   Q++    + +L+  +   +IS  ++T GDEFQ L + +
Sbjct: 3   YIALIGDIIQSKQLTDRSKVQKTLAAYLDDLNKTFAPYIISKLSLTLGDEFQGLFQVDTP 62

Query: 74  VFQIIDHIQLALKPVNVRFGLGTGNIITSINSNESIGADGPAYWHARSAINHIHDKNDYG 133
           +F +ID I   +  + +RFG+G G+I+T IN + SIGADGPAYWHAR AI +IH KNDYG
Sbjct: 63  IFHLIDLINHHMD-IPIRFGVGVGSILTDINPDISIGADGPAYWHAREAIRYIHQKNDYG 121

Query: 134 TVQVAICLDDEDQNLELTLNSLISAGDFIKSKWTTNHFQMLEHLILQDNYQEQFQHQKLA 193
              +A  L     N +  LNSL++AGD IK+ W  + +++ L+    Y+E F  Q+L
Sbjct: 122 NTTLA--LRTGHHNQDDVLNSLLAAGDAIKANWRASQWEIFDTLLDLGIYEEYFDQQRLG 179
```

```
Query: 194 QLENIEPSALTKRLKASGLKIYLRTRTQAADLL 226
          +  ++  SAL+KRLK+S +KIYLRTR  A + L
Sbjct: 180 KQLSLSSSALSKRLKSSHVKIYLRTRQSALNCL 212
```

A related GBS gene <SEQ ID 8637> and protein <SEQ ID 8638> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:      4.96
GvH: Signal Score (-7.5): -5.46
     Possible site: 49
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -1.28 threshold:  0.0
   INTEGRAL    Likelihood = -1.28   Transmembrane    3 -  19 (   1 -  19)
  PERIPHERAL   Likelihood =  5.99       74
 modified ALOM score:   0.76

*** Reasoning Step: 3

----- 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>
```

SEQ ID 8638 (GBS338) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 5; MW 30kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 11; MW 55kDa).

GBS338-GST was purified as shown in Figure 215, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 680

A DNA sequence (GBSx0722) was identified in *S.agalactiae* <SEQ ID 2095> which encodes the amino acid sequence <SEQ ID 2096>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.6082(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 681

A DNA sequence (GBSx0723) was identified in *S.agalactiae* <SEQ ID 2097> which encodes the amino acid sequence <SEQ ID 2098>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have a cleavable N-term signal seq.
```

```
    INTEGRAL    Likelihood = -9.87   Transmembrane  126 - 142 ( 124 - 154)
    INTEGRAL    Likelihood = -8.23   Transmembrane   45 -  61 (  41 -  66)
    INTEGRAL    Likelihood = -5.10   Transmembrane  241 - 257 ( 236 - 257)
    INTEGRAL    Likelihood = -4.04   Transmembrane  199 - 215 ( 198 - 218)
    INTEGRAL    Likelihood = -0.22   Transmembrane   96 - 112 (  96 - 112)

----- 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>
```

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

```
>GP:AAG28337 GB:U88582 SatE [Streptococcus mutans]
 Identities = 54/103 (52%), Positives = 70/103 (67%), Gaps = 2/103 (1%)

Query: 1   MISDFLRDNPILTLLFCAHFLADFQWQSQSLADSKSHSWRGLWRHLLIVFLPLAALMILI 60
           +IS FL  NP+LTLL  AHFLADFQWQSQ +AD KS +W  L RHL+IV LPL  L ++I
Sbjct: 6   VISQFLSGNPVLTLLLIAHFLADFQWQSQKMADLKSSNWTYLIRHLIIVALPLILLSVVI 65

Query: 61  PETTLLNLSIWGSHIVIDSIKKLSYPWVEEGHF--QKAAFIID 101
           P + L+   I+ SH++IDS K L   + ++  F   KA F+ID
Sbjct: 66  PHSFLVLSLIFLSHVLIDSGKLLLNSFYKDRSFIKTKAVFLID 108
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2099> which encodes the amino acid sequence <SEQ ID 2100>. Analysis of this protein sequence reveals the following:

```
     Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.59   Transmembrane  125 - 141 ( 120 - 144)
   INTEGRAL    Likelihood = -6.58   Transmembrane  222 - 238 ( 215 - 238)
   INTEGRAL    Likelihood = -5.04   Transmembrane   47 -  63 (  45 -  77)
   INTEGRAL    Likelihood = -4.62   Transmembrane  179 - 195 ( 178 - 199)
   INTEGRAL    Likelihood = -0.43   Transmembrane   67 -  83 (  67 -  83)

----- 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>
```

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

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 109/256 (42%), Positives = 146/256 (56%), Gaps = 28/256 (10%)

Query: 2   ISDFLRDNPILTLLFCAHFLADFQWQSQSLADSKSHSWRGLWRHLLIVFLPLAALMILIP 61
           +S +L   P LTL    H L+D+Q QSQ +AD K      L  HL+ V +PL  L ++IP
Sbjct: 5   VSHYLAQTPTLTLFLICHVLSDYQLQSQQVADLKEKHLTYLGYHLIGVSIPLICLTLIIP 64

Query: 62  ETTLLNLSIWGSHIVIDSIKKL---SYPWVEEGHFQKAAFIIDQLAHYTCIIVFYHALPT 118
           +  L++L +  SH +ID +K     S  W  E       F++DQ  H          L
Sbjct: 65  QAWLMSLLVMISHALIDWLKPKMANSLKWKREW-----IFLLDQCLHIAISSFAGLRLAG 119

Query: 119 YLPPNHWLLPIKHFIVIALVFIIITKPINIVFKIFFNKFQAKELSSLLTQEKTKIMKEKS 178
              PN WL PI    ++  L  ++ITKP NIVFK+FF K+Q  +   +
Sbjct: 120 VTLPN-WL-PIS-ILMTVLFILLITKPTNIVFKLFFIKYQPDQGEKM------------ 163

Query: 179 EDHEETIEGAGAMIGNLERLIMAILLISGQYAAIGLVFTAKSIARYDKISKSQVFAEYYL 238
              +TI GAGA IG LER+++ + +I GQ+A+IGLVFTAKSIARY+KIS+S  FAEYYL
Sbjct: 164 ----DTIIGAGATIGILERIVIGVCMIMGQFASIGLVFTAKSIARYNKISESPAFAEYYL 219

Query: 239 IGSLFSIISVLITHWL 254
           IGSLFSI+SV I  W+
Sbjct: 220 IGSLFSILSVFIAAWI 235
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 682

A DNA sequence (GBSx0724) was identified in *S.agalactiae* <SEQ ID 2101> which encodes the amino acid sequence <SEQ ID 2102>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> 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>
```

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

```
>GP:AAD17886 GB:AF100456 hyaluronate-associated protein precursor
           [Streptococcus equi]
 Identities = 358/521 (68%), Positives = 426/521 (81%), Gaps = 2/521 (0%)

Query: 1   MSSFNRKKLKFLGISLATLTATTVTLVACGNESKNSGDNKV-INWYIPTEISTLDISKNT 59
           M+    K  K LG++  TL A+   L+ACGN+   S D K  INWY PTEI TLDISKNT
Sbjct: 1   MTVLGTKACKRLGLAAVTL-ASVAALMACGNKQSASTDKKSEINWYTPTEIITLDISKNT 59

Query: 60  DAYSNLAIGNSGSNLLRIDKEGKPKPDLAKKVSVSSDGLTYTATLRDNLKWSDGSKLSAE 119
           D YS LAIGNSGSNLLR D +GK +PDLA+KV VS DGLTYTATLRD LKWSDGS L+AE
Sbjct: 60  DTYSALAIGNSGSNLLRADAKGKLQPDLAEKVDVSEDGLTYTATLRDGLKWSDGSDLTAE 119

Query: 120 DFVYTWRRIVDPKTASEYAYLATESHLLNADKINSGDIKDLNKLGVTAKGNQVTFKLTSP 179
           DFVY+W+R+VDPKTASEYAYLATESHL NA+ INSG   DL+ LGV A GN+V F LT P
Sbjct: 120 DFVYSWQRMVDPKTASEYAYLATESHLKNAEDINSGKNPDLDSLGVKADGNKVIFTLTEP 179

Query: 180 CPQFKYYLAFSNFMPQKQSYVEKVGKDYGTTSKNQIYSGPYLVKDWNGSNGKFKLVKNKY 239
            PQFK  L+FSNF+PQK+S+V+  GKDYGTTS+ QIYSGPY+VKDWNG++G FKLVKNK
Sbjct: 180 APQFKSLLSFSNFVPQKESFVKDAGKDYGTTSEKQIYSGPYIVKDWNGTSGTFKLVKNKN 239

Query: 240 YWDSKHVKTNSVIVQTIKKPDTAVQMYKQGQIDFAEISGTSAIYQANKNNKDVVDASDAR 299
           YWD+K+VKT +V VQT+KKPDTAVQMYKQG++DFA ISGTSAIY ANK +KDVV   +A
Sbjct: 240 YWDAKNVKTETVNVQTVKKPDTAVQMYKQGKLDFANISGTSAIYNANKKHKDVVPVLEAT 299

Query: 300 TTYIIYNQTGSVKALTNQKIRQALNLATDRKGVVKAAVDTGSTPAESLVPKKLAKLPNGE 359
           T YI+YNQTG+++ L + KIRQALNLATDRKG+V AAVDTGS PA +LVP  LAKL +G
Sbjct: 300 TAYIVYNQTGAIEGLNSLKIRQALNLATDRKGIVSAAVDTGSKPATALVPTGLAKLSDGT 359

Query: 360 DLSKYTAPGYTYNTSKAQKLFKEGLAEVGQSSLKLTITADSDSPAAKNAVDYVKSTWESA 419
           DL+++ APGY Y+  +A KLFKEGLAE+G+ +L +TITAD+D+PAAK+AVDY+K TWE+A
Sbjct: 360 DLTEHVAPGYKYDDKEAAKLFKEGLAELGKDALTITITADADAPAAKSAVDYIKETWETA 419

Query: 420 LPGLTVEEKFVTFKQRLEDAKNENFDVVLFSWGGDYPEGSTFYGLFTTNSAYNYGKFSSK 479
           LPGLTVEEKFV FKQRLED KN+NF+V +  WGGDYP+GSTFYGLF + SAYNYGKF++
Sbjct: 420 LPGLTVEEKFVPFKQRLEDTKNQNFEVAVVLWGGDYPKGSTFYGLFKSGSAYNYGKFTNA 479

Query: 480 EYDNAYQKAITTDALKPGDAANDYKTAEKALFDQSYYNPVY 520
           +YD AY KA+TTDAL     AA+DYK AEKAL+D + YNP+Y
Sbjct: 480 DYDAAYNKALTTDALNTDAAADDYKAAEKALYDNALYNPLY 520
```

There is also homology to SEQ ID 318. An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 138/524 (26%), Positives = 222/524 (42%), Gaps = 73/524 (13%)

Query: 7   KKLKFLG-ISLATLTATTVTLVACGNESKNSGDN--KVINWYIPTEISTLDISKNTDAYS 63
           KK K+L  +S+A L+ +   L+ ACGN++ + G     K   +    +LD        +
Sbjct: 5   KKSKWLAAVSVAILSVSA--LAACGNKNASGGSEATKTYKYVFVNDPKSLDYILTNGGGT 62
```

```
Query: 64   NLAIGNSGSNLLRIDKEGKPKPDLAKKVSVSSDGLTYTATLRDNLKW--SDGSK---LSA 118
              I      LL  D+ G   P LAK   VS DGLTYT TLRD + W  +DG +   ++A
Sbjct: 63   TDVITQMVDGLLENDEYGNLVPSLAKDWKVSKDGLTYTYTLRDGVSWYTADGEEYAPVTA 122

Query: 119 EDFVYTWRRIVDPKTASEYAYLATESHLLNADKINSGDIKDLNKLGVTAKGNQ-VTFKLT 177
           EDFV   +  VD K+ + Y     E  + N     +G++ D  ++GV A  ++ V + L
Sbjct: 123 EDFVTGLKHAVDDKSDALY---VVEDSIKNLKAYQNGEV-DFKEVGVKALDDKTVQYTLN 178

Query: 178 SPCPQFKYYLAFSNFMPQKQSYVEKVGKDYGTTSKNQI-YSGPYLVKDWNGSNGKFKLVK 236
             P   +     +S   P    +++  GKD+GTT  + I  +G Y +  +  S    + K
Sbjct: 179 KPESYWNSKTTYSVLFPVNAKFLKSKGKDFGTTDPSSILVNGAYFLSAFT-SKSSMEFHK 237

Query: 237 NKYYWDSKHVKTNSV--IVQTIKKPDTAVQMYKQGQIDFAEISGTSAIYQ-ANKNNKDVV 293
           N+ YWD+K+V   SV         P +  + +G+   A +      Y+ A KN  D +
Sbjct: 238 NENYWDAKNVGIESVKLTYSDGSDPGSFYKNFDKGEFSVARLYPNDPTYKSAKKNYADNI 297

Query: 294 D----ASDARTTYIIYN---------------QTGSVKALTNQKIRQALNLATDRKG--- 331
               D R  ++ +N                 Q    KAL N+  RQA+  A DR
Sbjct: 298 TYGMLTGDIR--HLTWNLNRTSFKNTKKDPAQQDAGKKALNNKDFRQAIQFAFDRASFQA 355

Query: 332 ---------------VVKAAVDTGSTPAESLVPKKLAKL-PNGEDLSKYTAPGYTYNTS 374
                        V   V  G +   S V K++AKL    +D++   A    YN
Sbjct: 356 QTAGQDAKTKALRNMLVPPTFVTIGESDFGSEVEKEMAKLGDEWKDVNLADAQDGFYNPE 415

Query: 375 KAQKLF---KEGLAEVGQS-SLKLTITADSDSPAAKNAVDYVKSTWESALPGLTV----- 425
           KA+  F   KE L   G +  ++L    D  + A         K + E++L    V
Sbjct: 416 KAKAEFAKAKEALTAEGVTFPVQLDYPVDQANAATVQEAQSFKQSVEASLGKENVIVNVL 475

Query: 426 EEKFVTFKQR---LEDAKNENFDVVLFSWGGDYPEGSTFYGLFT 466
           E +  T + +    E  + +++D++   WG DY +  T+  + +
Sbjct: 476 ETETSTHEAQGFYAETPEQQDYDIISSWWGPDYQDPRTYLDIMS 519
```

SEQ ID 2102 (GBS323) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 4; MW 61.3kDa).

The GBS323-His fusion product was purified (Figure 209, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 306), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 683

A DNA sequence (GBSx0725) was identified in *S.agalactiae* <SEQ ID 2103> which encodes the amino acid sequence <SEQ ID 2104>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.54   Transmembrane  199 - 215 ( 198 - 215)

----- Final Results -----
              bacterial membrane --- Certainty=0.1617(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 GENPEPT database:

```
>GP:AAC17173 GB:AF065141 unknown [Streptococcus mutans]
 Identities = 304/356 (85%), Positives = 334/356 (93%)

Query: 1   MKRELLLEKIDELKEIMPWYVLEYYQSKLSVPYSFTTLYEYLKEYRRFLEWLLDSGVANC 60
           M+RELLLEKIDELKE+MPWYVLEYYQSKL+VPYSFTTLYEYLKEYRRF EWL+DSGV+N
Sbjct: 1   MRRELLLEKIDELKELMPWYVLEYYQSKLTVPYSFTTLYEYLKEYRRFFEWLIDSGVSNA 60
```

```
Query: 61  HHIAEIELSVLENLTKKDMEAFILYLRERPLLNANTRQNGVSQTTINRTLSALSSLFKYL 120
           + +A+I L  LE+L+KKDME+FILYLRER LLN   ++ GVSQTTINRTLSALSSL+KYL
Sbjct: 61  NKLADIPLETLEHLSKKDMESFILYLRERTLLNTKNKRQGVSQTTINRTLSALSSLYKYL 120

Query: 121 TEEVENADGEPYFYRNVMKKVSTKKKKETLASRAENIKQKLFLGNETIEFLEYIDCEYQN 180
           TEEVENADGEPYFYRNVMKKVSTKKKKETLA+RAENIKQKLFLGNET+EFLEY+DCEY+
Sbjct: 121 TEEVENADGEPYFYRNVMKKVSTKKKKETLAARAENIKQKLFLGNETMEFLEYVDCEYEQ 180

Query: 181 KLSKRALAFFNKNKERDLAIIALLLASGVRLSEAVNLDLKDINLNVMVIDVTRKGGKRDS 240
           KLSKRAL+ F KNKERDLAIIALLLASGVRLSEAVNLDLKD+NLN+M+I+VTRKGGK DS
Sbjct: 181 KLSKRALSSFRKNKERDLAIIALLLASGVRLSEAVNLDLKDVNLNMMIIEVTRKGGKHDS 240

Query: 241 VNVASFAKPYLANYLDIRKNRYKAENQDIALFLSEYRGVPNRIDASSVEKMVAKYSQDFK 300
           VNVA FAKPYL NY+ IR+ RYKA+  D+A FLSEYRGVPNR+DASS+EKMVAKYSQDFK
Sbjct: 241 VNVAGFAKPYLENYITIRRGRYKAKKTDLAFFLSEYRGVPNRMDASSIEKMVAKYSQDFK 300

Query: 301 VRVTPHKLRHTLATRLYDATKSQVLVSHQLGHASTQVTDLYTHIVNDEQKNALDKL 356
           +RVTPHKLRHTLATRLYDATKSQVLVSHQLGHASTQVTDLYTHIVNDEQKNALDKL
Sbjct: 301 IRVTPHKLRHTLATRLYDATKSQVLVSHQLGHASTQVTDLYTHIVNDEQKNALDKL 356
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2105> which encodes the amino acid sequence <SEQ ID 2106>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.54   Transmembrane  211 - 227 ( 210 - 227)

----- Final Results -----
               bacterial membrane --- Certainty=0.1617(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 9139> which encodes the amino acid sequence <SEQ ID 9140>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 60
>>> Seems to have no N-terminal signal sequence
     INTEGRAL    Likelihood = -1.54   Transmembrane  199 - 215 ( 198 - 215)

----- Final Results -----
               bacterial membrane --- Certainty= 0.162(Affirmative) < succ>
                bacterial outside --- Certainty= 0.000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 283/356 (79%), Positives = 321/356 (89%)

Query: 1   MKRELLLEKIDELKEIMPWYVLEYYQSKLSVPYSFTTLYEYLKEYRRFLEWLLDSGVANC 60
           M+RELLLEKI+  K IMPWYVL+YYQSKL+VPYSFTTLYEYLKEY+RF +WL+D+ +
Sbjct: 13  MRRELLLEKIETYKAIMPWYVLDYYQSKLAVPYSFTTLYEYLKEYKRFFDWLMDADLTQA 72

Query: 61  HHIAEIELSVLENLTKKDMEAFILYLRERPLLNANTRQNGVSQTTINRTLSALSSLFKYL 120
             IA+I+LS LE+LTKKD+EAF+LYLRERP LN  + + G+SQTTINRTLSALSSL+KYL
Sbjct: 73  PKIADIDLSTLEHLTKKDLEAFVLYLRERPSLNTYSTKEGLSQTTINRTLSALSSLYKYL 132

Query: 121 TEEVENADGEPYFYRNVMKKVSTKKKKETLASRAENIKQKLFLGNETIEFLEYIDCEYQN 180
           TEEVEN  GEPYFYRNVMKKVSTKKKKETLASRAENIKQKLFLG+ET+ FL+Y+D EY+
Sbjct: 133 TEEVENDQGEPYFYRNVMKKVSTKKKKETLASRAENIKQKLFLGDETLAFLDYVDKEYEQ 192

Query: 181 KLSKRALAFFNKNKERDLAIIALLLASGVRLSEAVNLDLKDINLNVMVIDVTRKGGKRDS 240
           KLS RA + F KNKERDLAIIALLLASGVRLSEAVNLDLKD+NLN+M+I+V RKGGKRDS
Sbjct: 193 KLSNRAKSSFRKNKERDLAIIALLLASGVRLSEAVNLDLKDVNLNMMIIEVIRKGGKRDS 252

Query: 241 VNVASFAKPYLANYLDIRKNRYKAENQDIALFLSEYRGVPNRIDASSVEKMVAKYSQDFK 300
```

```
            VNVA FAK YL +YL +R+ RYKAE QD+A FL+EYRGVPNR+DASS+EKMV KYS+DFK
Sbjct: 253  VNVAGFAKGYLESYLAVRQRRYKAEKQDLAFFLTEYRGVPNRMDASSIEKMVGKYSEDFK 312

Query: 301  VRVTPHKLRHTLATRLYDATKSQVLVSHQLGHASTQVTDLYTHIVNDEQKNALDKL 356
            +RVTPHKLRHTLATRLYDATKSQVLVSHQLGH+STQVTDLYTHIVNDEQKNALD L
Sbjct: 313  IRVTPHKLRHTLATRLYDATKSQVLVSHQLGHSSTQVTDLYTHIVNDEQKNALDNL 368
```

SEQ ID 2104 (GBS420) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 5; MW 68kDa).

GBS420-GST was purified as shown in Figure 219, lane 9-10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 684

A DNA sequence (GBSx0726) was identified in *S.agalactiae* <SEQ ID 2107> which encodes the amino acid sequence <SEQ ID 2108>. This protein is predicted to be a sensor-like histidine kinase in idh 3'region. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.75   Transmembrane   10 -  26 (   8 -  34)
   INTEGRAL    Likelihood = -3.93   Transmembrane   37 -  53 (  35 -  54)

----- 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>
```

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

```
>GP:CAB16001 GB:Z99124 similar to two-component sensor histidine
           kinase [YxdJ] [Bacillus subtilis]
 Identities = 96/320 (30%), Positives = 172/320 (53%), Gaps = 16/320 (5%)

Query: 2    IRQFLREHLIWYILYIM--MFVLFFISFYLYHLPMPYLFNSLGLNVIVLLGISIWQYSRY 59
            ++ FLR H +  +L+++  +FV F+  F   H    +LF  LG+ +++L G  +++ +
Sbjct: 1    MKLFLRSHAVLILLFLLQGLFVFFYYWFAGLH-SFSHLFYILGVQLLILAGYLAYRWYKD 59

Query: 60   RKKMLHLKYFNSSQDPSFELQPSDYAYFNIITQLEA--REAQKVSETIEQTNHVALMIKM 117
            R     L       D  + L  S +       Q+E    + QK+ ET  + +     +
Sbjct: 60   RGVYHWLSSGQEGTDIPY-LGSSVFCSELYEKQMELIRLQHQKLHETEAKLDARVTYMNQ 118

Query: 118  WSHQMKVPLAAISLMAQTNHLDP--KEVEQQLLKLQHYLETLLAFLKFRQYRDDFRFEAV 175
            W HQ+K PL+ I+L+ Q    +P  +++++++ +++  LETLL   +   + DF+ EAV
Sbjct: 119  WVHQVKTPLSVINLIIQEED-EPVFEQIKKEVRQIEFGLETLLYSSRLDLFERDFKIEAV 177

Query: 176  SLREVVVEIIKSYKVICLSKSL--SIIIEGDNIWKTDKKWLTFALSQVLDNAIKYSNPES 233
            SL E++  +I+SYK   +   + + +  D+   TD KWL FA+ QV+ NA+KYS  +S
Sbjct: 178  SLSELLQSVIQSYKRFFIQYRVYPKMNVCDDHQIYTDAKWLKFAIGQVVTNAVKYSAGKS 237

Query: 234  -----KIIISIGEESIRIQDYGIGILEEDIPRLFEDGFTGYNGHEHQKATGMGLYMTKEV 288
                 +        + ++DYG+GI  +DI R+F+  +TG NG   Q++TG+GL++ KE+
Sbjct: 238  DRLELNVFCDEDRTVLEVKDYGVGIPSQDIKRVFDPYYTGENGRRFQESTGIGLHLVKEI 297

Query: 289  LSSLNLSISVDSKINYGTAV 308
               LN ++ + S    GT+V
Sbjct: 298  TDKLNHTVDISSSPGEGTSV 317
```

SEQ ID 2108 (GBS421) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 6; MW 63kDa).

GBS421-GST was purified as shown in Figure 219, lane 11.

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

Example 685

A DNA sequence (GBSx0727) was identified in *S.agalactiae* <SEQ ID 2111> which encodes the amino acid sequence <SEQ ID 2112>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1310(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 GENPEPT database:

```
>GP:AAD10258 GB:AF036964 putative response regulator [Lactobacillus
          sakei]
 Identities = 94/222 (42%), Positives = 140/222 (62%), Gaps = 8/222 (3%)

Query: 7   KIYIVEDDMTIVSLLKDHLSASYHVSSV--SNFRDVKQEIIAFQPDLILMDITLPYFNGF 64
           +I IVEDD TI +L+ ++L   + + ++   +F  +    +  +P L+L+DI LP ++GF
Sbjct: 3   EIMIVEDDPTIANLIAENLE-KWQLKAIIPDDFDTIFDRFLTDKPHLVLLDINLPVYDGF 61

Query: 65  YWTAELRKFLTIPIIFISSSNDEMDMVMALNMGGDDFISKPFSLAVLDAKLTAILRRSQQ 124
           YW  ++R+   +PIIFISS +  MDMVM++NMGGDDF++KPFS+ VL AK+ A+LRR+
Sbjct: 62  YWCRKIREVSKVPIIFISSRSTNMDMVMSMNMGGDDFVNKPFSMEVLIAKINALLRRTYN 121

Query: 125 FIQQE---LTFGGFTLT-REGLLSSQDKEVILSPTENKILSILLMHPKQVVSKESLLEKL 180
           ++ Q    +   G  +  + G     D  V LS  E K+L  L+    Q+VS+E LL  L
Sbjct: 122 YVDQNTDVIEHNGLLINLQSGGAQVGDTVVDLSKNEYKLLQFLMRQHGQIVSREKLLRAL 181

Query: 181 WENDSFIDQNTLNVNMTRLRKKIVPIGF-DYIHTVRGVGYLL 221
           W+++ F+D NTL VN+ RLRKKI   G  DYI T  G GY++
Sbjct: 182 WDDERFVDDNTLTVNINRLRKKIEQAGLEDYIQTKIGQGYII 223
```

There is also homology to SEQ ID 1182.

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

Example 686

A DNA sequence (GBSx0728) was identified in *S.agalactiae* <SEQ ID 2113> which encodes the amino acid sequence <SEQ ID 2114>. This protein is predicted to be permease OrfY. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.62   Transmembrane   55 -  71 (  49 -  75)
   INTEGRAL    Likelihood =-10.30   Transmembrane  197 - 213 ( 192 - 218)
   INTEGRAL    Likelihood = -9.13   Transmembrane  152 - 168 ( 141 - 172)
   INTEGRAL    Likelihood = -8.70   Transmembrane  624 - 640 ( 619 - 645)
   INTEGRAL    Likelihood = -8.44   Transmembrane  222 - 238 ( 219 - 250)
   INTEGRAL    Likelihood = -7.75   Transmembrane  283 - 299 ( 280 - 307)
   INTEGRAL    Likelihood = -7.70   Transmembrane  533 - 549 ( 526 - 552)
   INTEGRAL    Likelihood = -6.95   Transmembrane  108 - 124 (  99 - 140)
   INTEGRAL    Likelihood = -4.88   Transmembrane  585 - 601 ( 581 - 610)
   INTEGRAL    Likelihood = -3.82   Transmembrane   25 -  41 (  21 -  47)
```

```
    INTEGRAL    Likelihood = -0.48   Transmembrane  602 - 618 ( 602 - 618)

----- Final Results -----
               bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9809> which encodes amino acid sequence <SEQ ID 9810> was also identified.

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

```
>GP:AAF99695 GB:AF267498 permease OrfY [Streptococcus mutans]
 Identities = 154/665 (23%), Positives = 299/665 (44%), Gaps = 40/665 (6%)

Query: 4   MFYLKIAWHNLKHSIDQYIPFLLASLLLYSLTCSTLLILMSAVGRDMGTAAT---VLFLG 60
           MF  KI++HNL  +    +P+     +        +  L +    ++ TA     +L  G
Sbjct: 1   MFLPKISFHNLIVNKSLTLPYFAIMTIFSGFNYVLINFLTNPSFYNIPTARILIDILIFG 60

Query: 61  VIVLSIFAVVMEHYSYNILMKQRSSEFGLYNILGMNKRQVARVASLELFIIYIFLISIGS 120
            I++S+  ++   Y+   + +R+S  G++ +LGM K+Q+ ++  LE   ++      G
Sbjct: 61  FILISLLMLLYGRYANRFISDERNSNMGIFLMLGMGKKQLLKIIYLEKLYLFTGTFFGGL 120

Query: 121 LFSAFFAKFIYLIFVNIINYHALNLSLSLWPFIICIVIFTGIFLTLEVPVIRHVHLSSPL 180
           +F   ++K  +L   N+I    +   SL      +++   I+  + +   R +   S
Sbjct: 121 IFGFVYSKIFFLFIRNLIVIGDVREQYSLTAISWLLILTFFIYFIIYLSEYRLLKRQSIT 180

Query: 181 SLFRKKQQGEKEPKGNLILAILALVAIAIAYTMALTSGKAPALAVIY-RFFFAVLLVIAG 239
            +F  K + +   K ++ + +  L A+ + Y  ALTS   P +   + RF +A  LV  G
Sbjct: 181 VIFNSKAKRDNPRKTSVFVGLFGLFALLMGYHFALTS---PNVTTSFSRFIYAACLVTLG 237

Query: 240 TYLFYISFMTWYLKRLRQNKHYYYKSEHFVSTSQMIFRMKQNAVGLASITLLAVMALVTI 299
            + + S +   L  +++ +  YY    FV  + +  R++ NA+ LA+I + +   LV++
Sbjct: 238 IFCTFSSGVIMLLTVIKKRRAIYYNQRRFVVIASLFHRIRSNALSLATICIFSTATLVSL 297

Query: 300 ATTVSLYSNTQNVVTGLFPKSVSLSIDNSKGDAKNIFEEKILKKLGKSSKEAITYNQTMI 359
           +   SLY    N+V    P+ V++    S  D      E  L  +   +T  Q
Sbjct: 298 SVLASLYLAKDNMVRLSSPRDVTVL---STTDI-----EPNLMDIATKNHVTLTNRQ--- 346

Query: 360 SMPVSQSSELNITSKNVKHVDITKTGFMY------LITQNDFRRLGHQLPKLKDNQVAYF 413
           ++ VSQS   NI      H+ +   G M       +I+ + F    +LK++++  +
Sbjct: 347 NLKVSQSVYGNIKGS---HLSVDPNGGMANDYQITVISLDSFNASNNTHYRLKNHEILTY 403

Query: 414 VQKGDSRLKKINLLGNKFDVVKNLKEA-YVPETTNTYNPGLIIFANNKQI-DNIRKAYLP 471
           V  G +       G K   VK +K   ++    P   I  +N++I   I K  L
Sbjct: 404 VSNGAAAPSSYTTNGVKLTNVKQIKRINFIFSPLRSMQPNFFIITDNREIIQTILKEELT 463

Query: 472 YTKNINTFPKTFKAYLDLNSQEINSISKNDIIEVDG--KYVGNISTKQSFLKEGYQMFGG 529
           +       T   Y  +  +++N     D +E     ++  N+ + +        +FGG
Sbjct: 464 WG--------TMAGY-HVKGKKMNQKDFYDELETTNFRQFSANVVSIRQVKSMFNALFGG 514

Query: 530 LLFTGFLLGISFLLGIALIVYYKQYSEGHEDKRSYRILQEVGMSKKLVKRTINSQIMIFF 589
           LLF G + G  F +  A+ +YY+Q SEG  D+  Y+ + ++GM+ K ++ +I  QI   F
Sbjct: 515 LLFVGIIFGTIFAILTAITIYYQQLSEGIRDRDDYKAMIKLGMTNKTIQDSIKVQINFVF 574

Query: 590 FQPLVVAVIHFGVAIPMLKQMLLVFGVLNSTIVYVVSGLTVLAISIIYFIIYRITSRTYY 649
             P+  A+++   A+P+L +++  FG  ++ +     G  ++     Y+ I   TS+ YY
Sbjct: 575 ILPIAFALLNLIFALPILYKIMTTFGFNDAGLFLRAVGTCLIVYLFFYWFICHCTSKLYY 634

Query: 650 HIIER 654
            +I +
Sbjct: 635 RLISK 639
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2115> which encodes the amino acid sequence <SEQ ID 2116>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-13.59   Transmembrane  602 - 618 ( 592 - 630)
   INTEGRAL    Likelihood =-12.26   Transmembrane   59 -  75 (  50 -  81)
   INTEGRAL    Likelihood =-12.21   Transmembrane  235 - 251 ( 224 - 262)
   INTEGRAL    Likelihood = -9.82   Transmembrane  159 - 175 ( 146 - 177)
   INTEGRAL    Likelihood = -9.02   Transmembrane  201 - 217 ( 198 - 223)
   INTEGRAL    Likelihood = -8.97   Transmembrane  510 - 526 ( 507 - 540)
   INTEGRAL    Likelihood = -6.42   Transmembrane  569 - 585 ( 564 - 589)
   INTEGRAL    Likelihood = -5.95   Transmembrane  109 - 125 ( 102 - 138)
   INTEGRAL    Likelihood = -4.09   Transmembrane  294 - 310 ( 290 - 315)
   INTEGRAL    Likelihood = -1.86   Transmembrane  126 - 142 ( 126 - 142)

----- Final Results -----
               bacterial membrane --- Certainty=0.6434(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:

```
>GP:BAB03337 GB:AB035452 ABC transporter [Staphylococcus aureus]
 Identities = 141/657 (21%), Positives = 289/657 (43%), Gaps = 66/657 (10%)

Query: 5   ITKSNIKKNFSLYRIYFLATIGLLSIFIAFLNFISDKII--TEKIGDSGQALVIANGSL- 61
           I   N+++N   Y +Y      L S+F + + + S   +  T+ + +     +I  G+L
Sbjct: 6   IVFKNLRQNLKHYAMY------LFSLFFSIVLYFSFTTLQFTKGVNNDDSMAIIKKGALV 59

Query: 62  --IFLIVFLVVFLIYFNNFFVKKRSQELGVLAILGFSKRELTKLLTLENLVILVLSYLVS 119
             IFL + +V+FL+Y N+ FVK+R++E  +  ++G +++ + K+L LE +++ +++ +V
Sbjct: 60  GSIFLFIIIVIFLMYANHLFVKRRTREFALFQLIGLTRQNILKMLALEQMIVFLITGVVG 119

Query: 120 LLLGPTLYFLAVLAITHLLNLTMEVQWFITVNEIIESLGILVVVFLINVITNGLIISKQS 179
           +L G     L +  ++ L++L++ +         ++ ++ +L++ +++ +  + L + ++S
Sbjct: 120 VLCGIAGAQLLLSIVSKLMSLSINLSIHFEPMALVLTIFMLIIAYVLILFQSALFLKRRS 179

Query: 180 LIEFVNFSRKAE----KKIKIRKVRAIIAITALLLSYILCLATVFSSTRNMLLSIGMVPV 235
           ++  +  S K +    K      +  ++ I  + L Y   +AT    T   L      P
Sbjct: 180 ILSMMKDSIKTDATTAKVTTAEVISGVLGIAMIALGYY--MATEMFGTFKALTMAMTSP- 236

Query: 236 SLLIIVLVVLGTVFTIRYGLAFVVSLLKENKKRLYRPLSNIIYPKFNYRIATKNKLLTVL 295
             +I+ L V+G     R  ++ +  LK++K          +      YR+      LT++
Sbjct: 237 -FIILFLTVVGAYLFFRSSVSLIFKTLKKSKNGRVSITDVVFTSSIMYRMKKNAMSLTII 295

Query: 296 GGLLTVTVSVAGMMVMLYAYSLNGIERLTPSAIEYNVESENGQVNVTTILENDQVSL--- 352
             + VTV+V     +  + +   + P+  E+NV +         T L   Q++
Sbjct: 296 AIISAVTVTVLCFAALSKSNTDQTLTSMAPN--EFNVVATQDAKQFETKLSQQQITFSKN 353

Query: 353 ----VDVGLLRLNTIPEVTITDSGQTIPYFDIINYSDYKELMKAQGRTNSIEGSKSLPLL 408
               + V  ++   I     +DSG+T       N        K  G    I  +KSLP +
Sbjct: 354 AYETITVDNVKDQVITLENGSDSGRTNSILSANN--------KVTGNNAIITNTKSLPNI 405

Query: 409 INYYPTEISLGKTFNLGNAYDVT--VKQVSTNNVFSFSTSVTTLV--VSDKLYAKLSSRF 464
           IN     I L K   +    + T  V Q     V+  + S  + V  VS + Y +L +
Sbjct: 406 IN-----IHLNKDLVVKGTKNETFRVTQEDKGRVYPLNLSFNSPVVEVSPEKYQQLKT-- 458

Query: 465 PEKEMTIRTFNGTSIR------SSEAFYNQFSMVPDVISSYSKEHTVKTANIATYIFIT- 517
             +  + TF G  I+       ++A   QF    D + +Y +      A     IF+T
Sbjct: 459 ---QNNVHTFYGYDIKQTSQKEKAQAIAKQFG---DKVITYDEMKKEVDATNGILIFVTS 512

Query: 518 FLSILFIICTGSILYFTSLIEIMENKEEYGYLSKLGYSKKMIHRILRYETGILFLIPVFI 577
           FL + F++  G I+Y   + E  +    + L ++G++   + + L  +    F +P+ I
Sbjct: 513 FLGLAFLVAAGCIIYIKQMDETEDELSNFRILKRIGFTHTDMLKGLLLKITFNFGLPLLI 572

Query: 578 GIVNGGMLLIYYKYLFMDTLVAGNIIMLSLLLCLLFFLIIYGTFYVLTLRLVTSIIK 634
            I++     I +  L       GNI  + +++ ++ + +IY TF ++       +IK
Sbjct: 573 AILHAVFAAIAFMKLM------GNISFMPVIVVIVVYTLIYITFALIAFVHSNKLIK 623
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 145/678 (21%), Positives = 277/678 (40%), Gaps = 89/678 (13%)

Query: 13  NLKHSIDQYIPFLLASLLLYSLTCSTL-----LILMSAVGRDMGTAATVLFLGVIVLSIF 67
           N+K +   Y  + LA++ L S+  + L      I+   +G D G A  +    +I L +F
Sbjct: 9   NIKKNFSLYRIYFLATIGLLSIFIAFLNFISDKIITEKIG-DSGQALVIANGSLIFLIVF 67

Query: 68  AVVMEHYSYNILMKQRSSEFGLYNILGMNKRQVARVASLELFIIYIFLISIGSLFSAFFA 127
            VV   Y  N  +K+RS E G+  ILG +KR++ ++ +LE  +I +    +  L S
Sbjct: 68  LVVFLIYFNNFFVKKRSQELGVLAILGFSKRELTKLLTLENLVILV----LSYLVSLLLG 123

Query: 128 KFIYLIFVNIIINYHALNLSLSLWPFIICIVIFTGIFLTLEVPVIRHV------HLSSPLS 181
             +Y + V  I  H LNL++ +  FI    I   + + + V +I  +         S +
Sbjct: 124 PTLYFLAVLAIT-HLLNLTMEVQWFITVNEIIESLGILVVVFLINVITNGLIISKQSLIE 182

Query: 182 LFRKKQQGEKEPKGNLILAILALVAIAIAYTMAL------TSGKAPALAVIYRFFFAVLL 235
                ++ EK+ K   + AI+A+ A+ ++Y + L      T     ++ ++      ++L
Sbjct: 183 FVNFSRKAEKKIKIRKVRAIIAITALLLSYILCLATVFSSTRNMLLSIGMVPVSLLIIVL 242

Query: 236 VIAGTYLFYISFMIWYLKRLRQNKHYYYKSEHFVSTSQMIFRMKQNAVGLASITLLAVMA 295
           V+ GT       + + +  L++NK   Y+    +   + +R+   A     +T+L  +
Sbjct: 243 VVLGTVFTIRYGLAFVVSLLKENKKRLYRPLSNIIYPKFNYRI---ATKNKLLTVLGGLL 299

Query: 296 LVTIATT---VSLYSNTQNVVTGLFPKSVSLSIDNSKGDAKNIFEEKILKKLGKSSKEAI 352
            VT++     V LY+ + N +  L P ++  +++  G                +  I
Sbjct: 300 TVTSVSVAGMMVMLYAYSLNGIERLTPSAIEYNVESENGQV---------------NVTTI 344

Query: 353 TYNQTMISMPVSQSSELNITSKNVKHVDITKTG----FMYLITQNDFRRL------GHQL 402
             N  + + V       +    + V IT +G    +  +I  +D++ L       + +
Sbjct: 345 LENDQVSLVDVGL-----LRLNTIPEVTITDSGQTIPYFDIINYSDYKELMKAQGRTNSI 399

Query: 403 PKLKDNQVAYFVQKGDSRLKKINLLGNKFDVVKNLKEAYVPETTNTYNPGLIIFANNKQI 462
              K  +      + L K   LGN +DV   +K+       +       +  ++K
Sbjct: 400 EGSKSLPLLINYYPTEISLGKTFNLGNAYDVT--VKQVSTNNVFSFSTSVTTLVVSDKLY 457

Query: 463 DNIRKAYLPYTKNINTFPKT-------FKAYLDLNSQEINSISKNDIIEVDGKYVGNIST 515
             +   +      I TF  T       F     +  I+S SK   ++       NI+T
Sbjct: 458 AKLSSRFPEKEMTIRTFNGTSIRSSEAFYNQFSMVPDVISSYSKEHTVKT-----ANIAT 512

Query: 516 KQSFLKEGYQMFGGLLFTGFLLGISFLLGIALIVYYKQYSEGHEDKRSYRILQEVGMSKK 575
                      +F  FL  I F++    I+Y+    E  E+K  Y  L ++G SKK
Sbjct: 513 --------------YIFITFL-SILFIICTGSILYFTSLIEIMENKEEYGYLSKLGYSKK 557

Query: 576 LVKRTINSQIMIFFFQPLVVAVIHFGVAIPMLKQMLLVFGVLNSTIVYVVSGLTVLAISI 635
           ++ R +  +  I F  P+ + +++ G+ +   K  L +  ++   I+ +   L +L  I
Sbjct: 558 MIHRILRYETGILFLIPVFIGIVNGGMLLIYYK-YLFMDTLVAGNIIMLSLLLCLLFFLI 616

Query: 636 IYFIIYRITSRTYYHIIE 653
           IY   Y +T R    II+
Sbjct: 617 IYGTFYVLTLRLVTSIIK 634
```

A related GBS gene <SEQ ID 8639> and protein <SEQ ID 8640> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:    -11.64
GvH: Signal Score (-7.5): -3.52
     Possible site: 37
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 11 value: -11.62 threshold:  0.0
   INTEGRAL    Likelihood =-11.62   Transmembrane   55 -  71 (  49 -  75)
   INTEGRAL    Likelihood =-10.30   Transmembrane  197 - 213 ( 192 - 218)
   INTEGRAL    Likelihood = -9.13   Transmembrane  152 - 168 ( 141 - 172)
   INTEGRAL    Likelihood = -8.70   Transmembrane  624 - 640 ( 619 - 645)
   INTEGRAL    Likelihood = -8.44   Transmembrane  222 - 238 ( 219 - 250)
   INTEGRAL    Likelihood = -7.75   Transmembrane  283 - 299 ( 280 - 307)
```

```
    INTEGRAL    Likelihood = -7.70   Transmembrane  533 - 549 ( 526 - 552)
    INTEGRAL    Likelihood = -6.95   Transmembrane  108 - 124 (  99 - 140)
    INTEGRAL    Likelihood = -4.88   Transmembrane  585 - 601 ( 581 - 610)
    INTEGRAL    Likelihood = -3.82   Transmembrane   25 -  41 (  21 -  47)
    INTEGRAL    Likelihood = -0.48   Transmembrane  602 - 618 ( 602 - 618)
    PERIPHERAL  Likelihood =  1.16     129
 modified ALOM score:   2.82

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5649(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:

```
ORF02245(310 - 2262 of 2562)
GP|9802356|gb|AAF99695.1|AF267498_5|AF267498(1 - 639 of 640) permease OrfY {Streptococcus
mutans}
%Match = 10.2
%Identity = 24.0  %Similarity = 49.8
Matches = 147  Mismatches = 297  Conservative Sub.s = 158

123       153       183       213       243       273       303       333
QKTC*IYLKLLTWMDKLF*W*PIQQMLLVMPNAFYLSKMDVFFTNFIVVIRIIANSIKIFL*QCLPY*GVNNMFYLKIAW
                                                                    ||  ||::
                                                                    MFLPKISF

363       393       423       453       474       504       534       564
HNLKHSIDQYIPFLLASLLLYSLTCSTLLILMSAVGRDMGTAAT---VLFLGVIVLSIFAVVMEHYSYNILMKQRSSEFG
|||  :    :|::   ::  :   :  | :   :: ||    :| :| |::|:: ::   |:   :  :|:|  |
HNLIVNKSLTLPYFAIMTIFSGFNYVLINFLTNPSFYNIPTARILIDILIFGFILISLLMLLYGRYANRFISDERNSNMG
          20        30        40        50        60        70        80

594       624       654       684       714       744       774       804
LYNILGMNKRQVARVASLELFIIYIFLISIGSLFSAFFAKFIYLIFVNIINYHALNLSLSLWPFIICIVIFTGIFLTLEV
:: :||| |:|: ::  || : ::      | :|   ::|  :|   |:|    :   ||      :::   |:: : :
IFLMLGMGKKQLLKIIYLEKLYLFTGTFFGGLIFGFVYSKIFFLFIRNLIVIGDVREQYSLTAISWLLILTFFIYFIIYL
         100       110       120       130       140       150       160

834       864       894       924       954               1011      1041
PVIRHVHLSSPLSLFRKKQQGEKEPKGNLILAILALVAIAIAYTMALTSGKAPALAVIY-RFFFAVLLVIAGTYLFYISF
   | :   |   :|  | : :   | :: : :: | |: : |  ||||   | :   : || :|  ||  | : : |
SEYRLLKRQSITVIFNSKAKRDNPRKTSVFVGLFGLFALLMGYHFALTS---PNVTTSFSRFIYAACLVTLGIFCTFSSG
         180       190       200       210          220       230       240

1071      1101      1131      1161      1191      1221      1251      1281
MTWYLKRLRQNKHYYYKSEHFVSTSQMIFRMKQNAVGLASITLLAVMALVTIATTVSLYSNTQNVVTGLFPKSVSLSIDN
:   |  ::: :  ||    ||  : :  |:: ||: ||:| :::   ||:::   |||    |:|    |: |
VIMLLTVIKKRRAIYYNQRRFVVIASLFHRIRSNALSLATICIFSTATLVSLSVLASLYLAKDNMVRLSSPRDV------
            260       270       280       290       300       310

1311      1341      1371      1401      1431      1461
SKGDAKNIFEEKILKKLGKSSKEAITYNQTMISMPVSQSSELNITSKNVKHVDITKTGFM--------------------
                              |::|    : : ::| :||  || :|   :
----------------------------TVLSTTDIEPNLMDIATKN--HVTLTNRQNLKVSQSVYGNIKGSHLSVDPN
                          320       330       340       350       360

1464      1494      1524      1554      1584      1614      1641      1671
---------YLITQNDFRRLGHQLPKLKDNQVAYFVQKGDSRLKKINLLGNKFDVVKNLKEA-YVPETTNTYNPGLIIFA
         :|: : |    :   :||:::: :|  | :        | |:  || :|   ::     :  | : |
GGMANDYQITVISLDSFNASNNTHYRLKNHEILTYVSNGAAAPSSYTTNGVKLTNVKQIKRINFIFSPLRSMQPNFFIIT
         380       390       400       410       420       430       440

1698      1728      1758      1788      1818      1842      1872      1902
NNKQI-DNIRKAYLPYTKNINTFPKTFKAYLDLNSQEINSISKNDIIEVDG--KYVGNISTKQSFLKEGYQMFGGLLFTG
:|::|   | |  | :         |   |  :  :::|     | :|     ::  |: : :        :||||||
```

```
DNREIIQTILKEELTWG--------TMAGY-HVKGKKMNQKDFYDELETTNFRQFSANVVSIRQVKSMFNALFGGLLFVG
         460               470        480        490        500        510

1932       1962       1992       2022       2052       2082       2112       2142
FLLGISFLLGIALIVYYKQYSEGHEDKRSYRILQEVGMSKKLVKRTINSQIMIFFFQPLVVAVIHFGVAIPMLKQMLLVF
 ::|  | :  |: :||:| |||  |:  |: : ::||: | :: :|  ||   |  |:  |:::: |:|:| :::  |
IIFGTIFAILTAITIYYQQLSEGIRDRDDYKAMIKLGMTNKTIQDSIKVQINFVFILPIAFALLNLIFALPILYKIMTTF
        530        540        550        560        570        580        590

2172       2202       2232       2262       2292       2322       2352       2382
GVLNSTIVYVVSGLTVLAISIIYFIIYRITSRTYYHIIER*KGLVILPILLH**KPID*KICYTK*KKEISYYFRRGYVT
|  :: :     |  ::     |: |   ||: || :| :
GFNDAGLFLRAVGTCLIVYLFFYWFICHCTSKLYYRLISKK
        610        620        630        640
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 687

A DNA sequence (GBSx0729) was identified in *S.agalactiae* <SEQ ID 2117> which encodes the amino acid sequence <SEQ ID 2118>. This protein is predicted to be ABC transporter OrfX. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5121(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 GENPEPT database:

```
>GP:AAF99694 GB:AF267498 ABC transporter OrfX [Streptococcus mutans]
 Identities = 118/242 (48%), Positives = 175/242 (71%), Gaps = 1/242 (0%)

Query: 5   INHLEKVFRTRFSKEETRALQDVDFKVEQGEFIAIMGESGSGKTTLLNILATLEKPTNGQ 64
           ++HL+KV++T+        AL+D+ F V++GEFIAIMGESGSGK+TLLNILA ++ P++G
Sbjct: 6   VSHLKKVYKTQEGLTN-EALKDITFSVQEGEFIAIMGESGSGKSTLLNILACMDYPSSGH 64

Query: 65  VILNGEDITKIKEAKLASFRLKNLGFVFQDFNLLDTLSVRDNIYLPLVLDRKRYKEMDHR 124
           +I N    + K+K+ + A FR +++GF+FQ+FNLL+  + +DN+ +P+++    +    + R
Sbjct: 65  IIFNNYQLEKVKDEEAAVFRSRHIGFIFQNFNLLNIFNNKDNLLIPVIISGSKVNSYEKR 124

Query: 125 LSELSSHLRIDDLLDKRPFELSGGQKQRVAIARSLITNPQILLADEPTAALDYRNSEDLL 184
           L +L++ + I+ LL K P+ELSGGQ+QR+AIAR+LI NP ++LADEPT  LD + S+ +L
Sbjct: 125 LRDLAAVVGIESLLSKYPYELSGGQQQRLAIARALIMNPDLILADEPTGQLDSKTSQRIL 184

Query: 185 NLFETINLDGQTILMVTHSANAASHAKRVLFIKDGRIFHQLYRGNKNNSEFNKDISLTMS 244
           NL   IN   +TILMVTHS  AAS+A RVLFIKDG IF+QL RG K+   F   I +  +
Sbjct: 185 NLLSNINAKRKTILMVTHSPKAASYANRVLFIKDGVIFNQLVRGCKSREGFLDQIIMAQA 244

Query: 245 AI 246
           ++
Sbjct: 245 SL 246
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2119> which encodes the amino acid sequence <SEQ ID 2120>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
bacterial cytoplasm --- Certainty=0.2131(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 91/222 (40%), Positives = 142/222 (62%), Gaps = 2/222 (0%)

Query: 2   LLEINHLEKVFRTRFSKEETRALQDVDFKVEQGEFIAIMGESGSGKTTLLNILATLEKPT 61
           LL +  + K +       EE   L+ +D +V +G+F+AIMG SGSGK+TL+NI+  L+KP
Sbjct: 1   LLNLKDIRKSYH--LGTEEFAILKGIDLEVNEGDFLAIMGPSGSGKSTLMNIIGCLDKPG 58

Query: 62  NGQVILNGEDITKIKEAKLASFRLKNLGFVFQDFNLLDTLSVRDNIYLPLVLDRKRYKEM 121
           +G   + G D++ + + +LA  R + +GFVFQ+FNL+  L+   N+ LPL        KE
Sbjct: 59  SGSYAIEGRDVSSLSDNELADLRNQKIGFVFQNFNLMPKLTACQNVELPLTYMNVPKKER 118

Query: 122 DHRLSELSSHLRIDDLLDKRPFELSGGQKQRVAIARSLITNPQILLADEPTAALDYRNSE 181
             R  E+   + +++  + +P ELSGGQKQRVAIAR+L+TNP  +L DEPT ALD + S
Sbjct: 119 RKRALEMLKLVGLEERSEFKPMELSGGQKQRVAIARALVTNPSFILGDEPTGALDTKTSV 178

Query: 182 DLLNLFETINLDGQTILMVTHSANAASHAKRVLFIKDGRIFH 223
            +++LF+  N +G+TI+++TH    A+  K+ + ++DG I H
Sbjct: 179 QIMDLFKQFNDNGKTIIIITHEPEVAALCKKTVILRDGNIEH 220
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 688

A DNA sequence (GBSx0730) was identified in *S.agalactiae* <SEQ ID 2121> which encodes the amino acid sequence <SEQ ID 2122>. This protein is predicted to be nisin-resistance protein. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.16   Transmembrane    8 -  24 (   1 -  31)

----- 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>
```

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

```
>GP:AAB08491 GB:U25181 nisin-resistance protein [Lactococcus lactis]
 Identities = 108/318 (33%), Positives = 190/318 (58%), Gaps = 8/318 (2%)

Query: 3   RKIVLLFVVPMLIVLGILGVVVHYYGSALNIYLLPPSSERYGRVILDRVEQRGLYSQGRQ 62
           ++I+L  V    + LGI    ++++G   NIYL+PPS ++Y RV L  +++ GL++  ++
Sbjct: 5   KRILLGLVAVCALFLGI----IYFWGYKFNIYLVPPSPQKYVRVALKNMDELGLFTDSKE 60

Query: 63  WQIIRQRSEKKLKTSKSYQESRNIVQEAVRYGGGKHSQILSKETVRRDTLDSRYPEYRRL 122
           W   ++++ ++   +K+Y E+   +Q+A++  GGKHS I  +E + + ++          +
Sbjct: 61  WVETKKKTIEETSNAKNYAETIPFLQKAIKVAGGKHSFIEHEEDISKRSITKYIKPKAEI 120

Query: 123 NEDILLITIPSISKLDKRSISHYSGKLQNILMEKSYKGLILDLSNNTGGNMIPMIGGVAS 182
            + L++TIP  + D ++ S Y+  L++   + +Y G+I+DL  N GG++ PM+ G++
Sbjct: 121 EGNTLILTIPEFTGNDSQA-SDYANFLESSFHKNNYNGVIVDLRGNRGGDLSPMVLGLSP 179

Query: 183 ILPNDTLFHYTDKYGNKKTITMKNIPLEALKISRKTINTKHV---PIAIITNHKTASSAE 239
           +LP+ TLF Y DK  + K + ++N  + +  S K  + K +   PIA++ T  SS E
Sbjct: 180 LLPDGTLFTYVDKSSHSKPVELQNGEINSGGSSTKVSDNKKIKKAPIAVLIDNNTGSSGE 239

Query: 240 MTFLSFKGLPNVKSFGQATAGYTTVNETFMLYDGARLALTTGIVSDRQGYKYENTPILPD 299
           +T L FKG+PNVK  G  +AGYT+ N+T  LYDG+ L +T+  V DR     Y+N PI PD
```

```
Sbjct: 240 LTALCFKGIPNVKFLGSDSAGYTSANQTVYLYDGSTLQITSAFVKDRTNNIYKNFPISPD 299

Query: 300 QVTSLPLQESQSWLKSRI 317
           T+     +  W+KS+I
Sbjct: 300 IQTNNAKSSAIEWIKSQI 317
```

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

A related GBS gene <SEQ ID 8641> and protein <SEQ ID 8642> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:     12.71
GvH: Signal Score (-7.5): -5.64
     Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value: -13.16 threshold:  0.0
   INTEGRAL    Likelihood =-13.16   Transmembrane    8 -  24 (   1 -  31)
   PERIPHERAL  Likelihood =  4.03     174
 modified ALOM score:   3.13

*** Reasoning Step: 3

----- 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>
```

The protein has homology with the following sequences in the databases:

```
34.7/62.5% over 311aa
                                          Lactococcus lactis
  GP|805128| nisin-resistance protein Insert characterized

ORF01108(343 - 1254 of 1560)
GP|805128|gb|AAB08491.1||U25181(7  -  318  of  318)  nisin-resistance  protein  (Lactococcus
lactis)
%Match = 19.4
%Identity = 34.6  %Similarity = 62.4
Matches = 106  Mismatches = 112  Conservative Sub.s = 85

231       261       291       321       351                 393       423
LKLSNL*EIGLKM*GYSKPFCHIIDLKRKGEQEMRRKIVLLFVVPMLIVLGILGV------VVHYYGSALNIYLLPPSSE
                                        :   |:||:: |      :::::|  :||||:||| :
                                       MKIGKRILLGLVAVCALFLGIIYFWGYKFNIYLVPPSPQ
                                                10        20        30

453       483       513       543       573       603       633       663
RYGRVILDRVEQRGLYSQGRQWQIIRQRSEKKLKTSKSYQESRNIVQEAVRYGGGKHSQILSKETVRRDTLDSRYPEYRR
:| || |  ::: ||::  ::|   :::: ::   :|:| |:   :|:|:: ||||| |  :| : : ::
KYVRVALKNMDELGLFTDSKEWVETKKKTIEETSNAKNYAETIPFLQKAIKVAGGKHSFIEHEEDISKRSITKYIKPKAE
          50        60        70        80        90        100       110

693       723       753       783       813       843       873       903
LNEDILLITIPSISKLDKRSISHYSGKLQNILMEKSYKGLILDLSNNTGGNMIPMIGGVASILPNDTLFHYTDKYGNKKT
:  : |::|||  :  | :: | |:  |:: : : :| |:|:||  | ||:: ||: |:: :||: ||| | ||  : |
IEGNTLILTIPEFTGNDSQA-SDYANFLESSFHKNNYNGVIVDLRGNRGGDLSPMVLGLSPLLPDGTLFTYVDKSSHSKP
        130       140       150       160       170       180       190

933       963       984       1014      1044      1074      1104      1134
ITMKNIPLEALKISRKTINTKHV---PIAIITNHKTASSAEMTFLSFKGLPNVKSFGQATAGYTTVNETFMLYDGARLAL
: ::|  : :   | |  : | :   |||:: :: | || |:| | |||:|||| :|  :||||: |:|  ||||: | :
VELQNGEINSGGSSTKVSDNKKIKKAPIAVLIDNNTGSSGELTALCFKGIPNVKFLGSDSAGYTSANQTVYLYDGSTLQI
         210       220       230       240       250       260       270

1164      1194      1224      1254      1284      1314      1344      1374
TTGIVSDRQGYKYENTPILPDQVTSLPLQESQSWLKSRINQN*GIINKGELYVIRNQSLRKSFSYTFFKRRDKGSTRRRF
```

```
|:  | ||    |:| || ||  |:     :  |:||:|
TSAFVKDRTNNIYKNFPISPDIQTNNAKSSAIEWIKSQIK
```

SEQ ID 2122 (GBS38) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 7; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 12; MW 62kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 689

A DNA sequence (GBSx0731) was identified in *S.agalactiae* <SEQ ID 2123> which encodes the amino acid sequence <SEQ ID 2124>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> 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>
```

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

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2125> which encodes the amino acid sequence <SEQ ID 2126>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1369(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 31/49 (63%), Positives = 43/49 (87%)

Query: 6  KKLTKSLGPIGKLISIIPDTTELIGKAIDNSRPIIEKELDRRHEKKTDL 54
          K++ K+LG +GKL+SI+PDTTE+IGK IDNSRPIIEK ++++HEK+  L
Sbjct: 3  KRIRKALGVVGKLMSIVPDTTEIIGKTIDNSRPIIEKRMEQKHEKEMQL 51
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 690

A DNA sequence (GBSx0732) was identified in *S.agalactiae* <SEQ ID 2127> which encodes the amino acid sequence <SEQ ID 2128>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3644(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 2126.

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

Example 691

A DNA sequence (GBSx0733) was identified in *S.agalactiae* <SEQ ID 2129> which encodes the amino acid sequence <SEQ ID 2130>. This protein is predicted to be 28 kd outer membrane protein precursor (yaeC). Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> 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>
```

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

```
>GP:CAB59827 GB:AJ012388 hypothetical protein [Lactococcus lactis]
 Identities = 123/290 (42%), Positives = 178/290 (60%), Gaps = 18/290 (6%)

Query: 1   MKIKKLLGLTTTVVISALILGAC------GQSKNEDAKVVRVGTMVKSKTEKARWDKIEE 54
           +K +++L +T  +++  +I+G        G      +K+V++G M   K E   W ++++
Sbjct: 3   VKNRRIL-ITIIILVFIIIVGGIFAFSHSGNKSKVSSKIVKIGLMPGGKQEDVIWKQVQK 61

Query: 55  LVKKK-GVKLKFTEFTDYTQPNKALESDEIDINAFQHYNYLNNWNKANKTNLVSVAETYF 113
             K + G+ LKF  FTD  +PNKAL + E+D+NAFQHY YL +WNKAN  N+VS+ +T
Sbjct: 62  NAKDQFGITLKFVNFTDGDEPNKALVNHEVDLNAFQHYAYLKSWNKANNGNIVSIGDTII 121

Query: 114 TSFRLYSGTKNGKGKYQTVSEIPNKATITIPNDAVNESRSLYLLQSAGLLKLKVSGDALA 173
           T   LYS       KY+ V EIP+K+TI IPND  NESR+LY+L++AGL+KL  S   LA
Sbjct: 122 TPIHLYST------KYKKVDEIPDKSTIAIPNDITNESRALYVLKNAGLIKLDTSRGVLA 175

Query: 174 TMSDVVSNPKSLDLKEVDAAQTARSLDSTDAAVINNDFVTEAGINPKSAIFIEPKSKNAK 233
           T+ D+  NPKSL +KE+DA+QT R+LDS  AAVIN +F   A  + K +I+ EP ++++
Sbjct: 176 TVKDIRENPKSLIIKEIDASQTPRALDSVAAAVINYNFAISAKNSDKESIYQEPLNEDSA 235

Query: 234 QWYNLLVAQKGWQDKSKAKAIKEVVKAYHTDAVKKVIEKT-SQGLDQPVW 282
           QW N + A    Q      K  KEVVKAY    +  +I+K    G + P W
Sbjct: 236 QWINFIAAN---QSDKNNKVYKEVVKAYEQKNIADIIKKEYPDGGELPAW 282
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2131> which encodes the amino acid sequence <SEQ ID 2132>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1766(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 145/264 (54%), Positives = 203/264 (75%), Gaps = 2/264 (0%)

Query: 20  LGACGQSKNEDAKVVRVGTMVKSKTEKARWDKIEELVKKKGVKLKFTEFTDYTQPNKALE 79
           L AC + K +D    + +G M K+++++ARWDK+EEL+KK  + LK+ EFTDY+QPNKA+
Sbjct: 1   LVACSE-KQDDKNTLTIGVMTKTESDQARWDKVEELLKKDNITLKYKEFTDYSQPNKAVA 59
```

```
Query: 80  SDEIDINAFQHYNYLNNWNKANKTNLVSVAETYFTSFRLYSGT-KNGKGKYQTVSEIPNK 138
           + E+DINAFQHYN+LNNWNK NK +LV++A+TY +    L+SGT ++GK KY++V+++PN
Sbjct: 60  NGEVDINAFQHYNFLNNWNKENKEHLVAIADTYISPINLFSGTSQDGKAKYKSVADLPNG 119

Query: 139 ATITIPNDAVNESRSLYLLQSAGLLKLKVSGDALATMSDVVSNPKSLDLKEVDAAQTARS 198
             I +PNDA NESR+LY+LQSAGL+KL VSGD LAT++++  N K LD+KE+DA+QTAR+
Sbjct: 120 TQIAVPNDATNESRALYVLQSAGLIKLNVSGDQLATIANISENKKKLDIKELDASQTARA 179

Query: 199 LDSTDAAVINNDFVTEAGINPKSAIFIEPKSKNAKQWYNLLVAQKGWQDKSKAKAIKEVV 258
           L S DAAV+NN +   A I+ K+++F E    N+KQW N++  QK W+   KA AIK+++
Sbjct: 180 LVSADAAVVNNSYAVPAKIDYKTSLFKEKADDNSKQWINIIAGQKDWEKSEKADAIKKLI 239

Query: 259 KAYHTDAVKKVIEKTSQGLDQPVW 282
           KAY TD VKKV+EKTS G+D  VW
Sbjct: 240 KAYQTDEVKKVVEKTSNGIDVSVW 263
```

SEQ ID 2130 (GBS96) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 7; MW 32kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 22 (lane 3; MW 57.2kDa).

The GBS96-GST fusion product was purified (Figure 195, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 290), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 692

A DNA sequence (GBSx0734) was identified in *S.agalactiae* <SEQ ID 2133> which encodes the amino acid sequence <SEQ ID 2134>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5103(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9807> which encodes amino acid sequence <SEQ ID 9808> was also identified.

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

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 693

A DNA sequence (GBSx0735) was identified in *S.agalactiae* <SEQ ID 2135> which encodes the amino acid sequence <SEQ ID 2136>. This protein is predicted to be glucose-inhibited division protein (gid). Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0656(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 GENPEPT database:

```
>GP:CAB13486 GB:Z99112 glucose-inhibited division protein [Bacillus subtilis]
 Identities = 289/439 (65%), Positives = 352/439 (79%), Gaps = 10/439 (2%)

Query: 1   MSQSYINVIGAGLAGSEAAYQIAKRGIPVKLYEMRGVKSTPQHKTDNFAELVCSNSFRGD 60
            M+Q  +NVIGAGLAGSEAA+Q+AKRGI VKLYEMR VK TP H TD FAELVCSNS R +
Sbjct: 1   MNQQTVNVIGAGLAGSEAAWQLAKRGIQVKLYEMRPVKQTPAHHTDKFAELVCSNSLRSN 60

Query: 61  SLTNAVGLLKEEMRRLDSIIMRNGEAHRVPAGGAMAVDREGYSEAVTEEIHKHPLIEVIR 120
            +L NAVG+LKEEMR LDS I+   +   VPAGGA+AVDR  ++ +VT  +  HP + VI
Sbjct: 61  TLANAVGVLKEEMRALDSAIIAAADECSVPAGGALAVDRHEFAASVTNRVKNHPNVTVIN 120

Query: 121 DEITDIPGDAITVIATGPLTSDSLAAKIHELNGGDGFYFYDAAAPIVDKNTIDINKVYLK 180
            +E+T+IP +  T+IATGPLTS+SL+A++ EL G D  YFYDAAAPIV+K+++D++KVYLK
Sbjct: 121 EEVTEIP-EGPTIIATGPLTSESLSAQLKELTGEDYLYFYDAAAPIVEKDSLDMDKVYLK 179

Query: 181 SRYDKGEAAYLNCPMTKEEFMAFHEALTTAEEAPLNSFEKEKYFEGCMPIEVMAKRGIKT 240
            SRYDKGEAAYLNCPMT+EEF  FHEALT+AE  PL  FEKE +FEGCMPIEVMAKRG KT
Sbjct: 180 SRYDKGEAAYLNCPMTEEEFDRFHEALTSAETVPLKEFEKEIFFEGCMPIEVMAKRGKKT 239

Query: 241 MLYGPMKPVGLEYPEDYKGPRDGEFKTPYAVVQLRQDNAAGSLYNIVGFQTHLKWGEQKR 300
            ML+GPMKPVGLE+P   K          PYAVVQLRQD+AAG+LYNIVGFQTHLKWG+QK
Sbjct: 240 MLFGPMKPVGLEHPVTGK--------RPYAVVQLRQDDAAGTLYNIVGFQTHLKWGDQKE 291

Query: 301 VFQMIPGLENAEFVRYGVMHRNSYMDSPNLLNQTFATRKNPNLFFAGQMTGVEGYVESAA 360
            V ++IPGLEN E VRYGVMHRN++++SP+LL  T+  +   +LFFAGQMTGVEGYVESAA
Sbjct: 292 VLKLIPGLENVEIVRYGVMHRNTFINSPSLLKPTYQFKNRSDLFFAGQMTGVEGYVESAA 351

Query: 361 SGLVAGINAVRRFNGESEVVFPQTTAIGALPHYITHTDSKHFQPMNVNFGIIKELEGPRI 420
            SGLVAGINA +   GE  V+FPQ TAIG++ HYIT T+ K+FQPMN NFG++KEL   +I
Sbjct: 352 SGLVAGINAAKLVLGEELVIFPQETAIGSMAHYITTTNQKNFQPMNANFGLLKELP-VKI 410

Query: 421 RDKKERYEAIATRALKDLE 439
            ++KER E  A RA++ ++
Sbjct: 411 KNKKERNEQYANRAIETIQ 429
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2137> which encodes the amino acid sequence <SEQ ID 2138>. Analysis of this protein sequence reveals the following:

```
      Possible site: 30

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.44   Transmembrane   12 -  28 (   9 -  32)

----- 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>

RGD motif: 111-113
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB13486 GB:Z99112 glucose-inhibited division protein [Bacillus subtilis]
 Identities = 292/435 (67%), Positives = 350/435 (80%), Gaps = 10/435 (2%)

Query: 59  INVIGAGLAGSEAAYQIAKRGIPVKLYEMRGVKATPQHKTTNFAELVCSNSFRGDSLTNA 118
           +NVIGAGLAGSEAA+Q+AKRGI VKLYEMR VK TP H T  FAELVCSNS R ++L NA
Sbjct: 6   VNVIGAGLAGSEAAWQLAKRGIQVKLYEMRPVKQTPAHHTDKFAELVCSNSLRSNTLANA 65

Query: 119 VGLLKEEMRRLDSIIMRNGEANRVPAGGAMAVDREGYAESVTAELENHPLIEVIRGEITE 178
```

```
            VG+LKEEMR LDS I+   +   VPAGGA+AVDR  +A SVT  ++NHP + VI  E+TE
Sbjct: 66   VGVLKEEMRALDSAIIAAADECSVPAGGALAVDRHEFAASVTNRVKNHPNVTVINEEVTE 125

Query: 179 IPDDAITVIATGPLTSDALAEKIHALNGGDGFYFYDAAAPIIDKSTIDMSKVYLKSRYDK 238
            IP+   T+IATGPLTS++L+ ++  L G D  YFYDAAAPI++K ++DM KVYLKSRYDK
Sbjct: 126  IPEGP-TIIATGPLTSESLSAQLKELTGEDYLYFYDAAAPIVEKDSLDMDKVYLKSRYDK 184

Query: 239  GEAAYLNCPMTKEEFMAPHEALTTAEEAPLNAFEKEKYFEGCMPIEVMAKRGIKTMLYGP 298
            GEAAYLNCPMT+EEF  FHEALT+AE  PL  FEKE +FEGCMPIEVMAKRG KTML+GP
Sbjct: 185  GEAAYLNCPMTEEEFDRFHEALTSAETVPLKEFEKEIFFEGCMPIEVMAKRGKKTMLFGP 244

Query: 299  MKPVGLEYPDDYTGPRDGEFKTPYAVVQLRQDNAAGSLYNIVGFQTHLKWGEQKRVFQMI 358
            MKPVGLE+P   TG R       PYAVVQLRQD+AAG+LYNIVGFQTHLKWG+QK V ++I
Sbjct: 245  MKPVGLEHP--VTGKR------PYAVVQLRQDDAAGTLYNIVGFQTHLKWGDQKEVLKLI 296

Query: 359  PGLENAEFVRYGVMHRNSYMDSPNLLTETFQSRSNPNLFFAGQMTGVEGYVESAASGLVA 418
            PGLEN E VRYGVMHRN++++SP+LL  T+Q ++  +LFFAGQMTGVEGYVESAASGLVA
Sbjct: 297  PGLENVEIVRYGVMHRNTFINSPSLLKPTYQFKNRSDLFFAGQMTGVEGYVESAASGLVA 356

Query: 419  GINAARLFKREEALIFPQTTAIGSLPHYVTHADSKHFQPMNVNFGIIKELEGPRIRDKKE 478
            GINAA+L   EE +IFPQ TAIGS+ HY+T  + K+FQPMN NFG++KEL   +I++KKE
Sbjct: 357  GINAAKLVLGEELVIFPQETAIGSMAHYITTTNQKNFQPMNANFGLLKELP-VKIKNKKE 415

Query: 479  RYEAIASRALADLDT 493
            R E  A+RA+  + T
Sbjct: 416  RNEQYANRAIETIQT 430
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 395/439 (89%), Positives = 417/439 (94%)

Query: 4    SYINVIGAGLAGSEAAYQIAKRGIPVKLYEMRGVKSTPQHKTDNFAELVCSNSFRGDSLT 63
            +YINVIGAGLAGSEAAYQIAKRGIPVKLYEMRGVK+TPQHKT NFAELVCSNSFRGDSLT
Sbjct: 57   TYINVIGAGLAGSEAAYQIAKRGIPVKLYEMRGVKATPQHKTTNFAELVCSNSFRGDSLT 116

Query: 64   NAVGLLKEEMRRLDSIIMRNGEAHRVPAGGAMAVDREGYSEAVTEEIHKHPLIEVIRDEI 123
            NAVGLLKEEMRRLDSIIMRNGEA+RVPAGGAMAVDREGY+E+VT E+  HPLIEVIR EI
Sbjct: 117  NAVGLLKEEMRRLDSIIMRNGEANRVPAGGAMAVDREGYAESVTAELENHPLIEVIRGEI 176

Query: 124  TDIPGDAITVIATGPLTSDSLAAKIHELNGGDGFYFYDAAAPIVDKNTIDINKVYLKSRY 183
            T+IP DAITVIATGPLTSD+LA KIH LNGGDGFYFYDAAAPI+DK+TID++KVYLKSRY
Sbjct: 177  TEIPDDAITVIATGPLTSDALAEKIHALNGGDGFYFYDAAAPIIDKSTIDMSKVYLKSRY 236

Query: 184  DKGEAAYLNCPMTKEEFMAFHEALTTAEEAPLNSFEKEKYFEGCMPIEVMAKRGIKTMLY 243
            DKGEAAYLNCPMTKEEFMAFHEALTTAEEAPLN+FEKEKYFEGCMPIEVMAKRGIKTMLY
Sbjct: 237  DKGEAAYLNCPMTKEEFMAFHEALTTAEEAPLNAFEKEKYFEGCMPIEVMAKRGIKTMLY 296

Query: 244  GPMKPVGLEYPEDYKGPRDGEFKTPYAVVQLRQDNAAGSLYNIVGFQTHLKWGEQKRVFQ 303
            GPMKPVGLEYP+DY GPRDGEFKTPYAVVQLRQDNAAGSLYNIVGFQTHLKWGEQKRVFQ
Sbjct: 297  GPMKPVGLEYPDDYTGPRDGEFKTPYAVVQLRQDNAAGSLYNIVGFQTHLKWGEQKRVFQ 356

Query: 304  MIPGLENAEFVRYGVMHRNSYMDSPNLLNQTFATRKNPNLFFAGQMTGVEGYVESAASGL 363
            MIPGLENAEFVRYGVMHRNSYMDSPNLL +TF +R NPNLFFAGQMTGVEGYVESAASGL
Sbjct: 357  MIPGLENAEFVRYGVMHRNSYMDSPNLLTETFQSRSNPNLFFAGQMTGVEGYVESAASGL 416

Query: 364  VAGINAVRRFNGESEVVFPQTTAIGALPHYITHTDSKHFQPMNVNFGIIKELEGPRIRDK 423
            VAGINA R F  E  ++FPQTTAIG+LPHY+TH DSKHFQPMNVNFGIIKELEGPRIRDK
Sbjct: 417  VAGINAARLFKREEALIFPQTTAIGSLPHYVTHADSKHFQPMNVNFGIIKELEGPRIRDK 476

Query: 424  KERYEAIATRALKDLEKFL 442
            KERYEAIA+RAL DL+  L
Sbjct: 477  KERYEAIASRALADLDTCL 495
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 694

A DNA sequence (GBSx0736) was identified in *S.agalactiae* <SEQ ID 2139> which encodes the amino acid sequence <SEQ ID 2140>. This protein is predicted to be transcriptional regulator (GntRfamily). Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5103(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 GENPEPT database:

```
>GP:BAB04138 GB:AP001508 transcriptional regulator (GntR family)
          [Bacillus halodurans]
 Identities = 83/229 (36%), Positives = 133/229 (57%), Gaps = 1/229 (0%)

Query: 2   LPAYIKIHDAIKKEIDKGTWKIGQRLPSERDLADDYSVSRMTLRQSITLLVEEGILERRV 61
           LP Y +I + IK++I+ G  K G  L SER+ A+ Y VSRMT+RQ+I  LV +G + ++
Sbjct: 8   LPIYYQIEEQIKQQIESGVLKPGDMLKSEREYAEYYDVSRMTVRQAINNLVNQGYIYKKK 67

Query: 62  GSGTYVASHRVQEKMRGTTSFTEIVNSQGRKPSSKLISFQRKLANETEIQKLNLSQSDYV 121
           GSGTYV   ++++ + G TSFTE +  +G +PSS+L+ F+   A     ++LNL ++  V
Sbjct: 68  GSGTYVQEKKIEQALNGLTSFTEDMRKRGMEPSSRLLKFELIPATAKIAKELNLKENTPV 127

Query: 122 VRMERVRYADKVPLVYEVASIPENLIKGFEQSEVTEHFFKTLTEN-GYEIGKSQQTIYAR 180
             ++R+RY D VP+  E   +P NL+KG  +  + +  ++ + E     I  + Q I A
Sbjct: 128 TEIKRIRYGDGVPIAIERNLLPANLVKGLNEEIINQSLYQYIEEELNLRIADALQVIEAS 187

Query: 181 NASERVASHLEVNAGHAILALTQVSYFTDGKPFEYVHGQYVGDRFEFYL 229
            AS+  A  LE+  G  IL + + ++  DG   E V   Y  DR++F +
Sbjct: 188 TASKTEADLLEIQKGSPILLIERKTFLADGTVLELVKSAYRADRYKFMI 236
```

There is also homology to SEQ ID 1256.

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

Example 695

A DNA sequence (GBSx0737) was identified in *S.agalactiae* <SEQ ID 2141> which encodes the amino acid sequence <SEQ ID 2142>. This protein is predicted to be GMP synthase (guaA). Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.96   Transmembrane  228 - 244 ( 228 - 245)

----- Final Results -----
             bacterial membrane --- Certainty=0.1383(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 GENPEPT database:

```
>GP:AAD15805 GB:AF058326 GMP synthase [Lactococcus lactis]
 Identities = 416/511 (81%), Positives = 467/511 (90%), Gaps = 3/511 (0%)

Query: 10  IQKIIVLDYGSQYNQLIARRIREFGVFSELKSHKITADEIRDINPIGIVLSGGPNSVYAD 69
           ++KIIVLDYGSQYNQLIARRIRE GVFSEL SHK+TA EIR+INPIGI+LSGGPNSVY +
Sbjct: 6   LEKIIVLDYGSQYNQLIARRIREIGVFSELMSHKVTAKEIREINPIGIILSGGPNSVYDE 65
```

```
Query: 70  GAFGIDEEIFELGIPILGICYGMQLITHKLGGKVLPAGEAGHREYGQSALRLRSESALFA 129
           G+F ID EIFELG+P+LGICYGMQL+++KLGG V  AGE   REYG + L+L  +SALFA
Sbjct: 66  GSFDIDPEIFELGLPVLGICYGMQLMSYKLGGMVEAAGE---REYGVAPLQLTEKSALFA 122

Query: 130 GTPQEQLVLMSHGDAVTEIPEGFHLVGDSVDCPFAAMENTEKQFYGIQFHPEVRHSVYGN 189
           GTP+ Q VLMSHGD VT IPEGFH+VG S + PFAA+ENTE+  YGIQFHPEVRHSV+G
Sbjct: 123 GTPEVQDVLMSHGDRVTAIPEGFHVVGTSPNSPFAAVENTERNLYGIQFHPEVRHSVHGT 182

Query: 190 DILKNFAVNICGARGDWSMDNFIDMEIAKIRETVGDRKVLLGLSGGVDSSVVGVLLQRAI 249
           ++L+NFA+NICGA+G+WSM+NFIDM+I  IRE VGD+KVLLGLSGGVDSSVVGVLLQRAI
Sbjct: 183 EMLRNFALNICGAKGNWSMENFIDMQIKDIREKVGDKKVLLGLSGGVDSSVVGVLLQRAI 242

Query: 250 GDQLTCIFVDHGLLRKNEGDQVMDMLGGKFGLNIIRVDASKRFLDLLSGVEDPERKRKII 309
           GDQLT IFVDHG LRK E DQVM+ LGGKFGLNII+VDA KRF+D L G+ DPE +RKII
Sbjct: 243 GDQLTSIFVDHGFLRKGEADQVMETLGGKFGLNIIKVDAQKRFMDKLVGLSDPETQRKII 302

Query: 310 GNEFVYVFDDEASKLKGVDFLAQGTLYTDIIESGTETAQTIKSHHNVGGLPEDMQFELIE 369
           GNEFVYVFDDEA+KL+GVDFLAQGTLYTD+IESGT+TAQTIKSHHNVGGLPEDMQF+LIE
Sbjct: 303 GNEFVYVFDDEANKLEGVDFLAQGTLYTDVIESGTDTAQTIKSHHNVGGLPEDMQFQLIE 362

Query: 370 PLNTLFKDEVRALGTALGMPDEVVWRQPFPGPGLAIRVMGEITEEKLETVRESDAILREE 429
           PLNTLFKDEVRALGT LGMPDE+VWRQPFPGPGLAIRV+G++TEEKLETVRESDAILREE
Sbjct: 363 PLNTLFKDEVRALGTQLGMPDEIVWRQPFPGPGLAIRVLGDLTEEKLETVRESDAILREE 422

Query: 430 IAKAGLDRDVWQYFTVNTGVRSVGVMGDGRTYDYTIAIRAITSIDGMTADFAQLPWDVLK 489
           IA +GL+RDVWQYFTVNT V+SVGVMGD RTYDYT+AIRAITSIDGMTADFAQLPWD+L+
Sbjct: 423 IAASGLERDVWQYFTVNTDVKSVGVMGDQRTYDYTLAIRAITSIDGMTADFAQLPWDLLQ 482

Query: 490 KISTRIVNEVDHVNRIVYDITSKPPATVEWE 520
           KIS RIVNEVDHVNRIVYDITSKPPATVEW+
Sbjct: 483 KISKRIVNEVDHVNRIVYDITSKPPATVEWQ 513
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2143> which encodes the amino acid sequence <SEQ ID 2144>. Analysis of this protein sequence reveals the following:

```
     Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.96   Transmembrane  228 - 244 ( 228 - 245)

----- Final Results -----
               bacterial membrane --- Certainty=0.1383(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 203-205
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD15805 GB:AF058326 GMP synthase [Lactococcus lactis]
 Identities = 411/511 (80%), Positives = 464/511 (90%), Gaps = 3/511 (0%)

Query: 10  VQKIIVLDYGSQYNQLIARRIREFGVFSELKSHKITAQELREINPIGIVLSGGPNSVYAD 69
           ++KIIVLDYGSQYNQLIARRIRE GVFSEL SHK+TA+E+REINPIGI+LSGGPNSVY +
Sbjct: 6   LEKIIVLDYGSQYNQLIARRIREIGVFSELMSHKVTAKEIREINPIGIILSGGPNSVYDE 65

Query: 70  NAFGIDPEIFELGIPILGICYGMQLITHKLGGKVVPAGQAGNREYGQSTLHLRETSKLFS 129
            +F IDPEIFELG+P+LGICYGMQL+++KLGG V  AG+   REYG + L L E S LF+
Sbjct: 66  GSFDIDPEIFELGLPVLGICYGMQLMSYKLGGMVEAAGE---REYGVAPLQLTEKSALFA 122

Query: 130 GTPQEQLVLMSHGDAVTEIPEGFHLVGDSNDCPYAAIENTEKNLYGIQFHPEVRHSVYGN 189
           GTP+ Q VLMSHGD VT IPEGFH+VG S + P+AA+ENTE+NLYGIQFHPEVRHSV+G
Sbjct: 123 GTPEVQDVLMSHGDRVTAIPEGFHVVGTSPNSPFAAVENTERNLYGIQFHPEVRHSVHGT 182

Query: 190 DILKNFAISICGARGDWSMDNFIDMEIAKIRETVGDRKVLLGLSGGVDSSVVGVLLQKAI 249
           ++L+NFA++ICGA+G+WSM+NFIDM+I  IRE VGD+KVLLGLSGGVDSSVVGVLLQ+AI
Sbjct: 183 EMLRNFALNICGAKGNWSMENFIDMQIKDIREKVGDKKVLLGLSGGVDSSVVGVLLQRAI 242
```

```
Query: 250 GDQLTCIFVDHGLLRKDEGDQVMGMLGGKFGLNIIRVDASKRFLDLLADVEDPEKKRKII 309
           GDQLT IFVDHG LRK E DQVM  LGGKFGLNII+VDA KRF+D L  + DPE +RKII
Sbjct: 243 GDQLTSIFVDHGFLRKGEADQVMETLGGKFGLNIIKVDAQKRFMDKLVGLSDPETQRKII 302

Query: 310 GNEFVYVFDDEASKLKGVDFLAQGTLYTDIIESGTETAQTIKSHHNVGGLPEDMQFELIE 369
           GNEFVYVFDDEA+KL+GVDFLAQGTLYTD+IESGT+TAQTIKSHHNVGGLPEDMQF+LIE
Sbjct: 303 GNEFVYVFDDEANKLEGVDFLAQGTLYTDVIESGTDTAQTIKSHHNVGGLPEDMQFQLIE 362

Query: 370 PLNTLFKDEVRALGIALGMPEEIVWRQPFPGPGLAIRVMGAITEEKLETVRESDAILREE 429
           PLNTLFKDEVRALG  LGMP+EIVWRQPFPGPGLAIRV+G +TEEKLETVRESDAILREE
Sbjct: 363 PLNTLFKDEVRALGTQLGMPDEIVWRQPFPGPGLAIRVLGDLTEEKLETVRESDAILREE 422

Query: 430 IAKAGLDRDVWQYFTVNTGVRSVGVMGDGRTYDYTIAIRAITSIDGMTADFAQLPWDVLK 489
           IA +GL+RDVWQYFTVNT V+SVGVMGD RTYDYT+AIRAITSIDGMTADFAQLPWD+L+
Sbjct: 423 IAASGLERDVWQYFTVNTDVKSVGVMGDQRTYDYTLAIRAITSIDGMTADFAQLPWDLLQ 482

Query: 490 KISTRIVNEVDHVNRIVYDITSKPPATVEWE 520
           KIS RIVNEVDHVNRIVYDITSKPPATVEW+
Sbjct: 483 KISKRIVNEVDHVNRIVYDITSKPPATVEWQ 513
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 487/520 (93%), Positives = 505/520 (96%)

Query: 1   MTDISILNDIQKIIVLDYGSQYNQLIARRIREFGVFSELKSHKITADEIRDINPIGIVLS 60
           MT+ISILND+QKIIVLDYGSQYNQLIARRIREFGVFSELKSHKITA E+R+INPIGIVLS
Sbjct: 1   MTEISILNDVQKIIVLDYGSQYNQLIARRIREFGVFSELKSHKITAQELREINPIGIVLS 60

Query: 61  GGPNSVYADGAFGIDEEIFELGIPILGICYGMQLITHKLGGKVLPAGEAGHREYGQSALR 120
           GGPNSVYAD AFGID EIFELGIPILGICYGMQLITHKLGGKV+PAG+AG+REYGQS L
Sbjct: 61  GGPNSVYADNAFGIDPEIFELGIPILGICYGMQLITHKLGGKVVPAGQAGNREYGQSTLH 120

Query: 121 LRSESALFAGTPQEQLVLMSHGDAVTEIPEGFHLVGDSVDCPFAAMENTEKQFYGIQFHP 180
           LR  S LF+GTPQEQLVLMSHGDAVTEIPEGFHLVGDS DCP+AA+ENTEK  YGIQFHP
Sbjct: 121 LRETSKLFSGTPQEQLVLMSHGDAVTEIPEGFHLVGDSNDCPYAAIENTEKNLYGIQFHP 180

Query: 181 EVRHSVYGNDILKNFAVNICGARGDWSMDNFIDMEIAKIRETVGDRKVLLGLSGGVDSSV 240
           EVRHSVYGNDILKNFA++ICGARGDWSMDNFIDMEIAKIRETVGDRKVLLGLSGGVDSSV
Sbjct: 181 EVRHSVYGNDILKNFAISICGARGDWSMDNFIDMEIAKIRETVGDRKVLLGLSGGVDSSV 240

Query: 241 VGVLLQRAIGDQLTCIFVDHGLLRKNEGDQVMDMLGGKFGLNIIRVDASKRFLDLLSGVE 300
           VGVLLQ+AIGDQLTCIFVDHGLLRK+EGDQVM MLGGKFGLNIIRVDASKRFLDLL+ VE
Sbjct: 241 VGVLLQKAIGDQLTCIFVDHGLLRKDEGDQVMGMLGGKFGLNIIRVDASKRFLDLLADVE 300

Query: 301 DPERKRKIIGNEFVYVFDDEASKLKGVDFLAQGTLYTDIIESGTETAQTIKSHHNVGGLP 360
           DPE+KRKIIGNEFVYVFDDEASKLKGVDFLAQGTLYTDIIESGTETAQTIKSHHNVGGLP
Sbjct: 301 DPEKKRKIIGNEFVYVFDDEASKLKGVDFLAQGTLYTDIIESGTETAQTIKSHHNVGGLP 360

Query: 361 EDMQFELIEPLNTLFKDEVRALGTALGMPDEVVWRQPFPGPGLAIRVMGEITEEKLETVR 420
           EDMQFELIEPLNTLFKDEVRALG ALGMP+E+VWRQPFPGPGLAIRVMG ITEEKLETVR
Sbjct: 361 EDMQFELIEPLNTLFKDEVRALGIALGMPEEIVWRQPFPGPGLAIRVMGAITEEKLETVR 420

Query: 421 ESDAILREEIAKAGLDRDVWQYFTVNTGVRSVGVMGDGRTYDYTIAIRAITSIDGMTADF 480
           ESDAILREEIAKAGLDRDVWQYFTVNTGVRSVGVMGDGRTYDYTIAIRAITSIDGMTADF
Sbjct: 421 ESDAILREEIAKAGLDRDVWQYFTVNTGVRSVGVMGDGRTYDYTIAIRAITSIDGMTADF 480

Query: 481 AQLPWDVLKKISTRIVNEVDHVNRIVYDITSKPPATVEWE 520
           AQLPWDVLKKISTRIVNEVDHVNRIVYDITSKPPATVEWE
Sbjct: 481 AQLPWDVLKKISTRIVNEVDHVNRIVYDITSKPPATVEWE 520
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 696

A DNA sequence (GBSx0740) was identified in *S.agalactiae* <SEQ ID 2145> which encodes the amino acid sequence <SEQ ID 2146>. This protein is predicted to be branched chain amino acid ABC transporter, periplasmic amino acid-bind. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0957(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9409> which encodes amino acid sequence <SEQ ID 9410> was also identified.

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

```
>GP:AAD36211 GB:AE001771 branched chain amino acid ABC transporter,
          periplasmic amino acid-binding protein [Thermotoga maritima]
 Identities = 31/92 (33%), Positives = 51/92 (54%), Gaps = 4/92 (4%)

Query: 26  AKAFHDHYVKAYGEEPSMFSALSYDAVYMAAKSAKGAKTSID---IKKALAKLKDFKGVT 82
           AK F + Y + YG+EP+  +AL YDA YM    A     S D   I + + K ++F G +
Sbjct: 275 AKKFVEVYKEKYGKEPAALNALGYDA-YMVLLDAIERAGSFDREKIAEEIRKTRNFNGAS 333

Query: 83  GKMSIDKNHNVVKSAYVVKLEDGKTSSVNIIS 114
           G ++ID+N + +KS  V  +++G      +I+
Sbjct: 334 GIINIDENGDAIKSVVVNIVKNGSVDFEAVIN 365
```

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

SEQ ID 9410 (GBS660) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 135 (lane 8 & 9; MW 71.5kDa) + lane 10; MW 27kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 141 (lane 2; MW 46.5kDa) and in Figure 181 (lane 3; MW 46kDa).

GBS660-His was purified as shown in Figure 233, lane 5-6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 697

A DNA sequence (GBSx0741) was identified in *S.agalactiae* <SEQ ID 2147> which encodes the amino acid sequence <SEQ ID 2148>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood =-10.61   Transmembrane  140 - 156 ( 129 - 158)
    INTEGRAL    Likelihood = -9.55   Transmembrane   60 -  76 (  53 -  80)
    INTEGRAL    Likelihood = -7.59   Transmembrane  264 - 280 ( 257 - 285)
    INTEGRAL    Likelihood = -5.79   Transmembrane  232 - 248 ( 219 - 251)
    INTEGRAL    Likelihood = -2.23   Transmembrane  190 - 206 ( 190 - 207)
    INTEGRAL    Likelihood = -1.75   Transmembrane   90 - 106 (  90 - 110)

----- Final Results -----
            bacterial membrane --- Certainty=0.5246(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10059> which encodes amino acid sequence <SEQ ID 10060> was also identified.

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

```
>GP:AAD36212 GB:AE001771 branched chain amino acid ABC transporter,
          permease protein [Thermotoga maritima]
 Identities = 140/295 (47%), Positives = 200/295 (67%), Gaps = 7/295 (2%)

Query: 2   LQQLVNGLILGSIYALLALGYTMVYGIIKLINFAHGDIYMMGAFMGYYLINHLHLNFFLA 61
           LQ L NG++LG +YAL+A+GYTMVYGI++LINFAHGD+ MMG +  +Y    L LN   +
Sbjct: 5   LQNLFNGIMLGGLYALIAIGYTMVYGILRLINFAHGDVMMMGVYFAFYAATLLSLNPLFS 64

Query: 62  LLIAMLGSAFLGVVIEYLAYRPLRKSTRIAALITAIGVSFLLEYGMVYLVGADTRAFPQA 121
            ++A+LG+A LG +I+ +AY+PLR + RI+ALITAIGVSF LE   V + GA  ++F +
Sbjct: 65  AIVAILGAALLGFLIDRVAYKPLRNAPRISALITAIGVSFFLESLAVVVFGAIPKSFLKV 124

Query: 122 IHTVKYNLGPITITNVQL-----IILGIALLLMLTLQFIVQKTKMGKAMRALSVDSDAAQ 176
                    +T+   ++     +++ I  ++++ L FIV +TK+G AMRA+S+D
Sbjct: 125 FKDRTILNKVLTVAGARIPLLTFLVIFITAVILIVLFFIVYRTKIGMAMRAISMDIPTTA 184

Query: 177 LMGINVNRTISFTFALGSALAGAGGVLIGLYYNSVQPLMGVTPGLKAFVAAVLGGIGIIP 236
           LMG+NV+  I FTFALGSALA A G++  + + +V P MG  PGLKAF+AAV GGIG IP
Sbjct: 185 LMGVNVDAVIGFTFALGSALAAAAGIMWAMRFPNVHPYMGFMPGLKAFIAAVFGGIGSIP 244

Query: 237 GAAIGGFVIGILETLATAL--GVSDFRDGIVYAILILIFLIRPAGILGKNIKEKV 289
           GA +GG ++G++E    A    V  +RD   + ILI+I L++P+G+LGK I EKV
Sbjct: 245 GAVLGGVLLGLIEIFLAAYFPAVMGYRDAFAFIILIIILLVKPSGLLGKKIVEKV 299
```

There is also homology to SEQ ID 2150. A related sequence was also identified in GAS <SEQ ID 9171> which encodes the amino acid sequence <SEQ ID 9172>. Analysis of this protein sequence reveals the following:

```
     Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL    Likelihood =-12.74   Transmembrane  196 - 212 ( 191 - 219)
     INTEGRAL    Likelihood =-12.42   Transmembrane   12 -  28 (   5 -  36)
     INTEGRAL    Likelihood = -7.22   Transmembrane  106 - 122 ( 102 - 126)
     INTEGRAL    Likelihood = -4.78   Transmembrane  242 - 258 ( 240 - 260)
     INTEGRAL    Likelihood = -2.50   Transmembrane   61 -  77 (  60 -  77)
     INTEGRAL    Likelihood = -2.34   Transmembrane  293 - 309 ( 291 - 309)
     INTEGRAL    Likelihood = -1.44   Transmembrane  139 - 155 ( 138 - 156)
     INTEGRAL    Likelihood = -1.33   Transmembrane  317 - 333 ( 317 - 333)

----- Final Results -----
              bacterial membrane --- Certainty=0.609(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 35/147 (23%), Positives = 71/147 (47%), Gaps = 6/147 (4%)

Query: 134 ITNVQLIILGI--ALLLMLTLQFIVQKTKMGKAMRALSVDSDAAQLMGINVNRTISFTFA 191
           +TN   I +GI  A++ +  + F++ KT +G  +R++ ++  A++  G++  RTI  +
Sbjct: 197 LTNNSRINIGIFFAIIAIALIWFLLNKTTLGFEIRSVGLNPHASEYAGMSSKRTIILSMI 256

Query: 192 LGSALAGAGGVL--IGLYYNSVQPLMGVTPGLKAFVAAVLGGIGIIPGAAIGGFVIGILE 249
           +  ALAG GGV+  +G + N       +  G       ++L     + G     F+ G+L
Sbjct: 257 ISGALAGLGGVVEGLGTFENVFVQGSSLAVGFDGMAVSLLAANSPL-GIFFSSFLFGVLN 315

Query: 250 TLATALGVSDFRDGIVYAILI-LIFLI 275
            A  + ++     +V  +   +IF +
Sbjct: 316 IGAPGMNIAGIPPELVKVVTASIIFFV 342
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 698

A DNA sequence (GBSx0742) was identified in *S.agalactiae* <SEQ ID 2151> which encodes the amino acid sequence <SEQ ID 2152>. This protein is predicted to be branched chain amino acid ABC transporter, permease protein (livM). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.76   Transmembrane   90 - 106 (  84 - 113)
   INTEGRAL    Likelihood = -8.23   Transmembrane   12 -  28 (   5 -  33)
   INTEGRAL    Likelihood = -8.17   Transmembrane  205 - 221 ( 200 - 224)
   INTEGRAL    Likelihood = -7.86   Transmembrane  276 - 292 ( 273 - 300)
   INTEGRAL    Likelihood = -6.32   Transmembrane  159 - 175 ( 154 - 176)
   INTEGRAL    Likelihood = -6.05   Transmembrane  236 - 252 ( 232 - 264)
   INTEGRAL    Likelihood = -5.95   Transmembrane   42 -  58 (  38 -  60)
   INTEGRAL    Likelihood = -5.84   Transmembrane  120 - 136 ( 119 - 138)
   INTEGRAL    Likelihood = -4.35   Transmembrane  255 - 271 ( 253 - 274)
   INTEGRAL    Likelihood = -1.59   Transmembrane   66 -  82 (  66 -  85)

----- Final Results -----
              bacterial membrane --- Certainty=0.4503(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 GENPEPT database:

```
>GP:AAD36213 GB:AE001771 branched chain amino acid ABC transporter,
          permease protein [Thermotoga maritima]
 Identities = 119/332 (35%), Positives = 191/332 (56%), Gaps = 33/332 (9%)

Query: 12  LAIVVLDYLLISVLISMGIFNLYHIQIIETIGINVILAVGLNLIVGCSGQFSLGHAGFMA 71
           L +V L ++ + + ++    + Y ++++  I I  I+AV LNLI G +G FSLGHAGF+
Sbjct: 16  LTVVFLIFMALLLYLADRYMDSYKLRVVRLIAIYGIMAVSLNLINGITGIFSLGHAGFIL 75

Query: 72  IGAYAVAIIGVKMP----------------TYVGFLIAILVGTLVAGGIALGVGIPTLR 114
           IGAY  +++ +                    +  F  A + G ++A   A  +G P LR
Sbjct: 76  IGAYTASLLTLSPEQKAMSFIIEPIVPWLANAHTDFFTATVAGGVLAAVFAFLIGWPVLR 135

Query: 115 LKGDYLAIATLGVAEIIRILLVNGGDITNGAAGIMGIPPFTTWSLVYGVAVVSLILAMNF 174
           L GDYLAIA+LG AE+IRI+ +N   ITNG  G+ GIP ++     YG   V+++   +
Sbjct: 136 LSGDYLAIASLGFAEVIRIIALNAISITNGPLGLKGIPEYSNIWWCYGWLFVTVLFMASL 195

Query: 175 LRSPLGRNTIAIREDEIAAESMGVDTTKVKVIVFVFGAILASIAGSLQAGYVGTVMPKDF 234
           + S  GR   AIRED IAAE+MG++  K +++ FV GA  A ++GSL A ++ T+ P+
Sbjct: 196 VNSSYGRALKAIREDRIAAEAMGINVFKHQLLSFVIGAFFAGVSGSLYAHWLTTIDPRTT 255

Query: 235 SF--MMSVNVLIIVVLGGLGSMTGTVLAAILLGLLNMLLQD----------YASVR 278
           +   M++  VLI++VLGGLGS++G+++ A L  +L   L+D                +R
Sbjct: 256 TLGPMLTFYVLIMIVLGGLGSISGSLIGAALFAILFEWLRDLEEPFTFFGIHVPGIKGMR 315

Query: 279 MIIYALALILIMIFRPSGLLGTKELTLSHLFR 310
           +++ +   IL+MIF   G++G +ELT ++L+R
Sbjct: 316 ILVISAIFILVMIFWQRGIMGREELTWNNLYR 347
```

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

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

Example 699

A DNA sequence (GBSx0743) was identified in *S.agalactiae* <SEQ ID 2153> which encodes the amino acid sequence <SEQ ID 2154>. This protein is predicted to be branched chain amino acid ABC transporter, ATP-binding protein (livG). Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2057(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 GENPEPT database:

```
>GP:AAD36214 GB:AE001771 branched chain amino acid ABC transporter,
          ATP-binding protein [Thermotoga maritima]
 Identities = 136/271 (50%), Positives = 189/271 (69%), Gaps = 21/271 (7%)

Query: 3   LLEVKNLSKHFGGLTAVGDVSMKLHKGELIGLIGPNGAGKTTLFNLLTGVYLPSKGTISI 62
           LL + +++  FGGL AV D + ++ +GEL+GLIGPNGAGKTT+FN++TG+Y P+KG I
Sbjct: 11  LLLLDHVTMQFGGLVAVDDFTNEIREGELVGLIGPNGAGKTTVFNVITGIYTPTKGRIVF 70

Query: 63  DGKILNGRKPAKIASLGLGRTFQNIRLFKNMTVLDNVLVGLSNHHLSHPIASFLRLPK-- 120
           +   + G +P +I  LG+ RTFQNIRLF +MTVL+NVLV   +H LS+P A  + +
Sbjct: 71  NDIDITGLRPYQITHLGIARTFQNIRLFSDMTVLENVLVA-QHHVLSNPDADRILVKHGK 129

Query: 121 ------------------YYHSEKALRKKALELLEIFGLKAYQDALAKNLPYGKQRRLEI 162
                             Y   EK + ++A +L++  GL+      A +LPYG+QR+LEI
Sbjct: 130 PRKGHGRFWFWRAVTKIGYLKKEKEMVERAKDLIKRVGLEKVMYEKASSLPYGEQRKLEI 189

Query: 163 VRALATEPKILFLDEPAAGMNPQETAELTQLISQIKSDFDITIMLIEHDMNLVMQVTERI 222
            RALATEPK++ LDEPAAGMNP+ET +L + I QI+ DF++T++LIEHDM +VM + ERI
Sbjct: 190 ARALATEPKLILLDEPAAGMNPKETEDLMEFIKQIRKDFNLTVLLIEHDMKVVMGICERI 249

Query: 223 YVLEYGRLIAHGTPEEIKNNKRVIEAYLGGE 253
            V++YGR+IA GTP+EI+N+ RVIEAYLG E
Sbjct: 250 IVMDYGRIIAEGTPKEIQNDPRVIEAYLGRE 280
```

There is also homology to SEQ ID 644.

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

Example 700

A DNA sequence (GBSx0744) was identified in *S.agalactiae* <SEQ ID 2155> which encodes the amino acid sequence <SEQ ID 2156>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2216(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 GENPEPT database:

```
>GP:CAB52068 GB:AL109732 putative branched chain amino acid
          transport ATP-binding protein [Streptomyces coelicolor
          A3(2)]
 Identities = 136/233 (58%), Positives = 181/233 (77%)
```

```
Query: 3   MLKVENLSIHYGVIQAVNDVSFEVNQGEVVTLIGANGAGKTSILRTISGLVRPSQGSISF 62
           +L+VE+L + YG I+AV  +SF+V+ GEVVTLIG NGAGKT+ LRT+SGL++P  G I F
Sbjct: 4   LLEVEDLRVAYGKIEAVKGISFKVDAGEVVTLIGTNGAGKTTTLRTLSGLLKPVGGQIRF 63

Query: 63  MGKPIHKLAARKIVGNGLAQVPEGRHVFSSLSVMENLEMGAFLQKDREQNQKMLKKVFDR 122
            GK + K+ A +IV  GLA  PEGRH+F  +++ +NL +GAFL+ DR   +K +++ +D
Sbjct: 64  GGKSLKKVPAHQIVSLGLAHSPEGRHIFPRMTIEDNLRLGAFLRSDRPGIEKDIQRAYDL 123

Query: 123 FPRLEERKNQDAATLSGGEQQMLAMGRALMSRPKLLLLDEPSMGLAPIFIQEIFNIIEDI 182
           FP L ER+ Q A TLSGGEQQMLAMGRALMS+PKLL+LDEPSMGL+PI +Q+I   I ++
Sbjct: 124 FPILGERRKQAAGTLSGGEQQMLAMGRALMSQPKLLMLDEPSMGLSPIMMQKIMATIAEL 183

Query: 183 KKQGTTVLLVEQNANKALTIADKAYVLETGKVVLSGTGKELLVSDQVRKAYLG 235
           K QGTT+LLVEQNA  AL++AD  +V+E G +VLSG+G++LL  + VRKAYLG
Sbjct: 184 KSQGTTILLVEQNAQAALSLADHGHVMEVGNIVLSGSGQDLLHDESVRKAYLG 236
```

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

Example 701

A DNA sequence (GBSx0745) was identified in *S.agalactiae* <SEQ ID 2159> which encodes the amino acid sequence <SEQ ID 2160>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0415(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 GENPEPT database:

```
>GP:AAD36216 GB:AE001771 conserved hypothetical protein [Thermotoga maritima]
 Identities = 72/166 (43%), Positives = 116/166 (69%), Gaps = 2/166 (1%)

Query: 1   MPVKDFMTKKLVYVSPDTTVAEAADLLREHHLRRLPVVENDQLVGLVTEGTMAEAQPSKA 60
           M VKDFMT+  + ++P+T+ +EA  L++++ ++RL V++N+++VG+VTE  + A PSKA
Sbjct: 1   MLVKDFMTRNPITIAPETSFSEALKLMKQNKIKRLIVMKNEKIVGIVTEKDLLYASPSKA 60

Query: 61  TSLSIYEMNYLLNKTKIRDIMIKDIVTVSQYASLEDAIYLMMSRKIGVLPVVDN-GQLYG 119
           T+L+I+E++YLL+K KI +IM KD+VTV++   +EDA  +M  + I  LPVVD+ G+L G
Sbjct: 61  TTLNIWELHYLLSKLKIEEIMTKDVVTVNENTPIEDAARIMEEKDISGLPVVDDAGRLVG 120

Query: 120 IVTDRDVFKAFLEIAGYGQE-SYRLVILADEGIGVLSKVLNRLSSA 164
           I+T  D+FK F+EI G  +E + R  +   + G L +V  R+ A
Sbjct: 121 IITQTDIFKVFVEIFGTKREGTIRYTMEMPDKPGELLEVAKRIYEA 166
```

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

Example 702

A DNA sequence (GBSx0746) was identified in *S.agalactiae* <SEQ ID 2163> which encodes the amino acid sequence <SEQ ID 2164>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5585(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

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

Example 703

A DNA sequence (GBSx0747) was identified in *S.agalactiae* <SEQ ID 2165> which encodes the amino acid sequence <SEQ ID 2166>. This protein is predicted to be a transposase. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.65   Transmembrane   53 -  69 (  53 -  70)

----- 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>
```

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

```
>GP:AAA85003 GB:U28972 SpV1 ORF3; putative transposase [Spiroplasma citri]
 Identities = 49/154 (31%), Positives = 80/154 (51%), Gaps = 11/154 (7%)

Query: 39  WLEMDTVIGRIGGKVLLTFNVAFCNFIFAKLMDSKTAIETAKHIQ--VIKRTLYDNKRDF 96
           WLEMDTV+G+     +L          FA  +++ TA E   K  +  +IK  L    +
Sbjct: 174 WLEMDTVVGKDHKSAILVLVEQLSKKYFAIKLENHTAREVEKKFKDIIIKNNLIGKIKG- 232

Query: 97  FELFPVILTDNGGEFARVDDIEIDVCGQSQLFFCDPNRSDQKARIEKNHTLVRDILPKGT 156
                 I+TD G EF++  ++EI     ++Q++FCD      QK  IE  ++ +R   PKGT
Sbjct: 233 ------IITDRGKEFSKWREMEI--FAETQVYFCDAGSPQQKPLIEYMNSELRHWFPKGT 284

Query: 157 SFDNLTQEDINLALSHINSVKRQALNGKTAYELF 190
            F+ ++Q+ I+  ++ IN   R  LN  ++ E+F
Sbjct: 285 DFNKVSQKQIDWVVNVINDKLRPCLNWISSKEMF 318
```

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

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

Example 704

A DNA sequence (GBSx0748) was identified in *S.agalactiae* <SEQ ID 2167> which encodes the amino acid sequence <SEQ ID 2168>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3116(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10055> which encodes amino acid sequence <SEQ ID 10056> was also identified.

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

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

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

Example 705

A DNA sequence (GBSx0749) was identified in *S.agalactiae* <SEQ ID 2169> which encodes the amino acid sequence <SEQ ID 2170>. This protein is predicted to be thymidylate kinase (tmk). Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1876(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10053> which encodes amino acid sequence <SEQ ID 10054> was also identified.

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

```
>GP:BAB03761 GB:AP001507 thymidylate kinase [Bacillus halodurans]
 Identities = 112/210 (53%), Positives = 148/210 (70%), Gaps = 1/210 (0%)

Query: 17  MKKGLMISFEGPDGAGKTTVLEAVLPLLREKLSQDILTTREPGGVTISEEIRHIILDVKH 76
           M KG  I+ EG +GAGKT+ L+A+  +LRE      ++ TREPGG+ I+E+IR IILDV H
Sbjct: 1   MTKGCFITVEGGEGAGKTSALDAIEEMLRENGLSVVRTREPGGIPIAEQIRSIILDVDH 59

Query: 77  TQMDKKTELLLYMAARRQHLVEKVLPALEEGKIVLMDRFIDSSVAYQGSGRGLDKSHIKW 136
           T+MD +TE LLY AARRQHLVEKVLPALE G +VL DRFIDSS+AYQG  RG+    I
Sbjct: 60  TRMDPRTEALLYAAARRQHLVEKVLPALEAGHVVLCDRFIDSSLAYQGYARGIGFEDILA 119

Query: 137 LNDYATDSHKPDLTLYFDVPSEVGLERIQKSVQREVNRLDLEQLDMHQRVRQGYLELADS 196
           +N++A +   PDLTL F V  +VGL RI +   RE NRLD E L  HQ+V++GY  + ++
Sbjct: 120 INEFAIEGRYPDLTLLFRVDPDVGLSRIHRDQSREQNRLDQEALTFHQKVKEGYERIVET 179

Query: 197 EPNRIVTIDASQQLDEVIAETFSIILDRIN 226
            P R+V IDA+Q  D+V+A+   +I  R++
Sbjct: 180 YPERVVEIDANQSFDQVVADAVRMIKQRLS 209
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2171> which encodes the amino acid sequence <SEQ ID 2172>. Analysis of this protein sequence reveals the following:

```
     Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.75   Transmembrane  215 - 231 ( 215 - 231)

----- Final Results -----
              bacterial membrane --- Certainty=0.1298(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:

```
>GP:BAB03761 GB:AP001507 thymidylate kinase [Bacillus halodurans]
 Identities = 109/205 (53%), Positives = 148/205 (72%), Gaps = 1/205 (0%)

Query: 22  MITGKLITVEGPDGAGKTTVLEQLIPLLKQKVAQDILTTREPGGVAISEHIRELILDINH 81
```

```
            M  G  ITVEG +GAGKT+ L+ +  +L++      ++ TREPGG+ I+E IR +ILD++H
Sbjct: 1    MTKGCFITVEGGEGAGKTSALDAIEEMLREN-GLSVVRTREPGGIPIAEQIRSIILDVDH 59

Query: 82   TAMDPKTELLLYIAARRQHLVEKVLPALEAGQLVFIDRFIDSSVAYQGAGRGLIKADIQW 141
            T MDP+TE LLY AARRQHLVEKVLPALEAG +V  DRFIDSS+AYQG  RG+   DI
Sbjct: 60   TRMDPRTEALLYAAARRQHLVEKVLPALEAGHVVLCDRFIDSSLAYQGYARGIGFEDILA 119

Query: 142  LNEFATDGLEPDLTLYFDVPSEIGLARINANQQREVNRLDLETIEIHQRVRKGYLALAKE 201
            +NEFA +G  PDLTL F V  ++GL+RI+ +Q RE NRLD E +  HQ+V++GY  + +
Sbjct: 120  INEFAIEGRYPDLTLLFRVDPDVGLSRIHRDQSREQNRLDQEALTFHQKVKEGYERIVET 179

Query: 202  HPKRIVTIDATKPLKEVVSVALEHV 226
            +P+R+V IDA +   +VV+ A+  +
Sbjct: 180  YPERVVEIDANQSFDQVVADAVRMI 204
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 145/219 (66%), Positives = 181/219 (82%)

Query: 4    FDRIVVIINKGCTMKKGLMISFEGPDGAGKTTVLEAVLPLLREKLSQDILTTREPGGVTI 63
            FD+I ++ ++G  M  G +I+ EGPDGAGKTTVLE ++PLL++K++QDILTTREPGGV I
Sbjct: 9    FDKIELLKSEGNKMITGKLITVEGPDGAGKTTVLEQLIPLLKQKVAQDILTTREPGGVAI 68

Query: 64   SEEIRHIILDVKHTQMDKKTELLLYMAARRQHLVEKVLPALEEGKIVLMDRFIDSSVAYQ 123
            SE IR +ILD+ HT MD KTELLLY+AARRQHLVEKVLPALE G++V +DRFIDSSVAYQ
Sbjct: 69   SEHIRELILDINHTAMDPKTELLLYIAARRQHLVEKVLPALEAGQLVFIDRFIDSSVAYQ 128

Query: 124  GSGRGLDKSHIKWLNDYATDSHKPDLTLYFDVPSEVGLERIQKSVQREVNRLDLEQLDMH 183
            G+GRGL K+ I+WLN++ATD  +PDLTLYFDVPSE+GL RI  + QREVNRLDLE +++H
Sbjct: 129  GAGRGLIKADIQWLNEFATDGLEPDLTLYFDVPSEIGLARINANQQREVNRLDLETIEIH 188

Query: 184  QRVRQGYLELADSEPNRIVTIDASQQLDEVIAETFSIIL 222
            QRVR+GYL LA    P RIVTIDA++ L EV++      +L
Sbjct: 189  QRVRKGYLALAKEHPKRIVTIDATKPLKEVVSVALEHVL 227
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 706

A DNA sequence (GBSx0750) was identified in *S.agalactiae* <SEQ ID 2173> which encodes the amino acid sequence <SEQ ID 2174>. This protein is predicted to be DNA polymerase III delta' subunit (dnaZX). Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2603(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 GENPEPT database:

```
>GP:BAB03763 GB:AP001507 DNA polymerase III delta' subunit [Bacillus halodurans]
 Identities = 78/189 (41%), Positives = 113/189 (59%), Gaps = 3/189 (1%)

Query: 2    DLKRTQPKLLEKFNTILQSDRMSHAYLFSGNFAS--LDMALYLAQSQFCEKRQSGLPCQE 59
            +L + QP +       L   R++HAY+F GN  +    MAL+LA+S FC +R    PCQ
Sbjct: 5    NLAKNQPFVATMLKNSLAKGRLAHAYIFDGNRGTGKKRMALHLAKSFFCAQRAGVEPCQT 64

Query: 60   CRACRLIANGEFSDVKIIEPQGQLIKTETIKELTKDFSRSGFEGKSQVFIIKDCEKMHVN 119
            C+ C+ I +G   DV  IEP GQ IK   ++ L K+FS  G E   +V+I+   +KM  +
Sbjct: 65   CKECKRIEHGNHPDVHFIEPDGQSIKKHQVEHLQKEFSYRGMESAKKVYIVNHADKMTTS 124
```

```
Query: 120 AANSLLKFIEEPQSSSYVILLTNDENNVLPTIKSRTQIFRF-PKQLDMLVHQAEQAGLLK 178
           AANSLLKF+EEP + +  ILLT    N+LPTIKSR+Q+  F P ++       E+ G+ +
Sbjct: 125 AANSLLKFLEEPLADTVAILLTEQLQNMLPTIKSRSQVLSFAPLEVQAFAKLLEEEGISE 184

Query: 179 SQASLLAQV 187
           S ++LLA +
Sbjct: 185 SVSNLLASL 193
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2175> which encodes the amino acid sequence <SEQ ID 2176>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2685(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 151/290 (52%), Positives = 213/290 (73%), Gaps = 3/290 (1%)

Query: 1   MDLKRTQPKLLEKFNTILQSDRMSHAYLFSGNFASLDMALYLAQSQFCEKRQSGLPCQEC 60
           MDL +  P + + F TIL+ DR++HAYLFSG+FA+ +MAL+LA+  FCE+++    PC  C
Sbjct: 1   MDLAQKAPNVYQAFQTILKKDRLNHAYLFSGDFANEEMALFLAKVIFCEQKKDQTPCGHC 60

Query: 61  RACRLIANGEFSDVKIIEPQGQLIKTETIKELTKDFSRSGFEGKSQVFIIKDCEKMHVNA 120
           R+C+LI  G+F+DV ++EP GQ+IKT+ +KE+  +FS++G+E K QVFIIKDC+KMH+NA
Sbjct: 61  RSCQLIEQGDFADVTVLEPTGQVIKTDVVKEMMANFSQTGYENKRQVFIIKDCDKMHINA 120

Query: 121 ANSLLKFIEEPQSSSYVILLTNDENNVLPTIKSRTQIFRFPKQLDMLVHQAEQAGLLKSQ 180
           ANSLLK+IEEPQ  +Y+ LLTND+N VLPTIKSRTQ+F+FPK    L   A++ GLL  Q
Sbjct: 121 ANSLLKYIEEPQGEAYIFLLTNDDNKVLPTIKSRTQVFQFPKNEAYLYQLAQEKGLLNHQ 180

Query: 181 ASLLAQVADDPKHLEILLTNKKLLDYLNLSQQFVTTLAKDRQTAYLEVSRLTSQVVDKND 240
           A L+A++A +  HLE LL   KLL+ +  +++FV+   KD+  AYL ++RL     +K +
Sbjct: 181 AKLVAKLATNTSHLERLLQTSKLLELITQAERFVSIWLKDQLQAYLALNRLVQLATEKEE 240

Query: 241 QAFVFQWLTIMLAKE---GQLYDLENTYRAQQMWKSNVSFQNSLEYMVLS 287
           Q  V   LT++LA+E     L  LE  Y+A+ MW+SNV+FQN+LEYMV+S
Sbjct: 241 QDLVLTLLTLLLARERAQTPLTQLEAVYQARLMWQSNVNFQNTLEYMVMS 290
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 707

A DNA sequence (GBSx0751) was identified in *S.agalactiae* <SEQ ID 2177> which encodes the amino acid sequence <SEQ ID 2178>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2016(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 GENPEPT database:

```
>GP:BAB03765 GB:AP001507 unknown conserved protein in B. subtilis
          [Bacillus halodurans]
 Identities = 45/116 (38%), Positives = 62/116 (52%), Gaps = 8/116 (6%)
```

```
Query: 1   MDKKDLFDAFDDFSQNLLVGLSEIETMKKQIQKLLEENTVLRIENGKLRERLSVIEAET- 59
           M+KK +F        + +      E+  +K+Q+  L+EEN  L IEN  LRERL   E E
Sbjct: 1   MNKKAIFTQVSQLEERIGELHRELGGLKEQLAYLIEENHFLTIENEHLRERLGEPELEET 60

Query: 60  ---ETAVKNSK----QGRELLEGIYNDGFHICNTFYGQRRENDEECAFCIELLYRD 108
              E   K  K     +G + L   +Y +GFHICNT YG  R+N E+C FC+  L +D
Sbjct: 61  EEKEQVTKERKPFVGEGYDNLARLYQEGFHICNTHYGSLRKNGEDCLFCLSFLNQD 116
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2179> which encodes the amino acid sequence <SEQ ID 2180>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0700(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 75/107 (70%), Positives = 89/107 (83%), Gaps = 1/107 (0%)

Query: 1   MDKKDLFDAFDDFSQNLLVGLSEIETMKKQIQKLLEENTVLRIENGKLRERLSVIEAETE 60
           ++KK+LFDAFD FSQNL+V L+EIE MKKQ+Q L+EENT+LR+EN KLRERLS +E ET
Sbjct: 1   VNKKELFDAFDGFSQNLMVTLAEIEAMKKQVQSLVEENTILRLENTKLRERLSHLEHET- 59

Query: 61  TAVKNSKQGRELLEGIYNDGFHICNTFYGQRRENDEECAFCIELLYR 107
            A   SKQ ++ LEGIY++GFHICN FYGQRRENDEEC FC ELL R
Sbjct: 60  VAKNPSKQRKDHLEGIYDEGFHICNFFYGQRRENDEECMFCRELLDR 106
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 708

A DNA sequence (GBSx0752) was identified in *S.agalactiae* <SEQ ID 2181> which encodes the amino acid sequence <SEQ ID 2182>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.28   Transmembrane  119 - 135 ( 119 - 135)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 10051> which encodes amino acid sequence <SEQ ID 10052> was also identified.

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

```
>GP:BAB03768 GB:AP001507 unknown conserved protein [Bacillus halodurans]
 Identities = 138/287 (48%), Positives = 189/287 (65%), Gaps = 2/287 (0%)

Query: 4   MQVQKSFKSNIHYGTLYLVPTPIGNLDDMTFRAIRILREVDFICAEDTRNTGLLLKHFDI 63
           M+ Q+S++      GTLYLV TPIGNL+D+TFRAIR L+E D I AEDTR T  LL HFDI
Sbjct: 1   MKTQQSYQQRDDKGTLYLVATPIGNLEDVTFRAIRTLKEADQIAAEDTRQTKKLLNHFDI 60

Query: 64  TTKQISFHEHNAYDKISGLIDLLKEGKSLAQVSDAGMPSISDPGHDLVKAAIEGDIPVVS 123
            TK +S+HEHN        LID L EG+++A VSDAGMP+ISDPG++LV +AI+  I V+
Sbjct: 61  ATKLVSYHEHNKETMGKRLIDDLIEGRTIALVSDAGMPAISDPGYELVVSAIKEGIAVIP 120
```

```
Query: 124 IPGASAGITALIASGLAPQPHIFYGFLPRKKGQQITFFETKQDYPETQIFYESPFRVSDT 183
           IPGA+A +TALIASGL  +   F GFLPR+K Q+    E  +    T IFYESP R+ DT
Sbjct: 121 IPGANAAVTALIASGLPTESFQFIGFLPRQKKQRRQALEETKPTKATLIFYESPHRLKDT 180

Query: 184 LKHMKEIYGDRQVVLVRELTKLYEEYQRGTISQLLEHIEKVPLKGECLIIVDGKRDTERV 243
           L  M  I G+R V + RELTK YEE+ RGT+ + +    +  +KGE  +IV+G  +
Sbjct: 181 LDDMLLILGNRHVSICRELTKTYEEFLRGTLEEAVHWAREATIKGEFCLIVEGNGEKVEP 240

Query: 244 KDS--SQQDPLVLVKEYIANGDKTNQAIKKVAKEFNLNRQELYASFH 288
           ++       P+  V+ YIA G ++ +AIK+VA +  + ++++Y  +H
Sbjct: 241 EEVWWESLSPVQHVEHYIALGFRSKEAIKQVATDRGVPKRDIYNIYH 287
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2183> which encodes the amino acid sequence <SEQ ID 2184>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.09   Transmembrane  116 - 132 ( 116 - 134)

----- 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>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB03768 GB:AP001507 unknown conserved protein [Bacillus halodurans]
 Identities = 139/287 (48%), Positives = 189/287 (65%), Gaps = 2/287 (0%)

Query: 1   MQVQKSFKDKKTSGTLYLVPTPIGNLQDMTFRAVATLKEVDFICAEDTRNTGLLLKHFDI 60
           M+ Q+S++ +   GTLYLV TPIGNL+D+TFRA+ TLKE D I AEDTR T  LL HFDI
Sbjct: 1   MKTQQSYQQRDDKGTLYLVATPIGNLEDVTFRAIRTLKEADQIAAEDTRQTKKLLNHFDI 60

Query: 61  ATKQISFHEHNAYEKIPDLIDLLISGRSLAQVSDAGMPSISDPGHDLVKAAIDSDIAVVA 120
           ATK +S+HEHN        LID LI GR++A VSDAGMP+ISDPG++LV +AI   IAV+
Sbjct: 61  ATKLVSYHEHNKETMGKRLIDDLIEGRTIALVSDAGMPAISDPGYELVVSAIKEGIAVIP 120

Query: 121 LPGASAGITALIASGLAPQPHVFYGFLPRKAGQQKAFFEDKHHYPETQMFYESPYRIKDT 180
           +PGA+A +TALIASGL  +   F GFLPR+  Q++    E+       T +FYESP+R+KDT
Sbjct: 121 IPGANAAVTALIASGLPTESFQFIGFLPRQKKQRRQALEETKPTKATLIFYESPHRLKDT 180

Query: 181 LTNMLACYGDRQVVLVRELTKLFEEYQRGSISEILSYLEETPLKGECLLIVA--GAQADS 238
           L +ML   G+R V + RELTK +EE+ RG++ E + +  E  +KGE  LIV   G + + +
Sbjct: 181 LDDMLLILGNRHVSICRELTKTYEEFLRGTLEEAVHWAREATIKGEFCLIVEGNGEKVEP 240

Query: 239 EVELTADVDLVSLVQKEIQAGAKPNQAIKTIAKAYQVNRQELYQQFH 285
           E      +  V  V+  I  G +  +AIK +A    V ++++Y  +H
Sbjct: 241 EEVWWESLSPVQHVEHYIALGFRSKEAIKQVATDRGVPKRDIYNIYH 287
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 208/287 (72%), Positives = 238/287 (82%)

Query: 4   MQVQKSFKSNIHYGTLYLVPTPIGNLDDMTFRAIRILREVDFICAEDTRNTGLLLKHFDI 63
           MQVQKSFK     GTLYLVPTPIGNL DMTFRA+  L+EVDFICAEDTRNTGLLLKHFDI
Sbjct: 1   MQVQKSFKDKKTSGTLYLVPTPIGNLQDMTFRAVATLKEVDFICAEDTRNTGLLLKHFDI 60

Query: 64  TTKQISFHEHNAYDKISGLIDLLKEGKSLAQVSDAGMPSISDPGHDLVKAAIEGDIPVVS 123
            TKQISFHEHNAY+KI  LIDLL  G+SLAQVSDAGMPSISDPGHDLVKAAI+ DI VV+
Sbjct: 61  ATKQISFHEHNAYEKIPDLIDLLISGRSLAQVSDAGMPSISDPGHDLVKAAIDSDIAVVA 120

Query: 124 IPGASAGITALIASGLAPQPHIFYGFLPRKKGQQITFFETKQDYPETQIFYESPFRVSDT 183
           +PGASAGITALIASGLAPQPH+FYGFLPRK GQQ  FFE K  YPETQ+FYESP+R+ DT
Sbjct: 121 LPGASAGITALIASGLAPQPHVFYGFLPRKAGQQKAFFEDKHHYPETQMFYESPYRIKDT 180

Query: 184 LKHMKEIYGDRQVVLVRELTKLYEEYQRGTISQLLEHIEKVPLKGECLIIVDGKRDTERV 243
           L +M   YGDRQVVLVRELTKL+EEYQRG+IS++L ++E+ PLKGECL+IV G +    V
```

```
Sbjct: 181 LTNMLACYGDRQVVLVRELTKLFEEYQRGSISEILSYLEETPLKGECLLIVAGAQADSEV 240

Query: 244 KDSSQQDPLVLVKEYIANGDKTNQAIKKVAKEFNLNRQELYASFHDL 290
           + ++  D + LV++ I  G K NQAIK +AK + +NRQELY  FHDL
Sbjct: 241 ELTADVDLVSLVQKEIQAGAKPNQAIKTIAKAYQVNRQELYQQFHDL 287
```

A related GBS gene <SEQ ID 8643> and protein <SEQ ID 8644> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      -6.92
GvH: Signal Score (-7.5): -9.26
     Possible site: 48
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -1.28 threshold:  0.0
   INTEGRAL    Likelihood = -1.28   Transmembrane  118 - 134 ( 118 - 134)
   PERIPHERAL  Likelihood =  6.89       32
 modified ALOM score:   0.76

*** Reasoning Step: 3

----- 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>
```

The protein has homology with the following sequences in the databases:

```
ORF00263(310 - 1164 of 1470)
EGAD|17863|BS0036(2 - 289 of 292) hypothetical 33.0 kd protein in xpac-abrb intergenic
region (Bacillus subtilis) OMNI|NT01BS0044 conserved hypothetical protein
SP|P37544|YABC_BACSU HYPOTHETICAL 33.0 KDA PROTEIN IN XPAC-ABRB INTERGENIC REGION.
GP|467425|dbj|BAA05271.1||D26185 unknown (Bacillus subtilis)
GP|2632303|emb|CAB11812.1||Z99104 similar to hypothetical proteins (Bacillus subtilis)
PIR|S66065|S66065 conserved hypothetical protein yabC - Bacillus subtilis
%Match = 24.5
%Identity = 45.8  %Similarity = 65.7
Matches = 131  Mismatches = 97  Conservative Sub.s = 57

123       153       183       213       243       273       303       333
CSTH*KW*TS*ASERY*SRNRNCS*KF*TRKRITRRHLQ*WLSHL*YFLWSTS*K*RRMCFLY*III*RLMEMQVQKSFK
                                                                      :: | ||
                                                                     MLRRQMSFN

363       393       423       453       483       513       543       573
SNIHYGTLYLVPTPIGNLDDMTFRAIRILREVDFICAEDTRNTGLLLKHFDITTKQISFHEHNAYDKISGLIDLLKEGKS
     | ||||||||||:|||||||  |: || | ||||| |  |   ::| |  :|:||||       :|: || ||:
GKSDMGILYLVPTPIGNLEDMTFRAIDTLKSVDAIAAEDTRQTKKLCHVYEIETPLVSYHEHNKESSGHKIIEWLKSGKN
         20        30        40        50        60        70        80

603       633       663       693       723       753       783       813
LAQVSDAGMPSISDPGHDLVKAAIEGDIPVVSIPGASAGITALIASGLAPQPHIFYGFLPRKKGQQITFFETKQDYPETQ
:| |||||:|:||||| ::||   :    || :|||:| :|||||||: ||| |||||  |:| ::   :|  :   ||
IALVSDAGLPTISDPGAEIVKDFTDIGGYVVPLPGANAALTALIASGIVPQPFFFYGFLNRQKKEKKKELEALKKRQETI
         100       110       120       130       140       150       160

843       873       903       933       963       993       1023      1053
IFYESPFRVSDTLKHMKEIYGDRQVVLVRELTKLYEEYQRGTISQLLEHIEKVPLKGECLIIVDGKRDTERVKDSSQQDP
||||:| |: :||  | || |||:: : ||||| |||: |||||:::    :  ::||  ::|:|  : |  ::    :
IFYEAPHRLKETLSAMAEILGDREIAVTRELTKKYEEFIRGTISEVIGWANEDQIRGEFCLVVEGSNNEEVDEEEQWWET
         180       190       200       210       220       230       240

1074      1104      1134      1164      1194      1224      1254      1284
LVL---VKEYIANGDKTNQAIKKVAKEFNLNRQELYASFHDL*VII*KGCQRKIWQPFIISDLAIGIKK*DTSNFLKIFN
|     |: ||: |  : :|||| | : |: ::|:| ::|
LTAKEHVEHYISKGATSKEAIKKAAVDRNVPKREVYDAYHIKQ
         260       270       280       290
```

SEQ ID 8644 (GBS343) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 11; MW 35.4kDa).

The GBS343-His fusion product was purified (Figure 215, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 277), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 709

A DNA sequence (GBSx0753) was identified in *S.agalactiae* <SEQ ID 2185> which encodes the amino acid sequence <SEQ ID 2186>. This protein is predicted to be bA483F11.3 (cutC). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2568(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 GENPEPT database:

```
>GP:CAB88199 GB:AL133353 bA483F11.3 (CGI-32 protein ) [Homo sapiens]
 Identities = 79/203 (38%), Positives = 116/203 (56%), Gaps = 7/203 (3%)

Query: 3   LREFCAENLTDLTRLDKAIISRVELCDNLAVGGTTPSYGVIKEANQYLHEKGISVAVMIR 62
           L E C +++       ++    R+ELC  L+ GGTTPS GV++   Q +    I V VMIR
Sbjct: 27  LMEVCVDSVESAVNAERGGADRIELCSGLSEGGTTPSMGVLQVVKQSVQ---IPVFVMIR 83

Query: 63  PRGGNFVYNDLELRIMEEDILRAVELESDALVLGILTSNNHIDTEAIEQLLPATQGLPLV 122
           PRGG+F+Y+D E+ +M+ DI  A    +D LV G LT + HID E    L+   + LP+
Sbjct: 84  PRGGDFLYSDREIEVMKADIRLAKLYGADGLVFGALTEDGHIDKELCMSLMAICRPLPVT 143

Query: 123 FHMAFDVIPKSDQKKSIDQLVALGFTRILLHGSSNGEPIIENIKHIKALVEYANNRIEIM 182
           FH AFD++   D   +++ L+ LGF R+L  G  +    +E +  IK L+E A  RI +M
Sbjct: 144 FHRAFDMV--HDPMAALETLLTLGFERVLTSGCDSS--ALEGLPLIKRLIEQAKGRIVVM 199

Query: 183 VGGGVTAENYQYICQETGVKQAH 205
            GGG+T  N Q I + +G  + H
Sbjct: 200 PGGGITDRNLQRILEGSGATEFH 222
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2187> which encodes the amino acid sequence <SEQ ID 2188>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2372(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 143/208 (68%), Positives = 168/208 (80%)

Query: 2   ILREFCAENLTDLTRLDKAIISRVELCDNLAVGGTTPSYGVIKEANQYLHEKGISVAVMI 61
           +++EFCAENLT L  LD   ISRVELCDNLAVGGTTPSYGVIKEA Q LH+K ISVA MI
```

```
Sbjct: 1   MIKEFCAENLTLLPTLDAGQISRVELCDNLAVGGTTPSYGVIKEACQLLHDKKISVATMI 60

Query: 62  RPRGGNFVYNDLELRIMEEDILRAVELESDALVLGILTSNNHIDTEAIEQLLPATQGLPL 121
           RPRGG+FVYNDLEL+ MEEDIL+AVE  SDALVLG+LT+ N +DT+AIEQLLPATQGLPL
Sbjct: 61  RPRGGDFVYNDLELKAMEEDILKAVEAGSDALVLGLLTTENQLDTDAIEQLLPATQGLPL 120

Query: 122 VFHMAFDVIPKSDQKKSIDQLVALGFTRILLHGSSNGEPIIENIKHIKALVEYANNRIEI 181
           VFHMAFD IP   Q +++DQL+  GF R+L HGS    PI +N++ +K+LV YAN RIEI
Sbjct: 121 VFHMAFDRIPTDHQHQALDQLIDYGFVRVLTHGSPEATPITDNVEQLKSLVTYANKRIEI 180

Query: 182 MVGGGVTAENYQYICQETGVKQAHGTRI 209
           M+GGG+TAEN Q + Q TG    HGT+I
Sbjct: 181 MIGGGITAENCQSLSQLTGTAIVHGTKI 208
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 710

A DNA sequence (GBSx0754) was identified in *S.agalactiae* <SEQ ID 2189> which encodes the amino acid sequence <SEQ ID 2190>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1216(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 GENPEPT database:

```
>GP:BAA12206 GB:D84061 phosphoserine aminotransferase [Spinacia
           oleracea]
 Identities = 65/109 (59%), Positives = 79/109 (71%), Gaps = 1/109 (0%)

Query: 3   IYNFSAGPAVLPKPVLVKAQSELLNYQGSSMSVLEVSHRSKEFDDIIKGAERYLRDLMGI 62
           ++NF+AGPAVLP+ VL KAQSELLN++GS MSV+E+SHR KEF  II  AE  LR L+ I
Sbjct: 69  VFNFAAGPAVLPENVLQKAQSELLNWRGSGMSVMEMSHRGKEFTSIIDKAEADLRTLLNI 128

Query: 63  PDNYKVIFLQGGASLQFSMIPLNIARGRKAY-YHVAGSWGEKSLYRGCK 110
           P +Y V+FLQGGAS QFS IPLN+     A  Y V GSWG+K+     K
Sbjct: 129 PSDYTVLFLQGGASTQFSAIPLNLCTPDSAVDYIVTGSWGDKAAKEAAK 177
```

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

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

Example 711

A DNA sequence (GBSx0755) was identified in *S.agalactiae* <SEQ ID 2191> which encodes the amino acid sequence <SEQ ID 2192>. Analysis of this protein 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>
```

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

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

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

Example 712

A DNA sequence (GBSx0756) was identified in *S.agalactiae* <SEQ ID 2193> which encodes the amino acid sequence <SEQ ID 2194>. This protein is predicted to be phosphoserine aminotransferase (serC). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3380(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10049> which encodes amino acid sequence <SEQ ID 10050> was also identified.

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

```
>GP:AAF94318 GB:AE004196 phosphoserine aminotransferase [Vibrio cholerae]
 Identities = 104/210 (49%), Positives = 152/210 (71%), Gaps = 3/210 (1%)

Query: 4   NNTIEGTSLYDIPKTNEVPVIADMSSNILAVKYKVEDFAMIYAGAQKNIGPAGVTVVIIR 63
           N TI+G  + D+P T++ P++ADMSS IL+ +  V  + +IYAGAQKNIGPAG+ + I+R
Sbjct: 170 NETIDGIEINDLPVTDK-PIVADMSSTILSREIDVSKYGVIYAGAQKNIGPAGICIAIVR 228

Query: 64  EDMIN-EEPTLSSMLDYKIQSDAGSLYNTPPAYSIYIAKLVFEWVKSLGGVDAMEKANRE 122
           +D+++     L  +L+YKI ++  S++NTPP ++ Y++ LVF+W+K+ GGV A+E+ NR
Sbjct: 229 DDLLDLASDLLPGVLNYKILAEQESMFNTPPTFAWYLSGLVFQWLKAQGGVKAIEEVNRA 288

Query: 123 KSGLLYDYIDSSEFYSNPVRDKKSRSLCNIPFITINKDLDEKFVKEATERGFKNIKGHRS 182
           K+ LLY YIDSS+FY N +     +RSL N+PF      +LD+ F++ A  RG  ++KGHR
Sbjct: 289 KAALLYGYIDSSDFYRNEIH-PDNRSLMNVPFQLAKPELDDTFLELAEARGLVSLKGHRV 347

Query: 183 VGGMRASLYNAFPKQGVIELIDFMKTFEAE 212
           VGGMRAS+YNA P +GV  L+DFMK FEA+
Sbjct: 348 VGGMRASIYNAMPLEGVQALVDFMKEFEAQ 377
```

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 713

A DNA sequence (GBSx0757) was identified in *S.agalactiae* <SEQ ID 2195> which encodes the amino acid sequence <SEQ ID 2196>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0466(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10047> which encodes amino acid sequence <SEQ ID 10048> was also identified.

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

```
>GP:CAB73701 GB:AL139079 putative acetyltransferase [Campylobacter
          jejuni]
 Identities = 46/170 (27%), Positives = 78/170 (45%), Gaps = 13/170 (7%)

Query: 7   IRLAFPNEIDQIMLLIEEARAEIAKTGSDQWQKEDGYPNRNDIIDDILNGYAWVGIEDGM 66
           I+ A   +++ I+ + ++A   +      QW  ++ YPN     +DI     +V  E+
Sbjct: 6   IQKAVNKDLNSILEITKDALNAMKTMNFHQW--DENYPNEIVFQEDIQAQELYVFKENDE 63

Query: 67  LATYAAVIDGHE-EVYDAIYEGKWLHDNHRYLTFHRIAISNQFRGRGLAQTFLQGL---- 121
           +  +  + + E Y  +   K   D   YL  HR+A+    +G+G+AQ  L
Sbjct: 64  ILGFICINEKFKPEFYKQVIFNKNYDDKAFYL--HRLAVKQNAKGKGVAQKLLNFCENFA 121

Query: 122 IEGHKGPDFRCDTHEKNVTMQHILNKLGYQYCGKVPLDGVR---LAYQKI 168
           +E HK    R DTH KN  M  +  KL + +CG   +   +   LAY+KI
Sbjct: 122 LENHKA-SLRADTHSKNFPMNSLFKKLDFNFCGNFDIPNYQDPFLAYEKI 170
```

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 714

A DNA sequence (GBSx0758) was identified in *S.agalactiae* <SEQ ID 2197> which encodes the amino acid sequence <SEQ ID 2198>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2968(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 715

A DNA sequence (GBSx0759) was identified in *S.agalactiae* <SEQ ID 2199> which encodes the amino acid sequence <SEQ ID 2200>. This protein is predicted to be D-3-phosphoglycerate dehydrogenase (serA). Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3102(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10045> which encodes amino acid sequence <SEQ ID 10046> was also identified.

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

```
>GP:AAB99020 GB:U67544 phosphoglycerate dehydrogenase (serA)
          [Methanococcus jannaschii]
 Identities = 102/313 (32%), Positives = 168/313 (53%), Gaps = 21/313 (6%)

Query: 31  ENPDAYIIRSQNLHNQDF---PSNLKAIARAGAGTNNIPIEEASAQGIVVFNTPGANANA 87
           ++ D  ++RS     +D       LK I RAG G +NI +E A+ +GI+V N P A++ +
Sbjct: 40  KDADVLVVRSGTKVTRDVIEKAEKLKVIGRAGVGVDNIDVEAATEKGIIVVNAPDASSIS 99

Query: 88  VKEAVIAALLLSARDYLGANRWVNTLTGTDIPKQIEAGKKAFAGNEIAGKKLGVIGLGAI 147
           V E  +  +L +AR          N    T   K+ E  +K F G E+ GK LGVIGLG I
Sbjct: 100 VAELTMGLMLAAAR---------NIPQATASLKRGEWDRKRFKGIELYGKTLGVIGLGRI 150

Query: 148 GARIANDARRLGMTVLGYDPYVSIETAWNISSHVQRVKEIKDIFETCDYITIHVPLTNET 207
           G ++   A+  GM ++GYDPY+  E A ++   V+ V +I ++ +  D+IT+HVPLT +T
Sbjct: 151 GQQVVKRAKAFGMNIIGYDPYIPKEVAESMG--VELVDDINELCKRADFITLHVPLTPKT 208

Query: 208 KHTFDAKAFSIMKKGTTIINFARAELVNNQELFEAIETGVVKRYITDFGDKE------LL 261
           +H    +  ++MKK   I+N AR  L++ + L+EA++ G ++    D  ++E      LL
Sbjct: 209 RHIIGREQIALMKKNAIIVNCARGGLIDEKALYEALKEGKIRAAALDVFEEEPPKDNPLL 268

Query: 262 NQKGITVFPHVGGSTDEAELNCAIMASQTIRCFMETGEITNSVNFPNVHQIQTAPFR-IT 320
               +   PH G ST+EA+     + ++ I+  +     N VN PN+ Q +    +
Sbjct: 269 TLDNVIGTPHQGASTEEAQKAAGTIVAEQIKKVLRGELAENVVNMPNIPQEKLGKLKPYM 328

Query: 321 LINKNVPNIVAKI 333
           L+ + + NIV ++
Sbjct: 329 LLAEMLGNIVMQV 341
```

There is also homology to SEQ ID 124.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 716

A DNA sequence (GBSx0760) was identified in *S.agalactiae* <SEQ ID 2201> which encodes the amino acid sequence <SEQ ID 2202>. This protein is predicted to be methylated-DNA--protein-cysteine S-methyltransferase (ogt). Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2460(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 GENPEPT database:

```
>GP:AAF96913 GB:AE004427 methylated-DNA--protein-cysteine
          S-methyltransferase [Vibrio cholerae]
 Identities = 73/156 (46%), Positives = 99/156 (62%), Gaps = 9/156 (5%)

Query: 7   YQSPLGEIRLLADNLGLSGLYFVGQKYDMLAVNQEEIVNMSNSYTLLGK--KWLDAYFSQ 64
           Y SPLG + L A + GL G++F  Q          E + +       +L K  + LD YFS
Sbjct: 7   YSSPLGPMTLQASSQGLLGVWFATQ-----TTQPEHLGDYVKECPILNKTIRQLDEYFSG 61

Query: 65  QNLP-SIPLSLRGTAFQTRVWQELQKIPFGDTKTYGELAKEL-NCQSAQAVGGAIGKNSI 122
           Q     +PL+  GTAFQ  VW  L KIP+G+  +Y +LA+ + N ++ +AVG A GKN I
```

```
Sbjct: 62  QRTQFELPLAASGTAFQQSVWHALCKIPYGEIWSYQQLAEAIGNPKAVRAVGLANGKNPI 121

Query: 123 SLIIPCHRVLGRYGQLTGYAGGLERKSWLLEYEKEK 158
           S+I+PCHRV+G+ GQLTGYAGGLERK++LLE EK +
Sbjct: 122 SIIVPCHRVVGKNGQLTGYAGGLERKAFLLELEKRR 157
```

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

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

Example 717

A DNA sequence (GBSx0761) was identified in *S.agalactiae* <SEQ ID 2203> which encodes the amino acid sequence <SEQ ID 2204>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3137(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 GENPEPT database:

```
>GP:BAB07204 GB:AP001518 arsenate reductase [Bacillus halodurans]
 Identities = 56/107 (52%), Positives = 74/107 (68%), Gaps = 1/107 (0%)

Query: 3   TFYEYPKCTTCRSAKKELTELGLTFEAIDIKSNPPKVSLLKELLENSPYDLKKFFNTSGN 62
           TFY+YPKC TC+ AKK L + G+   ++ I   PP    LK+L E S  +LKKFFNTSG
Sbjct: 4   TFYQYPKCGTCQKAKKWLDQHGIEVNSVHIVEQPPSKEELKQLYEQSGLELKKFFNTSGK 63

Query: 63  SYRELGLKDKFDDLTLDQALDLLASDGMLIKRPLLVKDNKILQIGYR 109
            YRELGLKDK  + + D+ L+ LASDGMLIKRP+L   +K+  +G++
Sbjct: 64  KYRELGLKDKVKEASEDELLETLASDGMLIKRPILTDGDKV-TVGFK 109
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2205> which encodes the amino acid sequence <SEQ ID 2206>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3969(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 64/99 (64%), Positives = 79/99 (79%)

Query: 19  ELTELGLTFEAIDIKSNPPKVSLLKELLENSPYDLKKFFNTSGNSYRELGLKDKFDDLTL 78
           EL +L   FEAIDIK+NPPK   LK  +E S Y +K FFNTSGNSYRELGLKDK D L+L
Sbjct: 3   ELKQLVSDFEAIDIKANPPKAQDLKHWMETSGYTIKNFFNTSGNSYRELGLKDKIDQLSL 62

Query: 79  DQALDLLASDGMLIKRPLLVKDNKILQIGYRTKYKDLNL 117
           D+A +LLA+DGMLIKRP+L+KD  +LQ+GYR  Y++L+L
Sbjct: 63  DKAAELLATDGMLIKRPILIKDGNVLQVGYRKPYQELDL 101
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 718

A DNA sequence (GBSx0762) was identified in *S.agalactiae* <SEQ ID 2207> which encodes the amino acid sequence <SEQ ID 2208>. This protein is predicted to be exodeoxyribonuclease (exoA). Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1859(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 GENPEPT database:

```
>GP:AAA26879 GB:J04234 exodeoxyribonuclease [Streptococcus pneumoniae]
 Identities = 217/275 (78%), Positives = 245/275 (88%)

Query: 1   MKLISWNIDSLNAALTSESTRALMSRQVIDTLVAEDADIIAIQETKLSAKGPTKKHLEVL 60
           MKLISWNIDSLNAALTS+S RA +S++V+ TLVAE+ADIIAIQETKLSAKGPTKKH+E+L
Sbjct: 1   MKLISWNIDSLNAALTSDSARAKLSQEVLQTLVAENADIIAIQETKLSAKGPTKKHVEIL 60

Query: 61  ETYFPEYDLVWRSSVEPARKGYAGTMFLYRKGLNPIVSFPEIDAPTTMDNEGRIITLELE 120
           E  FP Y+  WRSS EPARKGYAGTMFLY+K L P +SFPEI AP+TMD EGRIITLE +
Sbjct: 61  EELFPGYENTWRSSQEPARKGYAGTMFLYKKELTPTISFPEIGAPSTMDLEGRIITLEFD 120

Query: 121 NCYITQVYTPNAGDGLKRLADRQIWDIKYAEYLATLDSQKPVLATGDYNVAHKEIDLANP 180
             ++TQVYTPNAGDGLKRL +RQ+WD KYAEYLA LD +KPVLATGDYNVAH EIDLANP
Sbjct: 121 AFFVTQVYTPNAGDGLKRLEERQVWDAKYAEYLAELDKEKPVLATGDYNVAHNEIDLANP 180

Query: 181 SSNRRSAGFTAEERQGFTNLLAKGFTDTFRYLHGDVPNVYSWWAQRSRTSKINNTGWRID 240
           +SNRRS GFT EER GFTNLLA GFTDTFR++HGDVP  Y+WWAQRS+TSKINNTGWRID
Sbjct: 181 ASNRRSPGFTDEERAGFTNLLATGFTDTFRHVHGDVPERYTWWAQRSKTSKINNTGWRID 240

Query: 241 YWLTSNRVADKITKSEMIHSGDRQDHTPIILEIEL 275
           YWLTSNR+ADK+TKS+MI SG RQDHTPI+LEI+L
Sbjct: 241 YWLTSNRIADKVTKSDMIDSGARQDHTPIVLEIDL 275
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2209> which encodes the amino acid sequence <SEQ ID 2210>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2181(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 221/275 (80%), Positives = 251/275 (90%)

Query: 1   MKLISWNIDSLNAALTSESTRALMSRQVIDTLVAEDADIIAIQETKLSAKGPTKKHLEVL 60
           MKLISWNIDSLNAALT ES RAL+SR V+DTLVA+DADIIAIQETKLSAKGPTKKH+E L
Sbjct: 1   MKLISWNIDSLNAALTGESPRALLSRAVLDTLVAQDADIIAIQETKLSAKGPTKKHIETL 60

Query: 61  ETYFPEYDLVWRSSVEPARKGYAGTMFLYRKGLNPIVSFPEIDAPTTMDNEGRIITLELE 120
            +YFP Y  VWRSSVEPARKGYAGTMFLY+  LNP+++FPEI APTTMD EGRIITLE E
Sbjct: 61  LSYFPNYLHVWRSSVEPARKGYAGTMFLYKNTLNPVITFPEIGAPTTMDAEGRIITLEFE 120

Query: 121 NCYITQVYTPNAGDGLKRLADRQIWDIKYAEYLATLDSQKPVLATGDYNVAHKEIDLANP 180
           + ++TQVYTPNAGDGL+RL DRQIWD KYA+YL  LD+QKPVLATGDYNVAHKEIDLANP
Sbjct: 121 DFFVTQVYTPNAGDGLRRLDDRQIWDHKYADYLTELDAQKPVLATGDYNVAHKEIDLANP 180
```

```
Query: 181 SSNRRSAGFTAEERQGFTNLLAKGFTDTFRYLHGDVPNVYSWWAQRSRTSKINNTGWRID 240
           +SNRRS GFT EERQGFTNLLA+GFTDTFR++HGD+P+VY+WWAQRS+TSKINNTGWRID
Sbjct: 181 NSNRRSPGFTDEERQGFTNLLARGFTDTFRHVHGDIPHVYTWWAQRSKTSKINNTGWRID 240

Query: 241 YWLTSNRVADKITKSEMIHSGDRQDHTPIILEIEL 275
           YWL SNR+ DK+ +SEMI SG+RQDHTPI+L+I+L
Sbjct: 241 YWLASNRLVDKVKRSEMISSGERQDHTPILLDIDL 275
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 719

A DNA sequence (GBSx0763) was identified in *S.agalactiae* <SEQ ID 2211> which encodes the amino acid sequence <SEQ ID 2212>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.96   Transmembrane   28 -  44 (  22 -  49)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 8645> which encodes amino acid sequence <SEQ ID 8646> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 5
McG: Discrim Score:     17.78
GvH: Signal Score (-7.5): -4.56
     Possible site: 55
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -7.96 threshold:  0.0
   INTEGRAL    Likelihood = -7.96   Transmembrane    8 -  24 (   2 -  29)
   PERIPHERAL  Likelihood =  9.28     138
 modified ALOM score:   2.09

*** Reasoning Step: 3

----- 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>
```

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

```
>GP:AAD11512 GB:U60828 unknown [Lactococcus lactis]
 Identities = 53/240 (22%), Positives = 102/240 (42%), Gaps = 24/240 (10%)

Query: 65  PTILIPGSSATQERFNSMLAQL----NQMGEKHSVLKLTVKKDNSIIYNGQISGNDHKPY 120
           PTI I GS        + ++ +L    N   +K  V+   + K+  +   GQIS ++  P
Sbjct: 64  PTIYIGGSGGNVTSIDWLVERLLPIKNISSQKSLVMTSNITKNYELKVEGQISQDNKYPI 123

Query: 121 IVIGFENNEDGYSNIKKQTKWLQIAMNDLQKKYKFKRFNAIGHSNGGLSWTIFLEDYYDS 180
           I          G ++ +  +K LQ  +  L + Y+     N +G+S+G       ++ D  ++
Sbjct: 124 IEFA---TVKGTNSGELFSKGLQKIIVYLTENYQVPWINLVGYSSGATGAVYYMMDTGNN 180

Query: 181 DEFD-MKSLLTMGTPFNFEES-----NTSN--------HTQMLKDLISNKGNIPSSLMVY 226
             F  +   +++   +N E +      + SN          T+M + +  N   + S   +
Sbjct: 181 PNFPPVNKYVSLDGEYNNETNLQLGESLSNVLKEGPIVKTEMYQYIADNYQKVSSKTQML 240

Query: 227 NLAGT--NSYDGDKIVPFASVETGKYIFQETAKHYTQLTVTGNNATHSDLPDNPEVIQYV 284
            L G   +    D  +P+A   + ++F+      T  T+      +HS  P NP V++YV
Sbjct: 241 LLEGNFNSEKQTDSAIPWADSFSIYHLFKKNGNEITT-TLYPTKTSHSQAPKNPTVVKYV 299
```

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

SEQ ID 8646 (GBS219) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 3; MW 31.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 7; MW 56kDa).

GBS219-GST was purified as shown in Figure 203, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 720

A DNA sequence (GBSx0764) was identified in *S.agalactiae* <SEQ ID 2213> which encodes the amino acid sequence <SEQ ID 2214>. This protein is predicted to be PTS system, cellobiose-specific IIC component. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.64   Transmembrane  263 - 279 ( 260 - 282)
   INTEGRAL    Likelihood = -6.26   Transmembrane  200 - 216 ( 197 - 226)
   INTEGRAL    Likelihood = -5.95   Transmembrane  157 - 173 ( 156 - 175)
   INTEGRAL    Likelihood = -5.79   Transmembrane  307 - 323 ( 306 - 332)
   INTEGRAL    Likelihood = -5.68   Transmembrane  131 - 147 ( 126 - 148)
   INTEGRAL    Likelihood = -4.73   Transmembrane  375 - 391 ( 370 - 396)
   INTEGRAL    Likelihood = -3.61   Transmembrane  101 - 117 (  98 - 119)
   INTEGRAL    Likelihood = -1.75   Transmembrane  326 - 342 ( 324 - 342)
   INTEGRAL    Likelihood = -0.37   Transmembrane   25 -  41 (  25 -  41)
   INTEGRAL    Likelihood = -0.16   Transmembrane   71 -  87 (  71 -  88)

----- Final Results -----
               bacterial membrane --- Certainty=0.4057(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 GENPEPT database:

```
>GP:AAC74807 GB:AE000268 PEP-dependent phosphotransferase enzyme II
          for cellobiose, arbutin, and salicin [Escherichia coli K12]
 Identities = 60/197 (30%), Positives = 83/197 (41%), Gaps = 12/197 (6%)

Query: 209 LAIFLTLSGLFVPDIL--FRPYSYFSVVSENLNAALSQHTDKIPYLYTFYTVKNSFAMFG 266
           LA+    +G+  P  L     Y  +  V   L A  + H    P L        +SF   G
Sbjct: 253 LALTALDNGIMTPWALENIATYQQYGSVEAALAAGKTFHIWAKPML-------DSFIFLG 305

Query: 267 GIGILLSLFLAVLYESRKLQSKNYYKLTLLTLTPLIFDQNLPFLVGLPVILQPILFIPMV 326
           G G  L L LA+   SR+    +Y ++  L L   IF  N P L GLP+I+ P++FIP V
Sbjct: 306 GSGATLGLILAIFIASRRA---DYRQVAKLALPSGIFQINEPILFGLPIIMNPVMFIPFV 362

Query: 327 LTTIFAEAFGALMLYLKFVDPAVYTVPSGTPSLLFGFLASNGDWRYLPVTAIILVVGFFI 386
           L      A      Y+  + P     P   P+ L  F  +NG    L V    L +   I
Sbjct: 363 LVQPILAAITLAAYYMGIIPPVTNIAPWTMPTGLGAFFNTNGSVAALLVALFNLGIATLI 422

Query: 387 YRPFVKIAFAKEEQYEK 403
           Y PFV +A   +   +K
Sbjct: 423 YLPFVVVANKAQNAIDK 439
```

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

Example 721

A DNA sequence (GBSx0765) was identified in *S.agalactiae* <SEQ ID 2217> which encodes the amino acid sequence <SEQ ID 2218>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1991(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 722

A DNA sequence (GBSx0766) was identified in *S.agalactiae* <SEQ ID 2219> which encodes the amino acid sequence <SEQ ID 2220>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.79   Transmembrane  188 - 204 ( 179 - 206)
   INTEGRAL    Likelihood = -5.36   Transmembrane  105 - 121 ( 104 - 127)
   INTEGRAL    Likelihood = -4.41   Transmembrane  212 - 228 ( 210 - 229)
   INTEGRAL    Likelihood = -3.45   Transmembrane   72 -  88 (  69 -  89)
   INTEGRAL    Likelihood = -0.48   Transmembrane  124 - 140 ( 124 - 140)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 8647> which encodes amino acid sequence <SEQ ID 8648> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 6
SRCFLG: 0
McG: Length of UR:   5
     Peak Value of UR:   2.99
     Net Charge of CR: 4
McG: Discrim Score:      6.88
GvH: Signal Score (-7.5): -2.86
     Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 5 value:  -5.79 threshold:  0.0
   INTEGRAL    Likelihood = -5.79   Transmembrane  179 - 195 ( 170 - 197)
   INTEGRAL    Likelihood = -5.36   Transmembrane   96 - 112 (  95 - 118)
   INTEGRAL    Likelihood = -4.41   Transmembrane  203 - 219 ( 201 - 220)
   INTEGRAL    Likelihood = -3.45   Transmembrane   63 -  79 (  60 -  80)
   PERIPHERAL  Likelihood =  0.10       18
 modified ALOM score:   1.66
icm1 HYPID: 7  CFP: 0.331

*** Reasoning Step: 3

----- 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>
```

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

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2221> which encodes the amino acid sequence <SEQ ID 2222>. Analysis of this protein sequence reveals the following:

```
      Possible site: 30

  >>> Seems to have a cleavable N-term signal seq.
     INTEGRAL    Likelihood =-11.20   Transmembrane  179 - 195 ( 173 - 201)
     INTEGRAL    Likelihood = -3.66   Transmembrane   96 - 112 (  95 - 113)
     INTEGRAL    Likelihood = -1.44   Transmembrane  203 - 219 ( 203 - 219)
     INTEGRAL    Likelihood = -0.96   Transmembrane  115 - 131 ( 115 - 131)
     INTEGRAL    Likelihood = -0.64   Transmembrane   63 -  79 (  63 -  79)

  ----- Final Results -----
               bacterial membrane --- Certainty=0.5479(Affirmative) < succ>

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

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

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 160/228 (70%), Positives = 185/228 (80%)

Query: 10  MSKKSHRQYQIYEGLRCAVALCFISGYINAFTYVTQGKRFAGVQTGNLLSFAIHLSNKHY 69
           MSKK  + YQ+YEGLRCA+ LCFISGY+NAFTY+TQGKRFAGVQTGNLLSFAI LS +
Sbjct: 1   MSKKKRKHYQVYEGLRCAMTLCFISGYVNAFTYMTQGKRFAGVQTGNLLSFAIRLSEQQL 60

Query: 70  SQALAFLLPIMVFMLGQSFTYFMNRWANKHQLHWYLLSSFALTQVAIVTIILTPFLPSSF 129
            +AL FLLP++VFMLGQSFTYFM+RWA K  LHWYLLSS  LT +A  T + TPFLPS+
Sbjct: 61  KEALQFLLPMIVFMLGQSFTYFMHRWATKKGLHWYLLSSVILTGIAFGTALFTPFLPSNV 120

Query: 130 TVAGLAFFASIQVDTFKSLRGAPYANMMMTGNIKNAAYLLTKGLYEKNSDIFLIARNTII 189
           TVA LAFFASIQVDTFK+LRGA YAN+MMTGNIKNAAYLLTKGLYEKN ++  I RNT+I
Sbjct: 121 TVAALAFFASIQVDTFKTLRGASYANVMMTGNIKNAAYLLTKGLYEKNHELTHIGRNTLI 180

Query: 190 IIGGFIFGVVCSTYFSSKLGEWSLSLILIPLLYVNLLLGHEFYNLQVE 237
           +I  F  GVVCST      GE++L  IL+PLLYVN LL  EFY++Q +
Sbjct: 181 VILAFAVGVVCSTLLCIAYGEYALMPILMPLLYVNYLLAQEFYHIQTK 228
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 723

A DNA sequence (GBSx0767) was identified in *S.agalactiae* <SEQ ID 2223> which encodes the amino acid sequence <SEQ ID 2224>. This protein is predicted to be tellurite resistance protein. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.00   Transmembrane  190 - 206 ( 190 - 206)

----- 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>
```

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

```
>GP:AAC22923 GB:U32807 tellurite resistance protein (tehB)
          [Haemophilus influenzae Rd]
 Identities = 164/282 (58%), Positives = 205/282 (72%), Gaps = 1/282 (0%)

Query: 7   LLPYKTMPVWTAQSIPKAFLEKHNTKEGTWAKLTILSGSLVFYQLSPDGEEISRHIFDAS 66
           L+ YK MPVWT  ++P+ F EKHNTK GTW KLT+L G L FY+L+ +G+ I+ HIF
Sbjct: 5   LICYKQMPVWTKDNLPQMFQEKHNTKVGTWGKLTVLKGKLKFYELTENGDVIAEHIFTPE 64

Query: 67  SDIPFVDPQVWHKVSPNSPDLSCYLTFYCQKEDYFHKKYGLTRTHSEVIASAPLLSEKSN 126
           S IPFV+PQ WH+V   S DL C L FYC+KEDYF KKY  T  H +V+ +A ++S
Sbjct: 65  SHIPFVEPQAWHRVEALSDDLECTLGFYCKKEDYFSKKYNTTAIHGDVVDAAKIISP-CK 123

Query: 127 ILDLGCGQGRNSLYLSLLGHQVTSVDSNGQSLVALENMALEEELPYNIKRYDINTAAIEG 186
           +LDLGCGQGRNSLYLSLLG+ VTS D N  S+  L     +E L  +   YDIN A I+
Sbjct: 124 VLDLGCGQGRNSLYLSLLGYDVTSWDHNENSIAFLNETKEKENLNISTALYDINAANIQE 183

Query: 187 HYDFILSTVVFMFLNPDCISDIILQMQSHTQIGGYNLIVSAMDTAENPCPLPFPFTFKEG 246
           +YDFI+STVVFMFLN + +  II  M+ HT +GGYNLIV+AM T + PCPLPF FTF E
Sbjct: 184 NYDFIVSTVVFMFLNRERVPSIIKNMKEHTNVGGYNLIVAAMSTDDVPCPLPFSFTFAEN 243

Query: 247 QLKSYYNDWEIIKYNENLGELHRVDENGNRLKLQFATLLARK 288
           +LK YY DWE ++YNEN+GELH+ DENGNR+K++FAT+LARK
Sbjct: 244 ELKEYYKDWEFLEYNENMGELHKTDENGNRIKMKFATMLARK 285
```

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

SEQ ID 2224 (GBS95) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 3; MW 35.6kDa) and in Figure 12 (lane 4; MW 35.6kDa). The GBS95-His fusion product was purified (Figure 191, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 292), which confirmed that the protein is immunoaccessible on GBS bacteria.

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

Example 724

A DNA sequence (GBSx0768) was identified in *S.agalactiae* <SEQ ID 2225> which encodes the amino acid sequence <SEQ ID 2226>. This protein is predicted to be methionyl-tRNA synthetase (metS). Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane  473 - 489 ( 473 - 489)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10043> which encodes amino acid sequence <SEQ ID 10044> was also identified.

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

```
>GP:CAB11814 GB:Z99104 methionyl-tRNA synthetase [Bacillus subtilis]
 Identities = 395/667 (59%), Positives = 501/667 (74%), Gaps = 12/667 (1%)

Query: 20  EKKSFYITTPIYYPSGKLHIGSAYTTIACDVLARYKRMMGFDVQYLTGLDEHGQKIQQKA 79
           E  +FYITTPIYYPSGKLHIG AYTT+A D +ARYKR+ GFDV+YLTG DEHGQKIQQKA
Sbjct: 4   ENNTFYITTPIYYPSGKLHIGHAYTTVAGDAMARYKRLKGFDVRYLTGTDEHGQKIQQKA 63
```

```
Query: 80  EEAGITPQEYVDGMAESVKTLWELLDISYDKFIRTTDTYHEEAVAKIFEQLLAQGDIYLG 139
           E+  ITPQEYVD  A  ++ LW+ L+IS D FIRTT+  H+  + K+F++LL  GDIYL
Sbjct: 64  EQENITPQEYVDRAAADIQKLWKQLEISNDDFIRTTEKRHKVVIEKVFQKLLDNGDIYLD 123

Query: 140 EYTGWYSVSDEEFFTESQLAEVYRDENGNMIGGVAP-SGHEVEKVSEESYFFRMSKYADR 198
           EY GWYS+ DE F+TE+QL ++ R+E G +IGG +P SGH VE + EESYFFRM KYADR
Sbjct: 124 EYEGWYSIPDETFYTETQLVDIERNEKGEVIGGKSPDSGHPVELIKEESYFFRMGKYADR 183

Query: 199 LKAYYAEHPEFIQPDGRMNEMLKNFIEPGLEDLAVSRTTYTWGVQVPSNPKHVIYVWIDA 258
           L  YY E+P FIQP+ R NEM+ NFI+PGLEDLAVSRTT+ WGV+VP NPKHV+YVWIDA
Sbjct: 184 LLKYYEENPTFIQPESRKNEMINNFIKPGLEDLAVSRTTFDWGVKVPENPKHVVYVWIDA 243

Query: 259 LMNYISALGYGWSDDLSQYHKFWPADIHMIGKDILRFHSIYWPIMLMALDLPLPKRLVAH 318
           L NY++ALGY   +D   Y K+WPAD+H++GK+I+RFH+IYWPIMLMALDLPLPK++ AH
Sbjct: 244 LFNYLTALGYDTEND-ELYQKYWPADVHLVGKEIVRFHTIYWPIMLMALDLPLPKQVFAH 302

Query: 319 GWFVMQDGKMSKSKGNVVYPEMLVERFGLDPLRYYLMRSLPVGSDGTFTPEDYVGRINYE 378
           GW +M+DGKMSKSKGNVV P  L+ER+GLD LRYYL+R +P GSDG FTPE +V RINY+
Sbjct: 303 GWLLMKDGKMSKSKGNVVDPVTLIERYGLDELRYYLLREVPFGSDGVFTPEGFVERINYD 362

Query: 379 LANDLGNLLNRTIAMVNKYFDGEVPRF-AVATDFDADLASVATDSIENYHKQMEAVDFPR 437
           LANDLGNLLNRT+AM+NKYFDG++  +    T+FD  L SVA ++++ Y K ME ++F
Sbjct: 363 LANDLGNLLNRTVAMINKYFDGQIGSYKGAVTEFDHTLTSVAEETVKAYEKAMENMEFSV 422

Query: 438 ALEAVWNLISRTNKYIDETAPWVLAKDETDRDKLAAVMSHLVASLRVVAHLIQPFMMETS 497
           AL  +W LISRTNKYIDETAPWVLAKD    ++L +VM HL  SLR+ A L+QPF+ +T
Sbjct: 423 ALSTLWQLISRTNKYIDETAPWVLAKDPAKEEELRSVMYHLAESLRISAVLLQPFLTKTP 482

Query: 498 DAIMEQLGL--GATFDLEKLT-FADLPEGVRVVAKGSPIFPRLDMEDEITYIKEQMNAGK 554
           + + EQLG+   +    + +T F  L +    V KG P+FPRL+ E+EI YIK +M  G
Sbjct: 483 EKMFEQLGITDESLKAWDSITAFGQLKD--TKVQKGEPLFPRLEAEEEIAYIKGKMQ-GS 539

Query: 555 APVEKEWVPEEVELTSSKGQIKFEDFDAVEIRVAEVIEVEKVEGSDKLLRFRLDAGDEGH 614
           AP ++E   EE +      +I  + F  VE+RVAEVIE E V+ +D+LL+ +LD G E
Sbjct: 540 APAKEETKEEEPQEVDRLPEITIDQFMDVELRVAEVIEAEPVKKADRLLKLQLDLGFE-K 598

Query: 615 RQILSGIAKFYPNEQELVGKKLQIVANLKPRKMMKKYVSQGMILSAEHDGKLTVLTVDSA 674
           RQ++SGIAK Y  E ELVGKKL  V NLKP K ++  +SQGMIL+ E DG L V+++D +
Sbjct: 599 RQVVSGIAKHYTPE-ELVGKKLVCVTNLKPVK-LRGELSQGMILAGEADGVLKVVSIDQS 656

Query: 675 VANGSII 681
           +  G+ I
Sbjct: 657 LPKGTRI 663
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2227> which encodes the amino acid sequence <SEQ ID 2228>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1245(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 516/665 (77%), Positives = 573/665 (85%), Gaps = 4/665 (0%)

Query: 21  KKSFYITTPIYYPSGKLHIGSAYTTIACDVLARYKRMMGFDVQYLTGLDEHGQKIQQKAE 80
           KK FYITTPIYYPSGKLHIGSAYTTIACDVLARYKR+MG +V YLTGLDEHGQKIQ KA+
Sbjct: 3   KKPFYITTPIYYPSGKLHIGSAYTTIACDVLARYKRLMGHEVFYLTGLDEHGQKIQTKAK 62

Query: 81  EAGITPQEYVDGMAESVKTLWELLDISYDKFIRTTDTYHEEAVAKIFEQLLAQGDIYLGE 140
           EAGITPQ YVD MA+ VK LW+LLDISYD FIRTTD YHEE VA +FE+LLAQ DIYLGE
Sbjct: 63  EAGITPQTYVDNMAKDVKALWQLLDISYDTFIRTTDDYHEEVVAAVFEKLLAQDDIYLGE 122
```

```
Query: 141 YTGWYSVSDEEFFTESQLAEVYRDENGNMIGGVAPSGHEVEKVSEESYFFRMSKYADRLK 200
           Y+GWYSVSDEEFFTESQL EV+RDE+G +IGG+APSGHEVE VSEESYF R+SKY DRL
Sbjct: 123 YSGWYSVSDEEFFTESQLKEVFRDEDGQVIGGIAPSGHEVEWVSEESYFLRLSKYDDRLV 182

Query: 201 AYYAEHPEFIQPDGRMNEMLKNFIEPGLEDLAVSRTTYTWGVQVPSNPKHVIYVWIDALM 260
           A++ E P+FIQPDGRMNEM+KNFIEPGLEDLAVSRTT+TWGV VPS+PKHV+YVWIDAL+
Sbjct: 183 AFFKERPDFIQPDGRMNEMVKNFIEPGLEDLAVSRTTFTWGVPVPSDPKHVVYVWIDALL 242

Query: 261 NYISALGYGWSDDLSQYHKFWPADI-HMIGKDILRFHSIYWPIMLMALDLPLPKRLVAHG 319
           NY +ALGY  ++  + + KFW   + HM+GKDILRFHSIYWPI+LM LDLP+P RL+AHG
Sbjct: 243 NYATALGYRQANH-ANFDKFWNGTVFHMVGKDILRFHSIYWPILLMMLDLPMPDRLIAHG 301

Query: 320 WFVMQDGKMSKSKGNVVYPEMLVERFGLDPLRYYLMRSLPVGSDGTFTPEDYVGRINYEL 379
           WFVM+DGKMSKSKGNVVYPEMLVERFGLDPLRYYLMRSLPVGSDGTFTPEDYVGRINYEL
Sbjct: 302 WFVMKDGKMSKSKGNVVYPEMLVERFGLDPLRYYLMRSLPVGSDGTFTPEDYVGRINYEL 361

Query: 380 ANDLGNLLNRTIAMVNKYFDGEVPRFA-VATDFDADLASVATDSIENYHKQMEAVDFPRA 438
           ANDLGNLLNRT+AM+NKYFDG VP +    T FDADL+ +    + +YHK MEAVD+PRA
Sbjct: 362 ANDLGNLLNRTVAMINKYFDGTVPAYVDNGTAFDADLSQLIDAQLADYHKHMEAVDYPRA 421

Query: 439 LEAVWNLISRTNKYIDETAPWVLAKDETDRDKLAAVMSHLVASLRVVAHLIQPFMMETSD 498
           LEAVW +I+RTNKYIDETAPWVLAK++ D+ +LA+VM+HL ASLR+VAH+IQPFMMETS
Sbjct: 422 LEAVWTIIARTNKYIDETAPWVLAKEDGDKAQLASVMAHLAASLRLVAHVIQPFMMETSA 481

Query: 499 AIMEQLGLGATFDLEKLTFADLPEGVRVVAKGSPIFPRLDMEDEITYIKEQMNAGKA-PV 557
           AIM QLGL    DL  L  AD P   +VVAKG+PIFPRLDME EI YIK QM    A
Sbjct: 482 AIMAQLGLEPVSDLSTLALADFPANTKVVAKGTPIFPRLDMEAEIDYIKAQMGDSSAISQ 541

Query: 558 EKEWVPEEVELTSSKGQIKFEDFDAVEIRVAEVIEVEKVEGSDKLLRFRLDAGDEGHRQI 617
           EKEWVPEEV L S K  I FE FDAVEIRVAEV EV KVEGS+KLLRFR+DAGD   RQI
Sbjct: 542 EKEWVPEEVALKSEKDVITFETFDAVEIRVAEVKEVSKVEGSEKLLRFRVDAGDGQDRQI 601

Query: 618 LSGIAKFYPNEQELVGKKLQIVANLKPRKMMKKYVSQGMILSAEHDGKLTVLTVDSAVAN 677
           LSGIAKFYPNEQELVGKKLQIVANLKPRKMMKKY+SQGMILSAEH  +LTVLTVDS+V N
Sbjct: 602 LSGIAKFYPNEQELVGKKLQIVANLKPRKMMKKYISQGMILSAEHGDQLTVLTVDSSVPN 661

Query: 678 GSIIG 682
           GSIIG
Sbjct: 662 GSIIG 666
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 725

A DNA sequence (GBSx0769) was identified in *S.agalactiae* <SEQ ID 2229> which encodes the amino acid sequence <SEQ ID 2230>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2633(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 726

A DNA sequence (GBSx0770) was identified in *S.agalactiae* <SEQ ID 2231> which encodes the amino acid sequence <SEQ ID 2232>. This protein is predicted to be branched chain amino acid transport system II carrier protein (brnQ). Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-14.91   Transmembrane  279 - 295 ( 269 - 303)
   INTEGRAL    Likelihood = -9.98   Transmembrane   82 -  98 (  74 - 102)
   INTEGRAL    Likelihood = -6.58   Transmembrane  345 - 361 ( 340 - 364)
   INTEGRAL    Likelihood = -6.00   Transmembrane  157 - 173 ( 153 - 179)
   INTEGRAL    Likelihood = -4.30   Transmembrane   48 -  64 (  45 -  66)
   INTEGRAL    Likelihood = -4.14   Transmembrane  251 - 267 ( 250 - 278)
   INTEGRAL    Likelihood = -4.09   Transmembrane  308 - 324 ( 305 - 326)
   INTEGRAL    Likelihood = -2.55   Transmembrane  218 - 234 ( 216 - 237)
   INTEGRAL    Likelihood = -1.38   Transmembrane  126 - 142 ( 126 - 142)

----- Final Results -----
              bacterial membrane --- Certainty=0.6965(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9407> which encodes amino acid sequence <SEQ ID 9408> was also identified.

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

```
>GP:AAC00400 GB:AF008220 branch-chain amino acid transporter
           [Bacillus subtilis]
 Identities = 130/367 (35%), Positives = 204/367 (55%), Gaps = 12/367 (3%)

Query: 1   MSEKFSPWFSLTFLVILYLTIGPLFAIPRTATVSFEIGVAPIVGHSP--IALLCFTACFF 58
           +++K  P F   F V+LYL+IGPLFAIPRT TVS+EIG  P +   P  ++LL FT  FF
Sbjct: 73  LADKAHPVFGTIFTVVLYLSIGPLFAIPRTGTVSYEIGAVPFLTGVPERLSLLIFTLIFF 132

Query: 59  AAAYYLAIRPNGILDSVGKILTPVFAFLILSLVVVGAIAYGNLESAKASADYAGKAFGSG 118
              YYLA+ P+ ++D VGKILTP+  F I+ ++V+ AI          +  Y G     G
Sbjct: 133 GVTYYLALNPSKVVDRVGKILTPI-KFTIILIIVLKAIFTPMGGLGAVTEAYKGTPVFKG 191

Query: 119 VLAGYNTLDALAAVAFCLVATETLKKFGFKTKKEYLSTIWIVGIVTSLAFSILYIGLGFL 178
            L GY T+DALA++ F +V    +K  G    K    +     G++ +L  + +Y+ L +L
Sbjct: 192 FLEGYKTMDALASIVFGVVVVNAVKSKGVTQSKALAAACIKAGVIAALGLTFIYVSLAYL 251

Query: 179 GNKFPVPADILADPNVNKGAYVLSQASYKLFGNFGRYFLSIMVTLTCFTTTVGLIVSVSE 238
           G          A   V +GA +LS +S+ LFG+ G   L   +T+ C TT++GL+ S  +
Sbjct: 252 G-----ATSTNAIGPVGEGAKILSASSHYLFGSLGNIVLGAAITVACLTTSIGLVTSCGQ 306

Query: 239 FFDKNFRFGNYKLFATVFTLIGFLIANLGLNAVITFSVPVLTLLYPIVIVIVLIILINKW 298
           +F K     +YK+  T+ TL   +IAN GL  +I FSVP+L+ +YP+ IVI+++  I+K
Sbjct: 307 YFSKLIPALSYKIVVTIVTLFSLIIANFGLAQIIAFSVPILSAIYPLAIVIIVLSFIDKI 366

Query: 299 LPLSKK---GMSLTIGLVTLVSFVEVLAGQWQEKTLTQLVGFLPFHTISMGWLVPMLIGI 355
               ++      +  GL +++  ++  AG          L   LP +++ +GW++P ++G
Sbjct: 367 FKERREVYIACLIGTGLFSILDGIKA-AGFSLGSSLDVFLNANLPLYSLGIGWVLPGIVGA 425

Query: 356 VFSLVLS 362
           V   VL+
Sbjct: 426 VIGYVLT 432
```

There is also homology to SEQ ID 2234.

A related GBS gene <SEQ ID 8649> and protein <SEQ ID 8650> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length of UR:   30
     Peak Value of UR:   2.99
     Net Charge of CR: 2
McG: Discrim Score:     13.17
GvH: Signal Score (-7.5): -3.3
     Possible site: 33
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 11 value: -14.91 threshold:  0.0
   INTEGRAL    Likelihood =-14.91   Transmembrane  347 - 363 ( 337 - 371)
   INTEGRAL    Likelihood = -9.98   Transmembrane  150 - 166 ( 142 - 170)
   INTEGRAL    Likelihood = -7.54   Transmembrane   40 -  56 (  36 -  61)
   INTEGRAL    Likelihood = -6.64   Transmembrane   79 -  95 (  76 -  97)
   INTEGRAL    Likelihood = -6.00   Transmembrane  225 - 241 ( 221 - 247)
   INTEGRAL    Likelihood = -4.30   Transmembrane  116 - 132 ( 113 - 134)
   INTEGRAL    Likelihood = -4.14   Transmembrane  319 - 335 ( 318 - 346)
   INTEGRAL    Likelihood = -4.09   Transmembrane  376 - 392 ( 373 - 394)
   INTEGRAL    Likelihood = -2.92   Transmembrane    7 -  23 (   6 -  28)
   INTEGRAL    Likelihood = -2.55   Transmembrane  286 - 302 ( 284 - 305)
   INTEGRAL    Likelihood = -1.38   Transmembrane  194 - 210 ( 194 - 210)
   PERIPHERAL  Likelihood =  2.49     402
 modified ALOM score:   3.48
icml HYPID: 7  CFP: 0.696

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6965(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:

```
ORF00247(304 - 1596 of 1941)
OMNI|NT01BS3447(19 - 446 of 459) branched chain amino acid transport system II carrier
protein
%Match = 21.7
%Identity = 38.8  %Similarity = 61.2
Matches = 166  Mismatches = 157  Conservative Sub.s = 96

93        123        153        183        213        243        273        303
VLTVDSAVANGSIIG*SKRALCSFFVFKKKVTE*LENYENDLEFIFIFDIIKDIDSKHLDRI**GEFMERV*IDYLH*WL

                                                                    LTEYFNIIIRRIFFMKHS
                                                                              10

333        363        393        423        453        483        513        543
LMVKKGFLTGLLLFGIFFGAGNLIFPPALGVASGQDFWPAILGFCLSGVGLAIITLLLGTLTNGGYKTEMSEKFSPWFSL
| ||   : |::|| :||||||:|:|| || |:|:: | || || |:|||| :: ::   ||    |  :::|  | |
LPVKDTIIIGFMLFALFFGAGNMIYPPELGQAAGHNVWKAIGGFLLTGVGLPLLGIIAIALTGKDAKG-LADKAHPVFGT
          30         40         50         60         70         80         90

573        603        633        657        687        717        747        777
TFLVILYLTIGPLFAIPRTATVSFEIGVAPIVGHSP--IALLCFTACFFAAAYYLAIRPNGILDSVGKILTPVFAFLILS
 | |:|||:|||||||||| |||:|||  | :   |  ::|| ||  ||   ||||: |: ::| |||||||:  | |:
IFTVVLYLSIGPLFAIPRTGTVSYEIGAVPFLTGVPERLSLLIFTLIFFGVTYYLALNPSKVVDRVGKILTPI-KFTIIL
          110        120        130        140        150        160        170

801        831        861        891        921        951        981       1011
LVVVGAI--AYGNLESAKASADYAGKAFGSGVLAGYNTLDALAAVAFCLVATETLKKFGFKTKKEYLSTIWIVGIVTSLA
::|: ||    | | :   :  | |     | | || |:||||:: | :|    :|  |    |   :     |:: :|
IIVLKAIFTPMGGLGA--VTEAYKGTPVFKGFLEGYKTMDALASIVFGVVVVNAVKSKGVTQSKALAAACIKAGVIAALG
           190        200        210        220        230        240        250

1041       1071       1101       1131       1161       1191       1221       1251
FSILYIGLGFLGNKFPVPADILADPNVNKGAYVLSQASYKLFGNFGRYFLSIMVTLTCFTTTVGLIVSVSEFFDKNFRFG
```

```
:: :|: | :||          |   | :|| :|| :|: |||::|   |   :|: |:||::||: |  ::| |
LTFIYVSLAYLG-----ATSTNAIGPVGEGAKILSASSHYLFGSLGNIVLGAAITVACLTTSIGLVTSCGQYFSKLIPAL
                  270       280       290       300       310       320

1281      1311      1341      1371      1401      1431      1461      1488
NYKLFATVFTLIGFLIANLGLNAVITFSVPVLTLLYPIVIVIVLIILINKWLPLSKKGMSLTIGLVTLVSFVEVLAG-QW
:||:  |: ||  ::|||:||  :| ||||:|: :||: |||::: :|:|   : |:   : |  :    :  :| | :
SYKIVVTIVTLFSLIIANFGLAQIIAFSVPILSAIYPLAIVIIVLSFIDK---IFKERREVYIACLIGTGLFSILDGIKA
       340       350       360       370           380       390       400

1518      1536      1566      1596      1626      1656      1686      1716
QEKTLTQLVGFL----PFHTISMGWLVPMLIGIVFSLVLSDKQKGQAFDLEKFEG*HYFNFIDMSKRLKLRF*PFLYQIF
   :|  |  ||    |:::: :||::| ::| |   ||:
AGFSLGSLDVFLNANLPLYSLGIGWVLPGIVGAVIGYVLTLFIGPSKQLNEIS
            420       430       440       450
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 727

A DNA sequence (GBSx0771) was identified in *S.agalactiae* <SEQ ID 2235> which encodes the amino acid sequence <SEQ ID 2236>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3291(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10041> which encodes amino acid sequence <SEQ ID 10042> was also identified.

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

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 728

A DNA sequence (GBSx0772) was identified in *S.agalactiae* <SEQ ID 2237> which encodes the amino acid sequence <SEQ ID 2238>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.33   Transmembrane  117 - 133 ( 112 - 136)
   INTEGRAL    Likelihood = -3.77   Transmembrane   53 -  69 (  53 -  70)
   INTEGRAL    Likelihood = -3.40   Transmembrane   98 - 114 (  97 - 115)

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

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

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

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

Example 729

A DNA sequence (GBSx0773) was identified in *S.agalactiae* <SEQ ID 2239> which encodes the amino acid sequence <SEQ ID 2240>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.19   Transmembrane   22 -  38 (  20 -  44)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 8651> which encodes amino acid sequence <SEQ ID 8652> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length of UR:   21
     Peak Value of UR:   3.11
     Net Charge of CR: 2
McG: Discrim Score:     11.30
GvH: Signal Score (-7.5): -5.35
     Possible site: 28
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 1 value:  -4.19 threshold:  0.0
   INTEGRAL    Likelihood = -4.19   Transmembrane    5 -  21 (   3 -  27)
   PERIPHERAL  Likelihood =  6.74       53
 modified ALOM score:   1.34
icm1 HYPID: 7  CFP: 0.268

*** 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 GENPEPT database:

```
>GP:CAB15623 GB:Z99122 spore coat protein (inner) [Bacillus subtilis]
 Identities = 71/359 (19%), Positives = 148/359 (40%), Gaps = 49/359 (13%)

Query: 127 ISYRGNTSRYFDKKSLKVKFVTNKLKEKKHRLAGMPKESEWVLHGPFLDRTLLRNYLSYN 186
           I+YRG+  R F KKS  + F   K       +          L+  + D +L+RN LS +
Sbjct: 47  IAYRGSHIRDFKKKSYHISFYQPKTFRGAREIH---------LNAEYKDPSLMRNKLSLD 97

Query: 187 IAGEIMSYAPNVRYCELFVNGEYQGVYLAVENIEQGEQRVPIEKSDKKLHKTPYIVAWDR 246
              E+ + +P   +  + +NG+ +GVYL +E++++          + +KL       A D
Sbjct: 98  FFSELGTLSPKAEFAFVKMNGKNEGVYLELESVDE------YYLAKRKLADGAIFYAVDD 151

Query: 247 EHKAKQKLDNYVHYTHQSGISALDVKYPGKQRLTSKQLEFINKD----INHIEKVLYSYD 302
           +       D       +   ++L++ Y  +++   +++ +F  +D    IN + K  +
Sbjct: 152 DANFSLMSD-----LERETKTSLELGY--EKKTGTEEDDFYLQDMIFKINTVPKAQFK-- 202

Query: 303 FSQYPKYIDRESFANYFVINEFFRNVDAGKFSTYLYKDLRDRA-KLVVWDFNNAFDNQIE 361
            S+  K++D + +  +      F  N D     +  LY+       +++WD++    I
Sbjct: 203 -SEVTKHVDVDKYLRWLAGIVFTSNYDGFVHNYALYRSGETGLFEVIPWDYDATWGRDIH 261

Query: 362 GRVDEADFTLTDAPWFNMLIKDKAFIDLVVHRYKELRKGVLATEYLSNYIDETRHFLGPA 421
```

```
           G    AD+       FN L             YK L +  L + +   Y++       P
Sbjct: 262 GERMAADYVRIQG--FNTLTARILDESEFRKSYKRLLEKTLQSLFTIEYME-------PK 312

Query: 422 IDRNYKKWGYVFDLKNTDPRNYLIPTERN-VTSYHKSVEQLKDFIKKRGRWMDRNIETL 479
           I   Y++           P   + P ++N +  + +  + + ++IK R +++  ++  L
Sbjct: 313 IMAMYER---------IRPFVLMDPYKKNDIERFDREPDVICEYIKNRSQYLKDHLSIL 362
```

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 730

A DNA sequence (GBSx0774) was identified in *S.agalactiae* <SEQ ID 2241> which encodes the amino acid sequence <SEQ ID 2242>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> 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>
```

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

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

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

Example 731

A DNA sequence (GBSx0775) was identified in *S.agalactiae* <SEQ ID 2243> which encodes the amino acid sequence <SEQ ID 2244>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -4.62   Transmembrane    5 -  21 (   3 -  24)

----- 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>
```

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

```
>GP:BAB05949 GB:AP001514 unknown [Bacillus halodurans]
 Identities = 199/697 (28%), Positives = 322/697 (45%), Gaps = 58/697 (8%)

Query: 57   KPFVVKGVDVESSLAGYHHNDFPITQKTYREWFHLISNMGANTVRVKVPMNVAFYDALYH 116
           K   + GV++     G    +  I +K Y  WF  I  MG N +RV       FY AL
Sbjct: 414  KKLQIHGVNLGMGKPGTFPGEAAIKEKDYYRWFEQIGEMGGNAIRVYTLHPPGFYHALKR 473

Query: 117  HNKASKRPLYLLQGIRIDSYRNNASITAFNDNYRGYLKREAKGVVDILHGRKQVWNTDLG 176
           +N+  + P+YL  G+ ID      ++ AF++      ++E K +VD++HG   V + + G
Sbjct: 474  YNEQHENPIYLFHGVWIDEEPLEDTLDAFDEETNEEFQQEMKRIVDVIHGNAVV-DPNPG 532

Query: 177  SRH--YHYDLSPWVLGYVVGDDWNSGTVAYTNHQEKKT-QYKGRYFKTSVAANPFEVMLA 233
             H  Y  D+SP+ +G+++G +W   TV  TN         Y G+Y +T   A PFE  LA
Sbjct: 533  HAHGVYQADVSPYTIGWIIGIEWYPHTVKATNKNNPDIGDYDGKYVETK-DAEPFEYWLA 591
```

```
Query: 234  QVMDELTHYETAKYGWQHLISFSNSPTTDPF-HYRKPFEAQAPKYVQLNVENIQANSNVK 292
              D L  YE  +Y W   +SF+N  TTD   H  +P E +    V  NV +++  +  +
Sbjct: 592  NQFDILLSYEIEQYNWIRPVSFTNWVTTDLLTHPAEPNEDEDLVGVDPNVIHLKGPA-TE 650

Query: 293  AGMFAAYKAIDFHPRYKDYLLFDKENISKEDRQKIKELSLSQGYVKLLNAYHKIPVLVTG 352
               FA+Y     +P Y D+L ++++ I   D +   EL+   GY+K L+  H +P+L+
Sbjct: 651  TNQFASYHV---YPYYPDFLNYEEDYIHYVDHR--GELNNYAGYLKDLHDAHDLPILIAE 705

Query: 353  YGYSTARGIA-QKEIDKRPLPINEKEQGQRLLEDYESFISSGSFGATINAWQDDWNARAW 411
             +G   +RG+  +    K    ++E+EQG+ ++E +E  I    G  I  WQD+W  R W
Sbjct: 706  FGVPASRGLTHENPFGKNQGFLSEEEQGKIVVELFEDIIEEKLLGGLIFTWQDEWFKRTW 765

Query: 412  NTSFATNKHSQFLWGDAQVFNQGYGLLGFKNAKHHYQVDGKRGKG-----EWKHPLMTSA 466
             NT    N   +  W +AQ   Q +GLL F   K     D +  +      E  HP +
Sbjct: 766  NTMDYDNPDRRPFWSNAQTNEQQFGLLSFDRLKVKVNGDDQDWEDASLLYEEDHPYVKR- 824

Query: 467  TGDDLYASSDESYLYLAIKTKPEKLKE-----KRLLPIDITPKSGSRKMNGSK-VTFSKS 520
              LY   DE YLY  I  K     +       +L +D  P  G+  +  + VTF
Sbjct: 825  ----LYMDHDERYLYFRIDMKSGSTDDFFKDGFPILVLDTLPGQGNEHIKEVEGVTFDHG 880

Query: 521  SDFVLSIDPNGKSELFVQERYNALKANYLRQLNGKDFYAFPPKKNSSNFEQINMVLRNTK 580
              DF++ +    +S + V   Y+     Y +     +  +  P+ N+  F++I+  L N +
Sbjct: 881  IDFIIELKGYDESRVKVDAYYDFFTYQYSQIYQMIEETSIEPQNNTGVFQKIHYAL-NQE 939

Query: 581  IVEDMEKVKATERFLP--THPTGLLKTGTTDRHQKTFDSQTD--ISFGKDFIEVRIPWQL 636
             I     ++ +T   +P   + TG L+ G  D     +DS  D  ++  K  IEVRIPW L
Sbjct: 940  I-----RIPSTNEVIPFSYYETGELRHGNGDPEADDYDSLADFFVNEEKGMIEVRIPWLL 994

Query: 637  LNFSDPSSQKIHDDYFKHYGVKELE-IESI-ALGLGANSKENTLIKMAD----------- 683
             L+F DPS +++    ++  G +  E IE + A  L    K++   ++ D
Sbjct: 995  LSFKDPSQREVMSAIYEGEGGETSEIIEGVRAAVLFVEPKDDDSYQVVDALPALDGDRLT 1054

Query: 684  ------YRLKNWERPDTKTFLKDSYYSIKKEWSKERE 714
                  Y  + W+ P  +  LK SY  +K+ ++  +E
Sbjct: 1055 DEVMNMYTWETWDIPLYEERLKQSYDLVKEAFTSIKE 1091
```

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

A related GBS gene <SEQ ID 8653> and protein <SEQ ID 8654> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 2
McG: Discrim Score:     12.00
GvH: Signal Score (-7.5): -5.46
     Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -4.62 threshold:  0.0
   INTEGRAL    Likelihood = -4.62   Transmembrane    5 -  21 (   3 -  24)
   PERIPHERAL  Likelihood =  7.32     223
 modified ALOM score:   1.42

*** Reasoning Step: 3

----- 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>
```

SEQ ID 2244 (GBS62) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 7; MW 80.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 4; MW 105kDa).

The GBS62-GST fusion product was purified (Figure 100A; see also Figure 193, lane 7) and used to immunise mice (lane 1 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure

100B), FACS (Figure 100C), 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 was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 732

A DNA sequence (GBSx0778) was identified in *S.agalactiae* <SEQ ID 2245> which encodes the amino acid sequence <SEQ ID 2246> in others. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.48   Transmembrane  310 - 326 ( 302 - 335)
   INTEGRAL    Likelihood = -7.32   Transmembrane  362 - 378 ( 361 - 380)
   INTEGRAL    Likelihood = -7.11   Transmembrane  334 - 350 ( 329 - 355)
   INTEGRAL    Likelihood = -2.28   Transmembrane  381 - 397 ( 380 - 397)

----- Final Results -----
               bacterial membrane --- Certainty=0.3994(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10039> which encodes amino acid sequence <SEQ ID 10040> was also identified.

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

```
>GP:BAB05950 GB:AP001514 unknown conserved protein in others
           [Bacillus halodurans]
 Identities = 143/405 (35%), Positives = 226/405 (55%), Gaps = 5/405 (1%)

Query: 11  IVPAYNESTTIVSSIDSLLHLDYEAYEIIVVDDGSSDNTSDVLKEEFALMKISNTIDSII 70
           +VPAYNE T I+ ++ SLL L Y   EI+VV+DGS+D T +V+ E F ++K+   I   I
Sbjct: 69  LVPAYNEETGIIETVRSLLSLKYPQTEIVVVNDGSTDQTLEVIIEHFQMVKVGKVIRKQI 128

Query: 71  ATQTCKDVFQRQVGKVKLTLIVKENGGKGDALNMGINAANYDYFLCLDADSMLQVDSLSQ 130
            T+  K V+Q  +    L L+ K NGGK DALN G+N + Y YF  +D DS+L+ D+L +
Sbjct: 129 ETEPIKGVYQSTIFP-HLLLVDKSNGGKADALNAGLNVSKYPYFCSIDGDSILETDALLK 187

Query: 131 ISKSIQV----DPTVIAVGGLVQVAQGVKIEQGKVASYRLPWRIIPCAQALEYDSSFLGA 186
           + K I      +  VIA GG V++A G  I+ G V S +L    +   Q +EY  +FL
Sbjct: 188 VMKPIVTSRDDEDEVIASGGNVRIANGSDIQMGSVLSVQLAKNPLVVMQVIEYLRAFLMG 247

Query: 187 RIFLDYLRANLIISGAFGLFKKDLVKAVGGYDTQTLGEDMELVMKLHFFCRNNNIPYRIC 246
           RI L      LIISGAF +F K  V   GGY  +T+GEDMELV++LH   +   + RI
Sbjct: 248 RIGLSRHNMVLIISGAFSVFAKKWVMEAGGYSKKTVGEDMELVVRLHRLVKEKRLKKRIT 307

Query: 247 YETDAVCWSQAPTNLGDLRKQRRRWYLGLYQCLKKYKSIFANYRFGAVGSISYIYYILFE 306
           +  D VCW++AP     L++QR RW+ GL + L  ++ +  N ++G VG+ S  Y+ + E
Sbjct: 308 FVPDPVCWTEAPATFRVLQRQRSRWHRGLMESLWLHRGMTFNPKYGLVGTASIPYFWIVE 367

Query: 307 LLTPFIECFGIVIIFLSLLFNQLNIPFFISLVSLYIFYCVLITLSSFLHRIYSQQLVIGI 366
              P +E  G + I  +  F  L + F ++L  L++ Y  + ++++ +    +S +    +
Sbjct: 368 FFGPVVELMGYLYIVFAFFFGGLYVEFALALFLLFVLYGTVFSMTAVILEGWSLKRYPKV 427

Query: 367 LDIVKVFYIAVFRYLILHPVLTFVKVASVIGYKNKKMVWGHITRE 411
            D+ ++   ++F  L   P+    + ++I    +  WG +TR+
Sbjct: 428 SDMSRLMIFSLFEALWYRPLTVLWRFGAIIEALFRSKAWGEMTRK 472
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2247> which encodes the amino acid sequence <SEQ ID 2248>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.04   Transmembrane   33 -  49 (  24 -  57)
   INTEGRAL    Likelihood =-10.77   Transmembrane  376 - 392 ( 370 - 399)
   INTEGRAL    Likelihood = -7.86   Transmembrane  344 - 360 ( 342 - 372)
   INTEGRAL    Likelihood = -4.94   Transmembrane   63 -  79 (  55 -  81)
   INTEGRAL    Likelihood = -2.07   Transmembrane  403 - 419 ( 403 - 419)

----- 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>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 84/397 (21%), Positives = 173/397 (43%), Gaps = 71/397 (17%)

Query: 6   FRRKSIVPAYNEST-TIVSSIDSLLHLDYEAYEIIVVDDGSSDNTSDVLKEEFALMKISN 64
           ++  +++P+YNE   +++ ++ S+L   Y   EI +VDDGSS+  +  L EE+    ++
Sbjct: 90  YKVAAVIPSYNEDAESLLETLKSVLAQTYPLSEIYIVDDGSSNTDAIQLIEEY----VNR 145

Query: 65  TIDSIIATQTCKDVFQRQVGKVKLTLIVKENGGKGDALNMGINAANYDYFLCLDADSMLQ 124
            +D       C++V       V  +L+   N GK  A        ++ D FL +D+D+ +
Sbjct: 146 EVD------ICRNVI------VHRSLV---NKGKRHAQAWAFERSDADVFLTVDSDTYIY 190

Query: 125 VDSLSQISKSIQVDPTVIAVGGLVQVAQGVKIEQGKVASYRLPWRIIPCAQALEYDSSFL 184
            ++L ++ KS   D TV A                G + +       ++      + YD++F
Sbjct: 191 PNALEELLKSFN-DETVYAA-------------TGHLNARNRQTNLLTRLTDIRYDNAF- 235

Query: 185 GARIFLDYLRANLII-SGAFGLFKKD-LVKAVGGYDTQT-------LGEDMELVMKLHFF 235
           G       L  N+++ SG   +++++ ++  +  Y  QT       +G+D  L
Sbjct: 236 GVERAAQSLTGNILVCSGPLSIYRREVIIPNLERYKNQTFLGLPVSIGDDRCLT------ 289

Query: 236 CRNNNIPY-RICYETDAVCWSQAPTNLGDLRKQRRRWYLGLY-QCLKKYKSIFANYRFGA 293
             N  I   R  Y++ A C +  P  L    KQ+ RW    + + +   K I +N
Sbjct: 290 --NYAIDLGRTVYQSTARCDTDVPFQLKSYLKQQNRWNKSFFKESIISVKKILSN----P 343

Query: 294 VGSISYIYYYILFELLTPFIECFGIVIIFLSLLFNQLNIPFFISLVSLYIFYCV--LITLS 351
           + ++  I+ ++  ++         +++    +LLFNQ       + L+ L+ F  +  ++ L
Sbjct: 344 IVALWTIFEVVMFMM--------LIVAIGNLLFNQ---AIQLDLIKLFAFLSIIFIVALC 392

Query: 352 SFLHRIYSQQLVIGILDIVKVFYIAVFRYLILHPVLT 388
             +H +         +  +  + ++ V + L L+ + T
Sbjct: 393 RNVHYMIKHPASFLLSPLYGILHLFVLQPLKLYSLCT 429
```

A related GBS gene <SEQ ID 8655> and protein <SEQ ID 8656> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     -5.18
GvH: Signal Score (-7.5): -4.91
     Possible site: 14
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 4 value:  -7.48 threshold:  0.0
   INTEGRAL    Likelihood = -7.48   Transmembrane  310 - 326 ( 302 - 335)
   INTEGRAL    Likelihood = -7.32   Transmembrane  362 - 378 ( 361 - 380)
   INTEGRAL    Likelihood = -7.11   Transmembrane  334 - 350 ( 329 - 355)
   INTEGRAL    Likelihood = -2.28   Transmembrane  381 - 397 ( 380 - 397)
   PERIPHERAL  Likelihood =  1.22     140
 modified ALOM score:   2.00

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3994(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:

```
ORF00238(331 - 1401 of 1866)
GP|5813901|gb|AAD52055.1|AF086783_3|AF086783(52 - 367 of 412) IcaA (Staphylococcus aureus)
%Match = 10.3
%Identity = 34.8  %Similarity = 55.9
Matches = 109  Mismatches = 128  Conservative Sub.s = 66

150       180       210       240       270       300       330       360
VAMRRSSKLNLGVRPPFACLR**AVFNTANISSKVVR*TPTRRLNRTSVNCLLAS*FIELLYHILFRRKSIVPAYNESTT
                                                                    ::  |||| |
                      MQFFNFLLFYPVFMSIYWIVGSIYFYFTREIRYSLNKKPDINVDELEGITFLLACYNESET
                               10        20        30        40        50        60

390       420       450       471       501       531       561       591
IVSSIDSLLHLDYEAYEIIVVDDGSSDNTSDVL---KEEFALMKISNTIDSIIATQTCKDVFQRQVGKVKLTLIVKENGG
|  :: ::| | ||  |||:::|||||||:::   ||   :: :                              ::|| |
IEDTLSNVLALKYEKKEIIIINDGSSDNTAELIYKIKENNDFIFVD--------------------------LQENRG
                 80        90       100                                   110

621       651       681       711       741       771       801       831
KGDALNMGINAANYDYFLCLDADSMLQVDSLSQISKSIQVDPTVIAVGGLVQVAQGVKIEQGKVASYRLPWRIIPCAQAL
| :||| ||  |:||| :|||||:::  |:   : :: : || : || |  ::     |              :   | :
KANALNQGIKQASYDYVMCLDADTIVDQDAPYYMIENFKHDPKLGAVTGNPRIRNKSSI-------------LGKIQTI
              130       140       150       160       170

861       891       918       948       978      1008      1038      1068
EYDSSFLGARIFLDYLRANL-IISGAFGLFKKDLVKAVGGYDTQTLGEDMELVMKLHFFCRNNNIPYRICYETDAVCWSQ
|| :|::|       |   :  ||| | ||||  |  || :||  : ||: :  |||          ||| ||  |:||
EY-ASLIGCIKRSQTLAGAVNTISGVFTLFKKSAVVDVGYWDTDMITEDIAVSWKLH------LRGYRIKYEPLAMCWML
         190       200       210       220       230              240       250

1098      1128      1155                1194      1224      1254      1284
APTNLGDLRKQRRRWYLGLYQCL-KKYKSIFANYRFG-------AVGSISYIYYILFELLTPFIECFGIVIIFLSLLFNQ
 |  || | ||| ||  | :: | : : |     ||         : || ::| :|: |   ||        ||   |
VPETLGGLWKQRVRWAQGGHEVLLRDFFSTMKTKRFPLYILMFEQIISILWVYIVLLYLGYLFI-----TANFLDYTFMT
               270       280       290       300       310             320

1311      1341      1371      1401      1431      1461      1491      1521
LNIP-FFISLVSLYIFYCVLITLSSFLHRIYSQQLVIGILDIVKVFYIAVFRYLILHPVLTFVKVASVIGYKNKKMVWGH
 :   |::|  ::     :  |:: |:   | :: : |::
YSFSIFLLSSFTMTFINVIQFTVALFIDSRYEKKNMAGLIFVSWYPTVYWIINAAVVLVAFPKALKRKRGGYATWSSPDR
         340       350       360       370       380       390       400
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 733

A DNA sequence (GBSx0779) was identified in *S.agalactiae* <SEQ ID 2249> which encodes the amino acid sequence <SEQ ID 2250>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2014(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 GENPEPT database:

```
>GP:CAA22725 GB:AL035161 hypothetical protein SC9C7.13c
          [Streptomyces coelicolor A3(2)]
 Identities = 35/153 (22%), Positives = 64/153 (40%), Gaps = 5/153 (3%)
```

```
Query: 5   IRRARLGDEVNLAYIQTESWKAAFGKILPEDIIQKTTEIEPAITMYQQLLHKEVGKGYIL 64
           +R   L D   ++ I+   W++A+  ++P+  +        A            G+  ++
Sbjct: 10  VREMTLADCDRVSLIRVRGWQSAYRGLMPQPYLDAMDPAADAERRRSLFARPPEGRVNLV 69

Query: 65  EVDSNPHCMAWWD----KSREDGMLDYAELICIHSLKEGWGKGYGSQMMNHVLSEIQQAG 120
             D     + W      +  E    D AEL  ++      +G G G  +    +  + AG
Sbjct: 70  AEDEGGEVVGWACHGPYRDGEARTAD-AELYALYVDAARFGAGIGRALAGESVRRCRAAG 128

Query: 121 YNKVILWVFTENTRARKFYDRFGFSFKGKSKTY 153
           + +++LWV   N RAR+FYDR GF   G  + +
Sbjct: 129 HARMLLWVLKGNVRARRFYDRAGFRPDGAEEPF 161
```

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

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

Example 734

A DNA sequence (GBSx0780) was identified in *S.agalactiae* <SEQ ID 2251> which encodes the amino acid sequence <SEQ ID 2252>. This protein is predicted to be a DNA-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1162(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 735

A DNA sequence (GBSx0781) was identified in *S.agalactiae* <SEQ ID 2253> which encodes the amino acid sequence <SEQ ID 2254>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2589(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10037> which encodes amino acid sequence <SEQ ID 10038> was also identified.

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

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2255> which encodes the amino acid sequence <SEQ ID 2256>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2767(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 80/86 (93%), Positives = 84/86 (97%)

Query: 6  LKTIKENNMTFEEILPGLKAKKKYVRTGWGGAENYVQLFDTLEVNGKVLQATPYFLINVT 65
          + +IKENNMTFEEILPGLKAKKKYVRTGWGGAENYVQLFDTLEV+GKVLQATPYFLI+VT
Sbjct: 3  ISSIKENNMTFEEILPGLKAKKKYVRTGWGGAENYVQLFDTLEVDGKVLQATPYFLIHVT 62

Query: 66 GEGEGFSMWAPTPCDVLAEDWIEVND 91
          G GEGFSMWAPTPCDVLAEDWIEVND
Sbjct: 63 GAGEGFSMWAPTPCDVLAEDWIEVND 88
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 736

A DNA sequence (GBSx0782) was identified in *S.agalactiae* <SEQ ID 2257> which encodes the amino acid sequence <SEQ ID 2258>. Analysis of this protein 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>
```

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

```
>GP:BAA85256 GB:AB021978 3-oxoacyl-[acyl carrier protein] reductase
          homolog [Moritella marina]
 Identities = 82/239 (34%), Positives = 125/239 (51%), Gaps = 15/239 (6%)

Query: 2   TKVVLVTGCASGIGYAQAQYFLKQGYQVVGVDKSDKPNLN-----GNFNF-IKLDLSSDL 55
           +K VLVTG + GIG A A++F K G  V G   S +          G+  F ++L+++S
Sbjct: 5   SKTVLVTGASRGIGRAIAEHFAKLGATVIGTATSAQGAERIGAYLGDAGFGLELNVTSQD 64

Query: 56  S------PLFTMVPTVDILCNTAGILDAYKPLLEVSDEELEHLFDINFFVTVRLTRHYLR 109
           S       + T V  +DIL N AGI  A    L + ++E  ++ D N      RL +  LR
Sbjct: 65  SVDALYAEIKTQVGHIDILVNNAGIT-ADNIFLRMKEDEWCNVIDTNLTSLYRLCKPCLR 123

Query: 110 RMVEKKSGIIINMCSIASFIAGGGGAAYTSSKHALAGFTRQLALDYAKDCIQIFGIAPGA 169
            M++++ G IIN+ S+      GG A Y ++K  L GFT+ LA + A    I +  +APG
Sbjct: 124 GMMKQRHGRIINIGSVVGTTGNGGQANYAAAKSGLLGFTKSLASEVASRGITVNAVAPGF 183

Query: 170 VQTAMTASDFEPGGLAEWVASETPIGRWTKPSEVAELTGFLASGKARSMQGEIVKIDGG 228
           ++T MTA   E     + + ++ P  R    +E+AE  GFLAS  A  + GE + ++GG
Sbjct: 184 IETDMTAELTEE--QKQTILAQVPTSRLGSTTEIAETVGFLASDGASYITGETIHVNGG 240
```

There is also homology to SEQ IDs 2628 and 7170.

A related sequence was also identified in GAS <SEQ ID 9107> which encodes the amino acid sequence <SEQ ID 9108>. Analysis of this protein 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>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 206/232 (88%), Positives = 224/232 (95%)

Query: 1   MTKVVLVTGCASGIGYAQAQYFLKQGYQVVGVDKSDKPNLNGNFNFIKLDLSSDLSPLFT 60
            MTKVVLVTGCASGIGYAQA+YFLKQG+ VVGVDKSDKP+L+GNF+FIKLDLSS+L+PLF
Sbjct: 4   MTKVVLVTGCASGIGYAQARYFLKQGHHVVGVDKSDKPDLSGNFHFIKLDLSSELAPLFK 63

Query: 61  MVPTVDILCNTAGILDAYKPLLEVSDEELEHLFDINFFVTVRLTRHYLRRMVEKKSGIII 120
            +VP+VDILCNTAGILDAYKPLL+VSDEE+EHLFDINFF TV+LTRHYLRRMVEK+SG+II
Sbjct: 64  VVPSVDILCNTAGILDAYKPLLDVSDEEVEHLFDINFFATVKLTRHYLRRMVEKQSGVII 123

Query: 121 NMCSIASFIAGGGGAAYTSSKHALAGFTRQLALDYAKDCIQIFGIAPGAVQTAMTASDFE 180
            NMCSIASFIAGGGG AYTSSKHALAGFTRQLALDYAKD I IFGIAPGAV+TAMTA+DFE
Sbjct: 124 NMCSIASFIAGGGGVAYTSSKHALAGFTRQLALDYAKDQIHIFGIAPGAVKTAMTANDFE 183

Query: 181 PGGLAEWVASETPIGRWTKPSEVAELTGFLASGKARSMQGEIVKIDGGWSLK 232
            PGGLA+WVA ETPIGRWTKP EVAELTGFLASGKARSMQGEIVKIDGGW+LK
Sbjct: 184 PGGLADWVARETPIGRWTKPDEVAELTGFLASGKARSMQGEIVKIDGGWTLK 235
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9063> which encodes amino acid sequence <SEQ ID 9064>. An alignment of the GAS and GBS sequences follows:

```
Score = 83.1 bits (202), Expect = 4e-18
Identities = 72/258 (27%), Positives = 106/258 (40%), Gaps = 36/258 (13%)

Query: 6   EVAFITGAASGIGKQIGETLLKEGKTVVFSDINQE-----KLDQVVADYTKEGYDAFSVV 60
            +V  +TG ASGIG     +  LK+G  V   D + +       + +  D + +    F++V
Sbjct: 3   KVVLVTGCASGIGYAQAQYFLKQGYQVVGVDKSDKPNLNGNFNFIKLDLSSDLSPLFTMV 62

Query: 61  CDVTKEEAINAAIDTVVEKYGRIDILVNNAG-LQHVAMIEDFPTEKFEFMIKIMLTAPFI 119
                                +DIL N AG L      + +    E+ E +  I
Sbjct: 63  ----------------------PTVDILCNTAGILDAYKPLLEVSDEELEHLFDINFFVTVR 102

Query: 120 AIKRAFPTMKAQKHGRIINMASINGVIGFAGKSAYNSAKHGLIGLTKVTALEAADSGITV 179
              +     M  +K G IINM SI   I   G +AY S+KH L G T+  AL+ A   I +
Sbjct: 103 LTRHYLRRMVEKKSGIIINMCSIASFIAGGGGAAYTSSKHALAGFTRQLALDYAKDCIQI 162

Query: 180 NAICPGYVDTPLVRGQFEDLSKTRGIPLENVLEEVLYPLVPQKRLIDVQEIADYVSFLAS 239
              I PG V T +    FE            L E +    P  R      E+A+   FLAS
Sbjct: 163 FGIAPGAVQTAMTASDFE----------PGGLAEWVASETPIGRWTKPSEVAELTGFLAS 212

Query: 240 DKAKGVTGQACILDGGYT 257
             KA+ + G+   +DGG++
Sbjct: 213 GKARSMQGEIVKIDGGWS 230
```

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 2259> which encodes the amino acid sequence <SEQ ID 2260>. An alignment of the GAS and GBS sequences follows:

```
Score =  427 bits (1086), Expect = e-122
Identities = 206/232 (88%), Positives = 224/232 (95%)

Query: 4   MTKVVLVTGCASGIGYAQARYFLKQGHHVVGVDKSDKPDLSGNFHFIKLDLSSELAPLFK 63
            MTKVVLVTGCASGIGYAQA+YFLKQG+ VVGVDKSDKP+L+GNF+FIKLDLSS+L+PLF
Sbjct: 1   MTKVVLVTGCASGIGYAQAQYFLKQGYQVVGVDKSDKPNLNGNFNFIKLDLSSDLSPLFT 60

Query: 64  VVPSVDILCNTAGILDAYKPLLDVSDEEVEHLFDINFFATVKLTRHYLRRMVEKQSGVII 123
            +VP+VDILCNTAGILDAYKPLL+VSDEE+EHLFDINFF TV+LTRHYLRRMVEK+SG+II
Sbjct: 61  MVPTVDILCNTAGILDAYKPLLEVSDEELEHLFDINFFVTVRLTRHYLRRMVEKKSGIII 120

Query: 124 NMCSIASFIAGGGGVAYTSSKHALAGFTRQLALDYAKDQIHIFGIAPGAVKTAMTANDFE 183
            NMCSIASFIAGGGG AYTSSKHALAGFTRQLALDYAKD I IFGIAPGAV+TAMTA+DFE
```

```
Sbjct: 121 NMCSIASFIAGGGGAAYTSSKHALAGFTRQLALDYAKDCIQIFGIAPGAVQTAMTASDFE 180

Query: 184 PGGLADWVARETPIGRWTKPDEVAELTGFLASGKARSMQGEIVKIDGGWTLK 235
           PGGLA+WVA ETPIGRWTKP EVAELTGFLASGKARSMQGEIVKIDGGW+LK
Sbjct: 181 PGGLAEWVASETPIGRWTKPSEVAELTGFLASGKARSMQGEIVKIDGGWSLK 232
```

SEQ ID 2258 (GBS251) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 2; MW 21.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 6; MW 52kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 737

A DNA sequence (GBSx0783) was identified in *S.agalactiae* <SEQ ID 2261> which encodes the amino acid sequence <SEQ ID 2262>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.82   Transmembrane   62 -  78 (  62 -  79)

----- 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>
```

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

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

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

Example 738

A DNA sequence (GBSx0784) was identified in *S.agalactiae* <SEQ ID 2263> which encodes the amino acid sequence <SEQ ID 2264>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1495(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 GENPEPT database:

```
>GP:CAA20397 GB:AL031317 SC6G4.19c, unknown, len: 190 aa; contains
           Pro-Ser- r ich domain at N-terminus [Streptomyces
           coelicolor A3(2)]
 Identities = 26/80 (32%), Positives = 44/80 (54%), Gaps = 5/80 (6%)

Query: 1   MDSNDEAICIIEITKVDIVPFKDVSADHAFKEGEGDKTLEWWRKAHIDFF-----KPYFE 55
           +DS +  + +IE+T+V +VP  +V   HA  EGEGD ++  WR  H  F+     +
Sbjct: 103 VDSRERPVAVIEVTEVRVVPLAEVDLAHAVDEGEGDTSVAGWRAGHERFWHGAEMRAALG 162

Query: 56  EFGLMFSEDSRIVLEEFQVV 75
           + G    + + +VLE F++V
Sbjct: 163 DPGFTVDDATPVVLERFRIV 182
```

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

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

Example 739

A DNA sequence (GBSx0785) was identified in *S.agalactiae* <SEQ ID 2265> which encodes the amino acid sequence <SEQ ID 2266>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.49   Transmembrane    3 -  19 (   3 -  19)

----- 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>
```

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

```
>GP:BAB06422 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 133/315 (42%), Positives = 191/315 (60%), Gaps = 4/315 (1%)

Query: 1   MKLAVLGTGMIVKEVLPVLQKIEGIDLVAILSTVRSLETAKDLAKEYNMSLATSEYKAVL 60
           MK+A +GTG IV+  L  L  I+G   VA+ S  R   TAK LA +YN+    + +  +L
Sbjct: 1   MKIATVGTGPIVEAFLSALDDIDGPMCVAMYS--RKETTAKPLADQYNIPTIYTHFDHML 58

Query: 61  DNEEIDTVYIGLPNHLHFDYAKEALLAGKHVICEKPFTLEASQLEELVSIANTRQLILLE 120
            +  ++ VY+  PN LH+ +A +AL   KHVICEKPFT  A +LE L+S+A   +L+L E
Sbjct: 59  ADPNVEVVYVASPNSLHYQHALQALEHRKHVICEKPFTSTARELEHLISVARKNELMLFE 118

Query: 121 AITNQYLPNFDLVKEHLSNLGDIKIVECNYSQYSSRYDAFKRGEIAPAFNPEMGGGALRD 180
           AIT  +LPN+ L+KE++  LG IK+++CNYSQYSSRYD F  GE    FNP   GGAL D
Sbjct: 119 AITTIHLPNYQLIKENIHKLGSIKMIQCNYSQYSSRYDRFLSGETPNVFNPAFSGGALMD 178

Query: 181 LNIYNLHLVIGLFGEPITAQYLPNIE-RGIDTSGVLVLDYGHFKTVCIGAKDCSAEVKST 239
           +N+YN+H V+ LFG P  A Y+ N    GIDTSGVLVL Y HF + C+G KD  +
Sbjct: 179 INVYNIHFVMNLFGPPEAAHYIANQHANGIDTSGVLVLKYPHFISECVGCKDTQSMNFVL 238

Query: 240 IQGDKGSIAILGPTNTMPKISLTMNGQESHVYQLNGDRHRMHDEFVIFEGIISNLDFKRA 299
           IQG+KG I +    N    + + ++ Q S +     D    ++E +    +F++
Sbjct: 239 IQGEKGYIHVENGANGCRNVKIYLDDQTSELNAQTNDNLLYYETRTFYE-MYQAKNFEKC 297

Query: 300 AQALEHSRTVMKVLD 314
            + L +S +VM+V++
Sbjct: 298 YELLSYSHSVMRVME 312
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 719> which encodes the amino acid sequence <SEQ ID 720>. 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>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 233/314 (74%), Positives = 269/314 (85%)

Query: 1   MKLAVLGTGMIVKEVLPVLQKIEGIDLVAILSTVRSLETAKDLAKEYNMSLATSEYKAVL 60
           MKLAVLGTGMIVKEVLPVLQKI+GIDLVAILSTVRSL TAKDLAK ++M LATS+Y+A+L
```

```
Sbjct: 1   MKLAVLGTGMIVKEVLPVLQKIDGIDLVAILSTVRSLTTAKDLAKAHHMPLATSKYEAIL 60

Query: 61  DNEEIDTVYIGLPNHLHFDYAKEALLAGKHVICEKPFTLEASQLEELVSIANTRQLILLE 120
            NEEIDTVYIGLPNHLHF YAKEALLAGKHVICEKPFT+ A +L+ELV IA  R+LILLE
Sbjct: 61  GNEEIDTVYIGLPNHLHFAYAKEALLAGKHVICEKPFTMTAGELDELVVIARKRKLILLE 120

Query: 121 AITNQYLPNFDLVKEHLSNLGDIKIVECNYSQYSSRYDAFKRGEIAPAFNPEMGGGALRD 180
            AITNQYL N   +KEHL  LGDIKIVECNYSQYSSRYDAFKRG+IAPAFNP+MGGGALRD
Sbjct: 121 AITNQYLSNMTFIKEHLDQLGDIKIVECNYSQYSSRYDAFKRGDIAPAFNPKMGGGALRD 180

Query: 181 LNIYNLHLVIGLFGEPITAQYLPNIERGIDTSGVLVLDYGHFKTVCIGAKDCSAEVKSTI 240
            LNIYN+H V+GLFG P T QYL N+E+GIDTSG+LV+DY  FK VCIGAKDC+AE+KSTI
Sbjct: 181 LNIYNIHFVVGLFGRPKTVQYLANVEKGIDTSGMLVMDYEQFKVVCIGAKDCTAEIKSTI 240

Query: 241 QGDKGSIAILGPTNTMPKISLTMNGQESHVYQLNGDRHRMHDEFVIFEGIISNLDFKRAA 300
            QG+KGS+A+LG TNT+P++ L+++G E  V   N   HRM++EFV F  +I   DF++
Sbjct: 241 QGNKGSLAVLGATNTLPQVQLSLHGHEPQVINHNKHDHRMYEEFVAFRDMIDQRDFEKVN 300

Query: 301 QALEHSRTVMKVLD 314
            QALEHSR VM VL+
Sbjct: 301 QALEHSRAVMAVLE 314
```

SEQ ID 2266 (GBS342) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 10; MW 36.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 2; MW 61kDa).

GBS342-GST was purified as shown in Figure 226, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 740

A DNA sequence (GBSx0786) was identified in *S.agalactiae* <SEQ ID 2267> which encodes the amino acid sequence <SEQ ID 2268>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0499(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 GENPEPT database:

```
>GP:CAB12535 GB:Z99107 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 41/127 (32%), Positives = 63/127 (49%), Gaps = 11/127 (8%)

Query: 1   MISSIGQVMLYVSNVEASADFWKNKVGFERVEKQTQGDYVTYI-VAPKLDSEVSFVLHDK 59
            MI  IG V +YV + + +  FW  KVGF+       G   +++ VAPK  +E   V++ K
Sbjct: 1   MIKQIGTVAVYVEDQQKAKQFWTEKVGFDIAADHPMGPEASWLEVAPK-GAETRLVIYPK 59

Query: 60  AIIAQMSPELDLATPSILFETTDIDSTYQELTAN--EVMTNP-IVDMGSMRVFNFSDNDN 116
            A    M    +   SI+FE  DI  TY+++  N  E +  P  ++ G+     F D D
Sbjct: 60  A----MMKGSEQMKASIVFECEDIFGTYEKMKTNGVEFLGEPNQMEWGTF--VQFKDEDG 113

Query: 117 NYFAIRE 123
            N F ++E
Sbjct: 114 NVFLLKE 120
```

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

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

Example 741

A DNA sequence (GBSx0787) was identified in *S.agalactiae* <SEQ ID 2269> which encodes the amino acid sequence <SEQ ID 2270>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3402(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 GENPEPT database:

```
>GP:BAB04569 GB:AP001510 unknown conserved protein in others
           [Bacillus halodurans]
 Identities = 46/144 (31%), Positives = 83/144 (56%), Gaps = 10/144 (6%)

Query: 1   MVKALETYIVTNGNGRQAVDFYKDVFQADLVNMMTWEEM--DPNC--LEDRKDLIINAQL 56
           M+  +  Y++ +G+G+ A++FY+D   A+++ + T+ ++   PN       KDLI++A L
Sbjct: 1   MILTMNPYLMLDGDGQAAIEFYQDALNAEVITIQTYGDLPEQPNSPMASVNKDLILHAHL 60

Query: 57  IFDGIRLQISDENPD-----FVYQAGKNVTAAIIVGSVEEAREIYEKLKKSAQEVQLELQ 111
              + L ISD+  D     F   +G  VT A+   +VE   E+++KL    +E+   L+
Sbjct: 61  KLGEMDLMISDQCLDVDPERFPQHSGSPVTIALTTNNVEMTTEVFQKLASGGEEIA-PLE 119

Query: 112 ETFWSPAYANLVDQFGVMWQISTE 135
           +TF+SP Y  + D+FG+ W +ST+
Sbjct: 120 KTFFSPLYGQVTDKFGITWHVSTQ 143
```

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

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

Example 742

A DNA sequence (GBSx0788) was identified in *S.agalactiae* <SEQ ID 2271> which encodes the amino acid sequence <SEQ ID 2272>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> 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>
```

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

```
>GP:BAB03784 GB:AP001507 UDP-N-acetylglucosamine pyrophosphorylase
           [Bacillus halodurans]
 Identities = 238/453 (52%), Positives = 322/453 (70%), Gaps = 1/453 (0%)

Query: 1   MSN-YAIILAAGKGTRMKSDLPKVMHKVSGITMLEHVFRSVQAIEPSKIVTVIGHKAELV 59
           MSN +A+ILAAG+GTRMKS L KV+H V G  M++HV   V A+   +IVT+IGH A+ V
Sbjct: 1   MSNRFAVILAAGQGTRMKSKLYKVLHSVCGKPMVQHVVDQVSALGFDEIVTIIGHGADAV 60

Query: 60  RDVLGDKSEFVMQTEQLGTGHAVMMAEEELATSKGHTLVIAGDTPLITGESLKNLIDFHV 119
           +  LG++  + +Q EQLGTGHAV+ AE  L   +G T+V+ GDTPL+T E++ +++ +H
```

```
Sbjct: 61  KSQLGERVSYALQEEQLGTGHAVLQAESALGGRRGVTIVLCGDTPLLTAETIDHVMSYHE 120

Query: 120 NHKNVATILTADAANPFGYGRIIRNSDDEVTKIVEQKDANDFEQQVKEINTGTYVFDNQS 179
             +  AT+LTA+ A+P GYGRI+RN    V +IVE KDA   E+Q+ E+NTGTY FDN++
Sbjct: 121 EEQAKATVLTAELADPTGYGRIVRNDKGLVERIVEHKDATSEEKQITEVNTGTYCFDNEA 180

Query: 180 LFEALKDINTNNAQGEYYLTDVIGIFKEAGKKVGAYKLRDFDESLGVNDRVALATAEKVM 239
             LF+ALK++  NNAQGEYYL DVI I +  G+KV AYK    +E+LGVNDRVALA AE+VM
Sbjct: 181 LFQALKEVGNNNAQGEYYLPDVIQILQTKGEKVAAYKTAHVEETLGVNDRVALAQAEQVM 240

Query: 240 RHRIARQHMVNGVTVVNPDSAYIDIDVEIGEESVIEPNVTLKGQTKIGKGTLLTNGSYLV 299
             + RI    M  GVT ++P+   Y+  D  IG+++VI P   + GQT IG+G +L   + L
Sbjct: 241 KRRINEAWMRKGVTFIDPEQTYVSPDATIGQDTVIYPGTMVLGQTTIGEGCVLGPHTELK 300

Query: 300 DAQVGNDVTITNSMVEESIISDGVTVGPYAHIRPGTSLAKGVHIGNFVEVKGSQIGENTK 359
             D+++GN   +  S+V  S + + V++GP++HIRP + +   V IGNFVEVK S IG+ +K
Sbjct: 301 DSKIGNKTAVKQSVVHNSEVGERVSIGPFSHIRPASMIHDDVRIGNFVEVKKSTIGKESK 360

Query: 360 AGHLTYIGNAEVGCDVNFGAGTITVNYDGQNKFKTEIGSNVFIGSNSTLIAPLEIGDNAL 419
             A HL+YIG+AEVG  VNF  G+ITVNYDG+NKF T+I  + FIG NS LIAP+ IG  AL
Sbjct: 361 ASHLSYIGDAEVGERVNFSCGSITVNYDGKNKFLTKIEDDAFIGCNSNLIAPVTIGKGAL 420

Query: 420 TAAGSTITDNVPIDSIAIGRGRQVNKEGYANKK 452
              AAGSTIT++VP D+++I R RQ NKE Y  KK
Sbjct: 421 IAAGSTITEDVPSDALSIARARQTNKEHYVTKK 453
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2273> which encodes the amino acid sequence <SEQ ID 2274>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0461(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 345/458 (75%), Positives = 398/458 (86%)

Query: 1   MSNYAIILAAGKGTRMKSDLPKVMHKVSGITMLEHVFRSVQAIEPSKIVTVIGHKAELVR 60
            M+NYAIILAAGKGTRM SDLPKV+HKVSG+TMLEHVFRSV+AI P K VTVIGHK+E+VR
Sbjct: 1   MTNYAIILAAGKGTRMTSDLPKVLHKVSGLTMLEHVFRSVKAISPEKSVTVIGHKSEMVR 60

Query: 61  DVLGDKSEFVMQTEQLGTGHAVMMAEEELATSKGHTLVIAGDTPLITGESLKNLIDFHVN 120
             VL D+S FV QTEQLGTGHAVMMAE +L   +GHTLVIAGDTPLITGESLK+LIDFHVN
Sbjct: 61  AVLADQSAFVHQTEQLGTGHAVMMAETQLEGLEGHTLVIAGDTPLITGESLKSLIDFHVN 120

Query: 121 HKNVATILTADAANPFGYGRIIRNSDDEVTKIVEQKDANDFEQQVKEINTGTYVFDNQSL 180
            HKNVATILTA A +PFGYGRI+RN D EV KIVEQKDAN++EQQ+KEINTGTYVFDN+ L
Sbjct: 121 HKNVATILTATAQDPFGYGRIVRNKDGEVIKIVEQKDANEYEQQLKEINTGTYVFDNKRL 180

Query: 181 FEALKDINTNNAQGEYYLTDVIGIFKEAGKKVGAYKLRDFDESLGVNDRVALATAEKVMR 240
             FEALK I TNNAQGEYYLTDV+ IF+   +KVGAY LRDF+ESLGVNDRVALA AE VMR
Sbjct: 181 FEALKCITTNNAQGEYYLTDVVAIFRANKEKVGAYILRDFNESLGVNDRVALAIAETVMR 240

Query: 241 HRIARQHMVNGVTVVNPDSAYIDIDVEIGEESVIEPNVTLKGQTKIGKGTLLTNGSYLVD 300
             RI ++HMVNGVT  NP++ YI+ DVEI  + +IE NVTLKG+T IG GT+LTNG+Y+VD
Sbjct: 241 QRITQKHMVNGVTFQNPETVYIESDVEIAPDVLIEGNVTLKGRTHIGSGTVLTNGTYIVD 300

Query: 301 AQVGNDVTITNSMVEESIISDGVTVGPYAHIRPGTSLAKGVHIGNFVEVKGSQIGENTKA 360
             +++G++  +TNSM+E S+++ GVTVGPYAH+RPGT+L + VHIGNFVEVKGS IGE TKA
Sbjct: 301 SEIGDNCVVTNSMIESSVLAAGVTVGPYAHLRPGTTLDREVHIGNFVEVKGSHIGEKTKA 360

Query: 361 GHLTYIGNAEVGCDVNFGAGTITVNYDGQNKFKTEIGSNVFIGSNSTLIAPLEIGDNALT 420
             GHLTYIGNA+VG  VN GAGTITVNYDGQNK++T IG + FIGSNSTLIAPLE+GD+ALT
```

```
Sbjct: 361 GHLTYIGNAQVGSSVNVGAGTITVNYDGQNKYETVIGDHAFIGSNSTLIAPLEVGDHALT 420

Query: 421 AAGSTITDNVPIDSIAIGRGRQVNKEGYANKKPHHPSQ 458
           AAGSTI+  VPIDSIAIGR RQV KEGYA +  HHPS+
Sbjct: 421 AAGSTISKTVPIDSIAIGRSRQVTKEGYAKRLAHHPSR 458
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 743

A DNA sequence (GBSx0790) was identified in *S.agalactiae* <SEQ ID 2275> which encodes the amino acid sequence <SEQ ID 2276>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1366(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 GENPEPT database:

```
>GP:CAB14293 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 92/177 (51%), Positives = 124/177 (69%), Gaps = 4/177 (2%)

Query: 4   EEKTINRQTVFDGQIIKVAVDDVELPNGLGQSKRELVFHGGAVATLAVTPEHKIVLVKQY 63
           EEKTI ++ +F G++I + V+DVELPNG   SKRE+V H GAVA LAVT E KI++VKQ+
Sbjct: 5   EEKTIAKEQIFSGKVIDLYVEDVELPNGKA-SKREIVKHPGAVAVLAVTDEGKIIMVKQF 63

Query: 64  RKAIEGISYEIPAGKLETGESGSKEEAALRELEEETGYTG-NLEILYSFYTAIGFCNEKI 122
           RK +E    EIPAGKLE GE    E  ALRELEEETGYT   L  + +FYT+ GF +E +
Sbjct: 64  RKPLERTIVEIPAGKLEKGE--EPEYTALRELEEETGYTAKKLTKITAFYTSPGFADEIV 121

Query: 123 VLYLATDLQKVENPRPQDDDEVLELLELSYEDCMQMVEKGMIQDAKTIIALQYYGLK 179
            ++LA +L  +E  R  D+DE +E++E++ ED +++VE   + DAKT  A+QY  LK
Sbjct: 122 HVFLAEELSVLEEKRELDEDEFVEVMEVTLEDALKLVESREVYDAKTAYAIQYLQLK 178
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2277> which encodes the amino acid sequence <SEQ ID 2278>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1120(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 136/182 (74%), Positives = 153/182 (83%)

Query: 1   MDFEEKTINRQTVFDGQIIKVAVDDVELPNGLGQSKRELVFHGGAVATLAVTPEHKIVLV 60
           M FEEKT+ RQTVFDG I KV VDDVELPN LGQSKREL+FH GAVA LA+TPE KIVLV
Sbjct: 1   MKFEEKTLKRQTVFDGHIFKVVVDDVELPNNLGQSKRELIFHRGAVAVLAITPERKIVLV 60

Query: 61  KQYRKAIEGISYEIPAGKLETGESGSKEEAALRELEEETGYTGNLEILYSFYTAIGFCNE 120
           KQYRKAIE +SYEIPAGKLE GE GSK +AA RELEEET YTG L  LY FYTAIGFCNE
Sbjct: 61  KQYRKAIERVSYEIPAGKLEIGEEGSKLKAAARELEEETAYTGTLTFLYEFYTAIGFCNE 120

Query: 121 KIVLYLATDLQKVENPRPQDDDEVLELLELSYEDCMQMVEKGMIQDAKTIIALQYYGLKM 180
           KI L+LATDL +V NP+PQDDDEV+E+LEL+Y++CM +V +G + DAKT+IALQYY L
Sbjct: 121 KITLFLATDLIQVANPKPQDDDEVIEVLELTYQECMDLVAQGKLADAKTLIALQYYALHF 180
```

```
Query: 181 GG 182
           GG
Sbjct: 181 GG 182
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 744

A DNA sequence (GBSx0791) was identified in *S.agalactiae* <SEQ ID 2279> which encodes the amino acid sequence <SEQ ID 2280>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-15.44   Transmembrane   70 -  86 (  64 -  88)

----- Final Results -----
               bacterial membrane --- Certainty=0.7177(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2281> which encodes the amino acid sequence <SEQ ID 2282>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-15.60   Transmembrane   65 -  81 (  58 -  83)

----- Final Results -----
               bacterial membrane --- Certainty=0.7241(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 39/89 (43%), Positives = 61/89 (67%), Gaps = 6/89 (6%)

Query: 1  MGKPLLTDDMIERSNRGEKVSGQTILDQETKIISTEDGMEQLTDENGKHIYKSRRIENAK 60
          MG+PLLTDD+IE++ R E        ++ +TK+++  +        ++   IYKSRRIENAK
Sbjct: 2  MGRPLLTDDIIEKARRMETFEPDDAVNFDTKVMTLPE------KDDKARIYKSRRIENAK 55

Query: 61 RNEFQRKLNLVLFILLILLALLFYAIFKL 89
          R++ Q KLN++L  +++L+A+L YAIF L
Sbjct: 56 RSQLQSKLNVILIAVMLLIAILVYAIFYL 84
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 745

A DNA sequence (GBSx0792) was identified in *S.agalactiae* <SEQ ID 2283> which encodes the amino acid sequence <SEQ ID 2284>. This protein is predicted to be pfs protein (pfs). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane   56 -  72 (  56 -  72)

----- Final Results -----
```

```
bacterial membrane --- Certainty=0.1128(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 GENPEPT database:

```
>GP:AAC22869 GB:U32801 pfs protein (pfs) [Haemophilus influenzae Rd]
 Identities = 100/229 (43%), Positives = 144/229 (62%)

Query: 1   MKIGIIAAMEEELKLLVENLEDKSQETVLSNVYYSGRYGEHELVLVQSGVGKVMSAMSVA 60
           MKIGI+ AM +E+++L   + D+++  V S V + G+     ++ L+QSG+GKV +A+
Sbjct: 1   MKIGIVGAMAQEVEILKNLMADRTETRVASAVIFEGKINGKDVALLQSGIGKVAAAIGTT 60

Query: 61  ILVESFKVDAIINTGSAGAVATGLNVGDVVVADTLVYHDVDLTAFGYDYGQMSMQPLYFH 120
            L++  K D +INTGSAG VA GL VGD+V++D   YHD D+TAFGY+ GQ+   P  F
Sbjct: 61  ALLQLAKPDCVINTGSAGGVAKGLKVGDIVISDETRYHDADVTAFGYEKGQLPANPAAFL 120

Query: 121 SDKTFVSTFEAVLSKEEMISKVGLIATGDSFIAGQEKIDVIKGHFPQVLAVEMEGAAIAQ 180
           SDK      + +  K+    K GLI +GDSFI  ++KI  IK  FP V  VEME  AIAQ
Sbjct: 121 SDKKLADLAQEIAEKQGQSVKRGLICSGDSFINSEDKIAQIKADFPNVTGVEMEATAIAQ 180

Query: 181 AAQATGKPFVVVRAMSDTAAHDANITFDEFIIEAGKRSAQVLMAFLKAL 229
              A   PFVVVRA+SD     A+++F+EF+  A K+S+ +++  +  L
Sbjct: 181 VCYAFNVPFVVVRAISDGGDGKASMSFEEFLPLAAKQSSALVLGMIDRL 229
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2285> which encodes the amino acid sequence <SEQ ID 2286>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1245(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 169/229 (73%), Positives = 189/229 (81%)

Query: 1   MKIGIIAAMEEELKLLVENLEDKSQETVLSNVYYSGRYGEHELVLVQSGVGKVMSAMSVA 60
           MKIGIIAAMEEEL LL+ NL D  +  VLS  YY+GR+G+HEL+LVQSGVGKVMSAM+VA
Sbjct: 1   MKIGIIAAMEEELSLLLANLLDAQEHQVLSKTYYTGRFGKHELILVQSGVGKVMSAMTVA 60

Query: 61  ILVESFKVDAIINTGSAGAVATGLNVGDVVVADTLVYHDVDLTAFGYDYGQMSMQPLYFH 120
           ILVE FK  AIINTGSAGAVA+ L +GDVVVAD LVYHDVD TAFGY YGQM+ QPLY+
Sbjct: 61  ILVEHFKAQAIINTGSAGAVASHLAIGDVVVADRLVYHDVDATAFGYAYGQMAGQPLYYD 120

Query: 121 SDKTFVSTFEAVLSKEEMISKVGLIATGDSFIAGQEKIDVIKGHFPQVLAVEMEGAAIAQ 180
            D  FV+ F+ VL  E+   +VGLIATGDSF+AGQ+KID IK  F  VLAVEMEGAAIAQ
Sbjct: 121 CDPQFVAIFKQVLKHEKTNGQVGLIATGDSFVAGQDKIDQIKTAFSDVLAVEMEGAAIAQ 180

Query: 181 AAQATGKPFVVVRAMSDTAAHDANITFDEFIIEAGKRSAQVLMAFLKAL 229
           AA   GKPF+VVRAMSDTAAHDANITFD+FIIEAGKRSAQ LM FL+ L
Sbjct: 181 AAHTAGKPFIVVRAMSDTAAHDANITFDQFIIEAGKRSAQTLMTFLENL 229
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 746

A DNA sequence (GBSx0793) was identified in *S.agalactiae* <SEQ ID 2287> which encodes the amino acid sequence <SEQ ID 2288>. This protein is predicted to be SloR. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3777(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9405> which encodes amino acid sequence <SEQ ID 9406> was also identified.

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

```
>GP:AAF81675 GB:AF232688 SloR [Streptococcus mutans]
 Identities = 97/175 (55%), Positives = 134/175 (76%)

Query: 1   MSEMIKKMISEQLIVKDKDLGYYLTKQGLLVVSDLYRKHRLVEVFLVNHLHYTADDIHEE 60
            +SEM+KK++ E L++KDK  GY LTK+G ++ S LYRKHRL+EVFL+NHL+YTAD+IHEE
Sbjct: 38  VSEMVKKLLLEDLVLKDKQAGYLLTKKGQILASSLYRKHRLIEVFLMNHLNYTADEIHEE 97

Query: 61  AEVLEHTVSTTFVDQLEKLLDFPQFCPHGGTIPKKGEFLVEINQMTLDQISQLGTYVISR 120
            AEVLEHTVS  FV++L+K L++P+ CPHGGTIP+ G+ LVE  + TL  ++++G Y++ R
Sbjct: 98  AEVLEHTVSDVFVERLDKFLNYPKVCPHGGTIPQHGQPLVERYRTTLKGVTEMGVYLLKR 157

Query: 121 VHDDFQLLKYLEQHRLHINDTIELTQIDPYAKTYHITYNDENLTIPERIASQIYV 175
            V D+FQLLKY+EQH L I D + L + D +A  Y I  + E L +   +ASQIY+
Sbjct: 158 VQDNFQLLKYMEQHHLKIGDELRLLEYDAFAGAYTIEKDGEQLQVTSAVASQIYI 212
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2289> which encodes the amino acid sequence <SEQ ID 2290>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2910(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 44/75 (58%), Positives = 59/75 (78%)

Query: 1   MSEMIKKMISEQLIVKDKDLGYYLTKQGLLVVSDLYRKHRLVEVFLVNHLHYTADDIHEE 60
            +SEMIKKMIS+  IVKDK  GY L  +G  +V++LYRK RL+EVFL++ L Y   ++H+E
Sbjct: 38  VSEMIKKMISQGWIVKDKAKGYLLKDKGYALVANLYRKLRLIEVFLIHQLGYNTQEVHQE 97

Query: 61  AEVLEHTVSTTFVDQ 75
            AEVLEHTVS +F+D+
Sbjct: 98  AEVLEHTVSDSFIDR 112
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 747

A DNA sequence (GBSx0794) was identified in *S.agalactiae* <SEQ ID 2291> which encodes the amino acid sequence <SEQ ID 2292>. This protein is predicted to be undecaprenyl pyrophosphate synthetase (uppS). Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3569(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9435> which encodes amino acid sequence <SEQ ID 9436> was also identified.

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

```
>GP:CAB13526 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 88/165 (53%), Positives = 118/165 (71%), Gaps = 4/165 (2%)

Query: 1   MNLPVKFFDKYVPELDKNNVRVQVIGDTHKLPKATYDAMQRACLRTKHNSGLVLNFALNY 60
           M LP +F + Y+PEL + NV+V++IGD   LP  T  A+++A   T  N G++LNFALNY
Sbjct: 100 MKLPEEFLNTYLPELVEENVQVRIIGDETALPAHTLRAIEKAVQDTAQNDGMILNFALNY 159

Query: 61  GGRSEITNAIKEIAQDVLEAKLNPDDITEDLVANHLMTNSLPYLYRDPDLIIRTSGELRL 120
           GGR+EI +A K +A+ V E  LN +DI E L + +LMT SL    +DP+L+IRTSGE+RL
Sbjct: 160 GGRTEIVSAAKSLAEKVKEGSLNIEDIDESLFSTYLMTESL----QDPELLIRTSGEIRL 215

Query: 121 SNFLPWQSAYSEFYFTPVLWPDFKKDELHKAIVDYNQRHRRFGSV 165
           SNF+ WQ AYSEF FT VLWPDFK+D   +A+ ++ QR RRFG +
Sbjct: 216 SNFMLWQVAYSEFVFTDVLWPDFKEDHFLQALGEFQQRGRRFGGI 260
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2293> which encodes the amino acid sequence <SEQ ID 2294>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2073(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 125/165 (75%), Positives = 145/165 (87%)

Query: 1   MNLPVKFFDKYVPELDKNNVRVQVIGDTHKLPKATYDAMQRACLRTKHNSGLVLNFALNY 60
           MNLPV FFDKYVP L +NNV++Q+IG+T +LP+ T  A+  A  +TK N+GL+LNFALNY
Sbjct: 85  MNLPVTFFDKYVPVLHENNVKIQMIGETSRLPEDTLAALNAAIDKTKRNTGLILNFALNY 144

Query: 61  GGRSEITNAIKEIAQDVLEAKLNPDDITEDLVANHLMTNSLPYLYRDPDLIIRTSGELRL 120
           GGR+EIT+A++ IAQDVL+AKLNP DITEDL+AN+LMT+ LPYLYRDPDLIIRTSGELRL
Sbjct: 145 GGRAEITSAVRFIAQDVLDAKLNPGDITEDLIANYLMTDHLPYLYRDPDLIIRTSGELRL 204

Query: 121 SNFLPWQSAYSEFYFTPVLWPDFKKDELHKAIVDYNQRHRRFGSV 165
           SNFLPWQSAYSEFYFTPVLWPDFKK EL KAI DYN+R RRFG V
Sbjct: 205 SNFLPWQSAYSEFYFTPVLWPDFKKAELLKAIADYNRRQRRFGKV 249
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 748

A DNA sequence (GBSx0795) was identified in *S.agalactiae* <SEQ ID 2295> which encodes the amino acid sequence <SEQ ID 2296>. This protein is predicted to be phosphatidate cytidylyltransferase (cdsA). Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
```

```
    INTEGRAL    Likelihood = -8.65   Transmembrane  201 - 217 ( 194 - 222)
    INTEGRAL    Likelihood = -7.96   Transmembrane  175 - 191 ( 170 - 197)
    INTEGRAL    Likelihood = -5.89   Transmembrane   81 -  97 (  74 -  99)
    INTEGRAL    Likelihood = -3.03   Transmembrane   26 -  42 (  23 -  42)
    INTEGRAL    Likelihood = -2.92   Transmembrane  136 - 152 ( 135 - 153)
    INTEGRAL    Likelihood = -2.02   Transmembrane   49 -  65 (  47 -  66)
    INTEGRAL    Likelihood = -0.64   Transmembrane  248 - 264 ( 248 - 264)

----- 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>
```

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

```
>GP:BAB06141 GB:AP001515 phosphatidate cytidylyltransferase
          [Bacillus halodurans]
 Identities = 116/266 (43%), Positives = 172/266 (64%), Gaps = 6/266 (2%)

Query: 1   MKERVIWGAVALAIFIPFLVMGGLPFQFLVGLLAMIGVSELLRMRRLEIFSFEGALAMIG 60
           MK+RV+   +   +F+ F+V+GGLPF   + ++A I +SELL+M+++  FS  GA +++
Sbjct: 1   MKQRVVTAIIFGLVFLTFVVVGGLPFTMFIIVVATIAMSELLKMKKIAPFSPMGAFSLLP 60

Query: 61  AFVLTVPLDSYLSFLPVDASLSAYGIVIFMILAGTVLNSNSYSFEDAAFPIASSFYVGIG 120
            ++L +P D +   +P    +  +   I  +L  TVL  N+++F++A F I SS Y+G G
Sbjct: 61  MWMLLLPNDWFKVVIPDFTKVEIFIFFILFLLLLTVLTKNTFTFDEAGFVILSSAYIGYG 120

Query: 121 FQNLVSARMA---GIDKVLLALFIVWATDIGAYMIGRQFGQRKLLPSVSPNKTIEGSLGG 177
           F  L+ +R     G+  V   LF++WATD GAY  GR FG+ KL P +SPNKTIEGS+GG
Sbjct: 121 FHFLLLSREIPEIGLPLVFFVLFVIWATDSGAYFAGRAFGKHKLWPHISPNKTIEGSIGG 180

Query: 178 IASAIVVAFFFMLFDKTVYAPHSFLVMLVLVAIFSIFGQFGDLVESSIKRHFGVKDSGKL 237
           I  A+++   F           S+ V L ++ + S+FGQ GDLVES++KRH+ VKDSG +
Sbjct: 181 IILAVIIGSLFYWIMPLF---SSYGVALAVIVVASVFGQLGDLVESALKRHYAVKDSGTV 237

Query: 238 IPGHGGILDRFDSMIFVFPIMHFFGL 263
           +PGHGGILDRFDS+I+V PI+H   L
Sbjct: 238 LPGHGGILDRFDSLIYVMPILHLLHL 263
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2297> which encodes the amino acid sequence <SEQ ID 2298>. Analysis of this protein sequence reveals the following:

```
     Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -9.98   Transmembrane  175 - 191 ( 170 - 197)
    INTEGRAL    Likelihood = -8.97   Transmembrane    5 -  21 (   4 -  42)
    INTEGRAL    Likelihood = -6.85   Transmembrane  201 - 217 ( 197 - 222)
    INTEGRAL    Likelihood = -6.53   Transmembrane   81 -  97 (  79 -  99)
    INTEGRAL    Likelihood = -4.73   Transmembrane   49 -  65 (  47 -  71)
    INTEGRAL    Likelihood = -3.40   Transmembrane  136 - 152 ( 135 - 153)
    INTEGRAL    Likelihood = -3.24   Transmembrane   26 -  42 (  22 -  42)
    INTEGRAL    Likelihood = -1.17   Transmembrane  248 - 264 ( 248 - 264)

----- 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>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB06141 GB:AP001515 phosphatidate cytidylyltransferase
          [Bacillus halodurans]
 Identities = 125/266 (46%), Positives = 177/266 (65%), Gaps = 6/266 (2%)

Query: 1   MKERVVWGGVAVAIFLPFLIIGNLPFQLFVGVLAMIGVSELLKMKRLEVFSFEGVFAMLA 60
           MK+RVV   +   +FL F+++G LPF +F+ V+A I +SELLKMK++  FS  G F++L
Sbjct: 1   MKQRVVTAIIFGLVFLTFVVVGGLPFTMFIIVVATIAMSELLKMKKIAPFSPMGAFSLLP 60
```

```
Query: 61   AFVLAVPMDHYLTFLPIDANVAFYSLMVFFILAGTVLNSRAYSFDDAAFPIATSFYVGIG 120
            ++L +P D +   +P    V  +   + F+L  TVL    ++FD+A F I +S Y+G G
Sbjct: 61   MWMLLLPNDWFKVVIPDFTKVEIFIFFILFLLLLTVLTKNTFTFDEAGFVILSSAYIGYG 120

Query: 121  FQHLINAR---LSGIDKVFLALFIVWATDIGAYLIGRQFGRRKLLPTVSPNKTIEGSLGG 177
            F  L+ +R     G+  VF  LF++WATD GAY  GR FG+ KL P +SPNKTIEGS+GG
Sbjct: 121  FHFLLLSREIPEIGLPLVFFVLFVIWATDSGAYFAGRAFGKHKLWPHISPNKTIEGSIGG 180

Query: 178  IACAVLVSFIFMVIDRSVYAPHHFLTMLVLVALFSIFAQFGDLVESALKRHFGVKDSGKL 237
            I  AV++  +F  I   +++ +      +++VA  S+F Q GDLVESALKRH+ VKDSG +
Sbjct: 181  IILAVIIGSLFYWI-MPLFSSYGVALAVIVVA--SVFGQLGDLVESALKRHYAVKDSGTV 237

Query: 238  IPGHGGILDRFDSMIFVFPIMHLFGL 263
            +PGHGGILDRFDS+I+V PI+HL  L
Sbjct: 238  LPGHGGILDRFDSLIYVMPILHLLHL 263
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 204/264 (77%), Positives = 243/264 (91%)

Query: 1    MKERVIWGAVALAIFIPFLVMGGLPFQFLVGLLAMIGVSELLRMRRLEIFSFEGALAMIG 60
            MKERV+WG VA+AIF+PFL++G LPFQ  VG+LAMIGVSELL+M+RLE+FSFEG  AM+
Sbjct: 1    MKERVVWGGVAVAIFLPFLIIGNLPFQLFVGVLAMIGVSELLKMKRLEVFSFEGVFAMLA 60

Query: 61   AFVLTVPLDSYLSFLPVDASLSAYGIVIFMILAGTVLNSNSYSFEDAAFPIASSFYVGIG 120
            AFVL VP+D YL+FLP+DA+++ Y +++F ILAGTVLNS +YSF+DAAFPIA+SFYVGIG
Sbjct: 61   AFVLAVPMDHYLTFLPIDANVAFYSLMVFFILAGTVLNSRAYSFDDAAFPIATSFYVGIG 120

Query: 121  FQNLVSARMAGIDKVLLALFIVWATDIGAYMIGRQFGQRKLLPSVSPNKTIEGSLGGIAS 180
            FQ+L++AR++GIDKV LALFIVWATDIGAY+IGRQFG+RKLLP+VSPNKTIEGSLGGIA
Sbjct: 121  FQHLINARLSGIDKVFLALFIVWATDIGAYLIGRQFGRRKLLPTVSPNKTIEGSLGGIAC 180

Query: 181  AIVVAFFFMLFDKTVYAPHSFLVMLVLVAIFSIFGQFGDLVESSIKRHFGVKDSGKLIPG 240
            A++V+F FM+ D++VYAPH FL MLVLVA+FSIF QFGDLVES++KRHFGVKDSGKLIPG
Sbjct: 181  AVLVSFIFMVIDRSVYAPHHFLTMLVLVALFSIFAQFGDLVESALKRHFGVKDSGKLIPG 240

Query: 241  HGGILDRFDSMIFVFPIMHFFGLF 264
            HGGILDRFDSMIFVFPIMH FGLF
Sbjct: 241  HGGILDRFDSMIFVFPIMHLFGLF 264
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 749

A DNA sequence (GBSx0796) was identified in *S.agalactiae* <SEQ ID 2299> which encodes the amino acid sequence <SEQ ID 2300>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.09   Transmembrane    2 -  18 (   1 -  25)
   INTEGRAL    Likelihood = -9.39   Transmembrane  394 - 410 ( 390 - 415)
   INTEGRAL    Likelihood = -8.01   Transmembrane  181 - 197 ( 173 - 198)
   INTEGRAL    Likelihood = -2.97   Transmembrane  343 - 359 ( 342 - 360)

----- 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 GENPEPT database:

```
>GP:AAD47948 GB:AF152237 Eep [Enterococcus faecalis]
 Identities = 229/425 (53%), Positives = 298/425 (69%), Gaps = 9/425 (2%)
```

```
Query: 1   MLGILTFIIIFGVIVVVHEFGHFYFAKKSGILVREFAIGMGPKIFSHIDKEGTTYTIRIL 60
           M  I+TFII+FG++V+VHEFGHFYFAK++GILVREFAIGMGPKIF+H  K+GTTYTIR+L
Sbjct: 1   MKTIITFIIVFGILVLVHEFGHFYFAKRAGILVREFAIGMGPKIFAHRGKDGTTYTIRLL 60

Query: 61  PLGGYVRMAGWGDDKTEIKTGTPASLTLNKEGIVTRINLSGKQLDNTSLPINVTAYDLED 120
           P+GGYVRMAG G+D TEI  G P S+ LN  G V +IN S K     S+P+ V  +DLE
Sbjct: 61  PIGGYVRMAGMGEDMTEITPGMPLSVELNAVGNVVKINTSKKVQLPHSIPMEVVDFDLEK 120

Query: 121 KLTITGLV---LSETKTYSVDHDATIIEEDGTEIRIAPLDMQYQNASVWGRLITNFAGPM 177
           +L I G V     E   Y VDHDATIIE DGTE+RIAPLD+Q+Q+A +  R++TNFAGPM
Sbjct: 121 ELFIKGYVNGNEEEETVYKVDHDATIIESDGTEVRIAPLDVQFQSAKLSQRILTNFAGPM 180

Query: 178 NNFILGLVVFIALAFIQGGVQDLSTNQV-RVSENGPAASAGLKNNDRILQIGSHKVSNWE 236
           NNFILG ++F     F+QGGV DL+TNQ+ +V  NGPAA AGLK ND++L I + K+  +E
Sbjct: 181 NNFILGFILFTLAVFLQGGVTDLNTNQIGQVIPNGPAAEAGLKENDKVLSINNQKIKKYE 240

Query: 237 QLTAAVEKSTRHLEKKQKLALKIKSKEVVKTINVKPQKVDKSYI--IGIMPALKTSFKDK 294
             T  V+K+     EK      ++    KE   T+  + QKV+K  I  +G+ P +KT    K
Sbjct: 241 DFTTIVQKNP---EKPLTFVVERNGKEEQLTVTPEKQKVEKQTIGKVGVYPYMKTDLPSK 297

Query: 295 LLGGLKLAWESFFRILNELKKLIAHFSINKLGGPVALYQASSQAAKNGFVTVLNLMGLIS 354
           L+GG++     S  +I   L  L   FS+NKLGGPV +++ S +A+  G  TV+ LM ++S
Sbjct: 298 LMGGIQDTLNSTTQIFKALGSLFTGFSLNKLGGPVMMFKLSEEASNAGVSTVVFLMAMLS 357

Query: 355 INLGIMNLIPIPALDGGKIVMNILEAIRRKPLKQETETYITLAGVAVMLVLMIAVTWNDI 414
           +NLGI+NL+PIPALDGGKIV+NI+E +R KP+  E E  ITL G   ++VLM+ VTWNDI
Sbjct: 358 MNLGIINLLPIPALDGGKIVLNIIEGVRGKPISPEKEGIITLIGFGFVMVLMVLVTWNDI 417

Query: 415 MRAFF 419
            R FF
Sbjct: 418 QRFFF 422
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2301> which encodes the amino acid sequence <SEQ ID 2302>. Analysis of this protein sequence reveals the following:

```
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.41   Transmembrane    2 -  18 (   1 -  25)
   INTEGRAL    Likelihood = -9.77   Transmembrane  394 - 410 ( 390 - 415)
   INTEGRAL    Likelihood = -9.61   Transmembrane  180 - 196 ( 173 - 201)
   INTEGRAL    Likelihood = -2.66   Transmembrane  347 - 363 ( 343 - 363)

----- Final Results -----
              bacterial membrane --- Certainty=0.5564(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:

```
>GP:AAD47948 GB:AF152237 Eep [Enterococcus faecalis]
 Identities = 230/427 (53%), Positives = 298/427 (68%), Gaps = 13/427 (3%)

Query: 1   MLGIITFIIIFGILVIVHEFGHFYFAKKSGILVREFAIGMGPKIFSHVDQGGTLYTLRML 60
           M  IITFII+FGILV+VHEFGHFYFAK++GILVREFAIGMGPKIF+H  + GT YT+R+L
Sbjct: 1   MKTIITFIIVFGILVLVHEFGHFYFAKRAGILVREFAIGMGPKIFAHRGKDGTTYTIRLL 60

Query: 61  PLGGYVRMAGWGDDKTEIKTGTPASLTLNEQGFVKRINLSQSKLDPTSLPMHVTGYDLED 120
           P+GGYVRMAG G+D TEI  G P S+ LN  G V +IN S+     P S+PM V  +DLE
Sbjct: 61  PIGGYVRMAGMGEDMTEITPGMPLSVELNAVGNVVKINTSKKVQLPHSIPMEVVDFDLEK 120

Query: 121 QLSITGLV---LEETKTYKVAHDATIVEEDGTEIRIAPLDVQYQNASIGGRLITNFAGPM 177
           +L I G V     EE   YKV HDATI+E DGTE+RIAPLDVQ+Q+A +  R++TNFAGPM
Sbjct: 121 ELFIKGYVNGNEEEETVYKVDHDATIIESDGTEVRIAPLDVQFQSAKLSQRILTNFAGPM 180

Query: 178 NNFILGIVVFILLVFLQGGMPDFSSNHV-RVQENGAAAKAGLRDNDQIVAINGYKVTSWN 236
           NNFILG ++F L VFLQGG+ D ++N + +V  NG AA+AGL++ND++++ING K+  +
Sbjct: 181 NNFILGFILFTLAVFLQGGVTDLNTNQIGQVIPNGPAAEAGLKENDKVLSINNQKIKKYE 240
```

```
Query: 237 DLTEAVDLATRDLGPSQTIKVTYKSHQRLKTVAVKPQKH-AKTYTI---GVKASLKTGFK 292
           D T  V         P + +    + + + + + V P+K   +  TI   GV   +KT
Sbjct: 241 DFTTIV-----QKNPEKPLTFVVERNGKEEQLTVTPEKQKVEKQTIGKVGVYPYMKTDLP 295

Query: 293 DKLLGGLELAWSRAFTILNALKGLITGFSLNKLGGPVAMYDMSNQAAQNGLESVLSLMAM 352
            KL+GG++   +    I  AL  L TGFSLNKLGGPV M+ +S +A+  G+ +V+ LMAM
Sbjct: 296 SKLMGGIQDTLNSTTQIFKALGSLFTGFSLNKLGGPVMMFKLSEEASNAGVSTVVFLMAM 355

Query: 353 LSINLGIFNLIPIPALDGGKILMNIIEAIRRKPIKQETEAYITLAGVAIMVVLMIAVTWN 412
           LS+NLGI NL+PIPALDGGKI++NIIE +R KPI  E E  ITL G   ++VLM+ VTWN
Sbjct: 356 LSMNLGIINLLPIPALDGGKIVLNIIEGVRGKPISPEKEGIITLIGFGFVMVLMVLVTWN 415

Query: 413 DIMRVFF 419
           DI R FF
Sbjct: 416 DIQRFFF 422
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 306/419 (73%), Positives = 359/419 (85%)

Query: 1   MLGILTFIIIFGVIVVVHEFGHFYFAKKSGILVREFAIGMGPKIFSHIDKEGTTYTIRIL 60
           MLGI+TFIIIFG++V+VHEFGHFYFAKKSGILVREFAIGMGPKIFSH+D+ GT YT+R+L
Sbjct: 1   MLGIITFIIIFGILVIVHEFGHFYFAKKSGILVREFAIGMGPKIFSHVDQGGTLYTLRML 60

Query: 61  PLGGYVRMAGWGDDKTEIKTGTPASLTLNKEGIVTRINLSGKQLDNTSLPINVTAYDLED 120
           PLGGYVRMAGWGDDKTEIKTGTPASLTLN++G V RINLS  +LD TSLP++VT YDLED
Sbjct: 61  PLGGYVRMAGWGDDKTEIKTGTPASLTLNEQGFVKRINLSQSKLDPTSLPMHVTGYDLED 120

Query: 121 KLTITGLVLSETKTYSVDHDATIIEEDGTEIRIAPLDMQYQNASVWGRLITNFAGPMNNF 180
           +L+ITGLVL ETKTY V HDATI+EEDGTEIRIAPLD+QYQNAS+ GRLITNFAGPMNNF
Sbjct: 121 QLSITGLVLEETKTYKVAHDATIVEEDGTEIRIAPLDVQYQNASIGGRLITNFAGPMNNF 180

Query: 181 ILGLVVFIALAFIQGGVQDLSTNQVRVSENGPAASAGLKNNDRILQIGSHKVSNWEQLTA 240
           ILG+VVFI L F+QGG+ D S+N VRV ENG AA AGL++ND+I+ I  +KV++W  LT
Sbjct: 181 ILGIVVFILLVFLQGGMPDFSSNHVRVQENGAAAKAGLRDNDQIVAINGYKVTSWNDLTE 240

Query: 241 AVEKSTRHLEKKQKLALKIKSKEVVKTINVKPQKVDKSYIIGIMPALKTSFKDKLLGGLK 300
           AV+ +TR L   Q + +  KS + +KT+ VKPQK  K+Y IG+  +LKT FKDKLLGGL+
Sbjct: 241 AVDLATRDLGPSQTIKVTYKSHQRLKTVAVKPQKHAKTYTIGVKASLKTGFKDKLLGGLE 300

Query: 301 LAWESFFRILNELKKLIAHFSINKLGGPVALYQASSQAAKNGFVTVLNLMGLISINLGIM 360
           LAW   F ILN LK LI  FS+NKLGGPVA+Y  S+QAA+NG  +VL+LM ++SINLGI
Sbjct: 301 LAWSRAFTILNALKGLITGFSLNKLGGPVAMYDMSNQAAQNGLESVLSLMAMLSINLGIF 360

Query: 361 NLIPIPALDGGKIVMNILEAIRRKPLKQETETYITLAGVAVMLVLMIAVTWNDIMRAFF 419
           NLIPIPALDGGKI+MNI+EAIRRKP+KQETE YITLAGVA+M+VLMIAVTWNDIMR FF
Sbjct: 361 NLIPIPALDGGKILMNIIEAIRRKPIKQETEAYITLAGVAIMVVLMIAVTWNDIMRVFF 419
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 750

A DNA sequence (GBSx0797) was identified in *S.agalactiae* <SEQ ID 2303> which encodes the amino acid sequence <SEQ ID 2304>. This protein is predicted to be prolyl-tRNA synthetase (proS). Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane  473 - 489 ( 473 - 490)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

A related GBS nucleic acid sequence <SEQ ID 10181> which encodes amino acid sequence <SEQ ID 10182> was also identified.

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

```
>GP:CAB13530 GB:Z99112 prolyl-tRNA synthetase [Bacillus subtilis]
 Identities = 301/608 (49%), Positives = 410/608 (66%), Gaps = 52/608 (8%)

Query: 1   MKQSKMLIPTLREMPSDAQVISHALMVRAGYVRQVSAGIYAYLPLANRTIEKFKTIMRQE 60
           M+QS  LIPTLRE+P+DA+  SH L++RAG++RQ ++G+Y+Y+PLA + I+  + I+R+E
Sbjct: 1   MRQSLTLIPTLREVPADAEAKSHQLLLRAGFIRQNTSGVYSYMPLAYKVIQNIQQIVREE 60

Query: 61  FEKIGAVEMLAPALLTADLWRESGRYETYGEDLYKLKNRDQSDFILGPTHEETFTTLVRD 120
            EKI AVEML PAL  A+ W+ESGR+ TYG +L +LK+R   +F LG THEE  T+LVRD
Sbjct: 61  MEKIDAVEMLMPALQQAETWQESGRWYTYGPELMRLKDRHGREFALGATHEEVITSLVRD 120

Query: 121 AVKSYKQLPLNLYQIQSKYRDEKRPRNGLLRTREFIMKDGYSFHKDYEDLDVTYEDYRKA 180
            VKSYK+LPL LYQIQSK+RDEKRPR GLLR REFIMKD YSFH   E LD TY+   +A
Sbjct: 121 EVKSYKRLPLTLYQIQSKFRDEKRPRFGLLRGREFIMKDAYSFHASAESLDETYQKMYEA 180

Query: 181 YEAIFTRAGLDFKGIIGDGGAMGGKDSQEFMAVTPNRTDLNRWLVLDKTIPSIDDIPEDV 240
           Y  IF R G++ + +I D GAMGGKD+ EFMA++
Sbjct: 181 YSNIFARCGINVRPVIADSGAMGGKDTHEFMALS------------------------- 214

Query: 241 LEEIKVELSAWLVSGEDTIAYSTESSYAANLEMATNEYKPSTKAATFEEVTKVETPNCKS 300
                      GEDTIAYS ES YAAN+EMA   ++        + + KV TPN K+
Sbjct: 215 ------------AIGEDTIAYSDESQYAANIEMAEVLHQEVPSDEEPKALEKVHTPNVKT 262

Query: 301 IDEVAGFLSIDENQTIKTLLFIADEQPVVALLVGNDQVNDVKLKNYLAADFLEPASEEQA 360
           I+E+  FL +     IK++LF AD++ V+ L+ G+ +VND+K+KN L A+ +E A+ E+
Sbjct: 263 IEELTAFLQVSAEACIKSVLFKADDRFVLVLVRGDHEVNDIKVKNLLHAEVVELATHEEV 322

Query: 361 KEIFGAGFGSLGPVNLPDSVKIIADRKVQDLANAVSGANQDGYHFTGVNPERDFTA-EYV 419
            + G   G +GPV +   V++ AD+ V+ + NAV+GAN+  +H+  VN  RD    E+
Sbjct: 323 IQQLGTEPGFVGPVGIHQDVEVYADQAVKAMVNAVAGANEGDHHYKNVNVNRDAQIKEFA 382

Query: 420 DIREVKEGEISPDGKGTLKFARGIEIGHIFKLGTRYSDSMGANILDENGRSNPIVMGCYG 479
           D+R +KEG+ SPDGKGT++FA GIE+G +FKLGTRYS++M A  LDENGR+ P++MGCYG
Sbjct: 383 DLRFIKEGDPSPDGKGTIRFAEGIEVGQVFKLGTRYSEAMNATYLDENGRAQPMLMGCYG 442

Query: 480 IGVSRILSAVIEQHARLFVNKTPKGAYRFAWGINFPEELAPFDVHLITVNVKDQESQDLT 539
           IGVSR LSA+ EQH                 G+ +P+ +AP+D+H++ +N+K    ++L
Sbjct: 443 IGVSRTLSAIAEQH-------------HDEKGLIWPKSVAPYDLHILALNMKNDGQRELA 489

Query: 540 EKIEADLMLKGYEVLTDDRNERVGSKFSDSDLIGLPIRVTVGKKASEGIVEVKIKASGDT 599
           EK+ ADL  +GYEVL DDR ER G KF+DSDLIGLPIR+TVGK+A EGIVEVKI+ +G++
Sbjct: 490 EKLYADLKAEGYEVLYDDRAERAGVKFADSDLIGLPIRITVGKRADEGIVEVKIRQTGES 549

Query: 600 IEVHADNL 607
            E+  D L
Sbjct: 550 TEISVDEL 557
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2305> which encodes the amino acid sequence <SEQ ID 2306>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane  473 - 489 ( 473 - 490)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 535/617 (86%), Positives = 584/617 (93%)

Query: 1   MKQSKMLIPTLREMPSDAQVISHALMVRAGYVRQVSAGIYAYLPLANRTIEKFKTIMRQE 60
           MKQSK+LIPTLREMPSDAQVISHALMVRAGYVRQVSAGIYAYLPLANRTIEKFKTIMR+E
Sbjct: 1   MKQSKLLIPTLREMPSDAQVISHALMVRAGYVRQVSAGIYAYLPLANRTIEKFKTIMREE 60

Query: 61  FEKIGAVEMLAPALLTADLWRESGRYETYGEDLYKLKNRDQSDFILGPTHEETFTTLVRD 120
           FEKIGAVEMLAPALLTADLWRESGRYETYGEDLYKLKNRD SDFILGPTHEETFTTLVRD
Sbjct: 61  FEKIGAVEMLAPALLTADLWRESGRYETYGEDLYKLKNRDNSDFILGPTHEETFTTLVRD 120

Query: 121 AVKSYKQLPLNLYQIQSKYRDEKRPRNGLLRTREFIMKDGYSFHKDYEDLDVTYEDYRKA 180
           AVKSYKQLPLNLYQIQSKYRDEKRPRNGLLRTREFIMKDGYSFH +YEDLDVTYEDYR+A
Sbjct: 121 AVKSYKQLPLNLYQIQSKYRDEKRPRNGLLRTREFIMKDGYSFHHNYEDLDVTYEDYRQA 180

Query: 181 YEAIFTRAGLDFKGIIGDGGAMGGKDSQEFMAVTPNRTDLNRWLVLDKTIPSIDDIPEDV 240
           YEAIFTRAGLDFKGIIGDGGAMGGKDSQEFMA+TP RTDL+RW+VLDK+I S+DDIP++V
Sbjct: 181 YEAIFTRAGLDFKGIIGDGGAMGGKDSQEFMAITPARTDLDRWVVLDKSIASMDDIPKEV 240

Query: 241 LEEIKVELSAWLVSGEDTIAYSTESSYAANLEMATNEYKPSTKAATFEEVTKVETPNCKS 300
           LE+IK EL+AW++SGEDTIAYSTESSYAANLEMATNEYKPS+K A  + + +VETP+CK+
Sbjct: 241 LEDIKAELAAWMISGEDTIAYSTESSYAANLEMATNEYKPSSKVAAEDALAEVETPHCKT 300

Query: 301 IDEVAGFLSIDENQTIKTLLFIADEQPVVALLVGNDQVNDVKLKNYLAADFLEPASEEQA 360
           IDEVA FLS+DE QTIKTLLF+AD +PVVALLVGND +N VKLKNYLAADFLEPASEE+A
Sbjct: 301 IDEVAAFLSVDETQTIKTLLFVADNEPVVALLVGNDHINTVKLKNYLAADFLEPASEEEA 360

Query: 361 KEIFGAGFGSLGPVNLPDSVKIIADRKVQDLANAVSGANQDGYHFTGVNPERDFTAEYVD 420
           +  FGAGFGSLGPVNL    +I+ADRKVQ+L NAV+GAN+DG+H TGVNP RDF AEYVD
Sbjct: 361 RAFFGAGFGSLGPVNLAQGSRIVADRKVQNLTNAVAGANKDGFHMTGVNPGRDFQAEYVD 420

Query: 421 IREVKEGEISPDGKGTLKFARGIEIGHIFKLGTRYSDSMGANILDENGRSNPIVMGCYGI 480
           IREVKEGE+SPDG G L+FARGIE+GHIFKLGTRYSDSMGA ILDENGR+ PIVMGCYGI
Sbjct: 421 IREVKEGEMSPDGHGVLQFARGIEVGHIFKLGTRYSDSMGATILDENGRTVPIVMGCYGI 480

Query: 481 GVSRILSAVIEQHARLFVNKTPKGAYRFAWGINFPEELAPFDVHLITVNVKDQESQDLTE 540
           GVSRILSAVIEQHARLFVNKTPKG YR+AWGINFP+ELAPFDVHLITVNVKDQ +QDLT
Sbjct: 481 GVSRILSAVIEQHARLFVNKTPKGDYRYAWGINFPKELAPFDVHLITVNVKDQVAQDLTA 540

Query: 541 KIEADLMLKGYEVLTDDRNERVGSKFSDSDLIGLPIRVTVGKKASEGIVEVKIKASGDTI 600
           K+EADLM KGY+VLTDDRNERVGSKFSDSDLIGLPIRVTVGKKA+EGIVE+KIKA+GD+I
Sbjct: 541 KLEADLMAKGYDVLTDDRNERVGSKFSDSDLIGLPIRVTVGKKAAEGIVEIKIKATGDSI 600

Query: 601 EVHADNLIETLEILTKK 617
           EV+A+NLIETLEILTK+
Sbjct: 601 EVNAENLIETLEILTKE 617
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 751

A DNA sequence (GBSx0798) was identified in *S.agalactiae* <SEQ ID 2307> which encodes the amino acid sequence <SEQ ID 2308>. This protein is predicted to be peptidoglycan hydrolase (flgJ). Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.86   Transmembrane    9 -  25 (   9 -  25)

----- 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>
```

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

```
>GP:CAB94815 GB:AJ245582 peptidoglycan hydrolase [Streptococcus thermophilus]
 Identities = 101/201 (50%), Positives = 122/201 (60%), Gaps = 9/201 (4%)

Query: 2   KSRKKDKLVLRLTT-----TLLVFGL----GGVWFYNYKNDNVEPTVTSASDQTTTFIQT 52
           KS+KK K VL          +L+  GL    G +   N+     +E  +T    + T  FI
Sbjct: 16  KSKKKKKSVLLFPKFFQKWSLIFIGLFSLLGLLASLNFPRLTMEKNMTPTDETTVAFIAE 75

Query: 53  ISPTAIEISKTYDLYASVLLAQAILESSSGQSDLSKAPNYNLFGIKGEYKGKSVQMPTLE 112
           I  T+  ++   DLYASV++AQAILES SGQS LS+ P YN FGIKGEY G+SV +PT E
Sbjct: 76  IGETSRYLAARNDLYASVMIAQAILESDSGQSQLSQKPLYNFFGIKGEYNGQSVTLPTWE 135

Query: 113 DDGKGNMTQIQAPFRAYPNYSASLYDYAELVSSQKYASVWKSNTSSYKDATAALTGLYAT 172
           DDGKGN   I A FR+Y +   SL DY E +    Y  V +S T SYKDATAALTG+YAT
Sbjct: 136 DDGKGNPYHIDAAFRSYGSVENSLQDYVEFLEGSYYVGVHRSKTRSYKDATAALTGVYAT 195

Query: 173 DTAYASKLNQIIETYSLDAYD 193
           DT Y  KLN IIE Y L  YD
Sbjct: 196 DTTYGDKLNSIIEQYQLTIYD 216
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2309> which encodes the amino acid sequence <SEQ ID 2310>. 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>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB94815 GB:AJ245582 peptidoglycan hydrolase [Streptococcus thermophilus]
 Identities = 103/189 (54%), Positives = 126/189 (66%), Gaps = 4/189 (2%)

Query: 4   KKGKLVLISLFVLAACLGAYSAMRQSHKTSNVSAETIASSSTRHFIDEIGPTASTIGQER 63
           +K  L+ I LF L   L + +  R + + +     T    +T  FI EIG T+  +
Sbjct: 32  QKWSLIFIGLFSLLGLLASLNFPRLTMEKNM----TPTDETTVAFIAEIGETSRYLAARN 87

Query: 64  DLYASVMIAQAILESSNGKSSLSQAPYYNFFGIKGAYNGSSVTMSTWEDDGNGNTYTIDQ 123
           DLYASVMIAQAILES +G+S LSQ P YNFFGIKG YNG SVT+ TWEDDG GN Y ID
Sbjct: 88  DLYASVMIAQAILESDSGQSQLSQKPLYNFFGIKGEYNGQSVTLPTWEDDGKGNPYHIDA 147

Query: 124 AFRAYPSIADSLNDYADLLSSSTYIGARKSNTLSYQDATAALTGLYATDTSYNLKLNNII 183
           AFR+Y S+ +SL DY + L  S Y+G  +S T SY+DATAALTG+YATDT+Y  KLN+II
Sbjct: 148 AFRSYGSVENSLQDYVEFLEGSYYVGVHRSKTRSYKDATAALTGVYATDTTYGDKLNSII 207

Query: 184 ATYGLTAYD 192
             Y LT YD
Sbjct: 208 EQYQLTIYD 216
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 108/192 (56%), Positives = 124/192 (64%), Gaps = 2/192 (1%)

Query: 3   SRKKDKLVL-RLTTTLLVFGLGGVWFYNYKNDNVEPTVTSASDQTTTFIQTISPTAIEIS 61
           ++KK KLVL  L       G       ++K  NV    T AS  T  FI  I PTA  I
Sbjct: 2   TKKKGKLVLISLFVLAACLGAYSAMRQSHKTSNVSAE-TIASSSTRHFIDEIGPTASTIG 60

Query: 62  KTYDLYASVLLAQAILESSSGQSDLSKAPNYNLFGIKGEYKGKSVQMPTLEDDGKGNMTQ 121
           +  DLYASV++AQAILESS+G+S LS+AP YN FGIKG Y G SV M T EDDG GN
Sbjct: 61  QERDLYASVMIAQAILESSNGKSSLSQAPYYNFFGIKGAYNGSSVTMSTWEDDGNGNTYT 120

Query: 122 IQAPFRAYPNYSASLYDYAELVSSQKYASVWKSNTSSYKDATAALTGLYATDTAYASKLN 181
```

```
            I   FRAYP+ + SL DYA+L+SS  Y    KSNT SY+DATAALTGLYATDT+Y  KLN
Sbjct: 121 IDQAFRAYPSIADSLNDYADLLSSSTYIGARKSNTLSYQDATAALTGLYATDTSYNLKLN 180

Query: 182 QIIETYSLDAYD 193
            II TY L AYD
Sbjct: 181 NIIATYGLTAYD 192
```

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9073> which encodes the amino acid sequence <SEQ ID 9074>. Analysis of this protein sequence reveals the following:

```
      Possible site: 58

>>> 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>
```

An alignment of the GAS and GBS sequences follows:

```
 Score =  130 bits (323), Expect = 2e-32
 Identities = 68/169 (40%), Positives = 96/169 (56%), Gaps = 3/169 (1%)

Query: 30  MWTLKLGNQRLAPY---ADHETLTFVRKISHAAQSVAQKKQLYSSVMMAQAILESNNGKS 86
           +W     N  + P    A  +T TF++ IS  A  +++    LY+SV++AQAILES++G+S
Sbjct: 25  VWFYNYKNDNVEPTVTSASDQTTTFIQTISPTAIEISKTYDLYASVLLAQAILESSSGQS 84

Query: 87  QLSQKPYYNFFGIKGSYKERSVIFPTLEDDGQGNLYQIDAAFRSYGSLTACFLDYARVLN 146
            LS+ P YN FGIKG YK +SV  PTLEDDG+GN+ QI A FR+Y + +A   DYA +++
Sbjct: 85  DLSKAPNYNLFGIKGEYKGKSVQMPTLEDDGKGNMTQIQAPFRAYPNYSASLYDYAELVS 144

Query: 147 DPLYDKTHKKFWSHYQXXXXXXXXXXXXXXXXXXKLNELIEWYQLTNFD 195
              Y    K   S Y+                   KLN++IE Y L  +D
Sbjct: 145 SQKYASVWKSNTSSYKDATAALTGLYATDTAYASKLNQIIETYSLDAYD 193
```

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9075> which encodes the amino acid sequence <SEQ ID 9076>. An alignment of the GAS and GBS sequences follows:

```
 Score = 69.1 bits (166), Expect = 1e-13
 Identities = 52/151 (34%), Positives = 79/151 (51%), Gaps = 10/151 (6%)

Query: 2   TFLDKIKQGCLDGWAKYKILPSLTAAQAILESGWGKH----APHNALFGIKADSSWTGKS 57
           TF+  I    ++    Y +  S+  AQAILES  G+     AP+  LFGIK +  + GKS
Sbjct: 48  TFIQTISPTAIEISKTYDLYASVLLAQAILESSSGQSDLSKAPNYNLFGIKGE--YKGKS 105

Query: 58  FDTKTQEEYQAGVVTDIVDRFRAYDSWDESIADHGQFLVDNPRYEAV--IGETDYKKACY 115
              T E+   G +T I   FRAY ++  S+ D+ + LV + +Y +V     + YK A
Sbjct: 106 VQMPTLEDDGKGNMTQIQAPFRAYPNYSASLYDYAE-LVSSQKYASVWKSNTSSYKDATA 164

Query: 116 AIKAAGYATASSYVELLIQLIEENDLQSWDR 146
           A+    YAT ++Y   L Q+IE   L ++D+
Sbjct: 165 ALTGL-YATDTAYASKLNQIIETYSLDAYDK 194
```

SEQ ID 2308 (GBS275) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 4; MW 22.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 4; MW 47.5kDa).

The GBS275-GST fusion product was purified (Figure 208, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 276), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 752

A DNA sequence (GBSx0799) was identified in *S.agalactiae* <SEQ ID 2311> which encodes the amino acid sequence <SEQ ID 2312>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16   Transmembrane  876 - 892 ( 876 - 892)

----- 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>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2313> which encodes the amino acid sequence <SEQ ID 2314>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16   Transmembrane  873 - 889 ( 873 - 889)

----- 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>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB94815 GB:AJ245582 peptidoglycan hydrolase [Streptococcus thermophilus]
 Identities = 96/202 (47%), Positives = 127/202 (62%), Gaps = 10/202 (4%)

Query: 4   KKRRRRAKSSV---------NRLVLGLV-LLNLIVSMWTLKLGNQRLAPYADHETLTFVR 53
           KK +++ KS +          + + +GL  LL L+ S+    +L  ++      D  T+ F+
Sbjct: 15  KKSKKKKKSVLLFPKFFQKWSLIFIGLFSLLGLLASLNFPRLTMEKNMTPTDETTVAFIA 74

Query: 54  KISHAAQSVAQKKQLYSSVMMAQAILESNNGKSQLSQKPYYNFFGIKGSYKERSVIFPTL 113
           +I   ++ +A +  LY+SVM+AQAILES++G+SQLSQKP YNFFGIKG Y  +SV  PT
Sbjct: 75  EIGETSRYLAARNDLYASVMIAQAILESDSGQSQLSQKPLYNFFGIKGEYNGQSVTLPTW 134

Query: 114 EDDGQGNLYQIDAAFRSYGSLTACFLDYARVLNDPLYDKTHKKFWSHYQDATATLTGTYA 173
           EDDG+GN Y IDAAFRSYGS+     DY   L    Y   H+     Y+DATA LTG YA
Sbjct: 135 EDDGKGNPYHIDAAFRSYGSVENSLQDYVEFLEGSYYVGVHRSKTRSYKDATAALTGVYA 194

Query: 174 TDTTYHTKLNELIEWYQLTNFD 195
           TDTTY  KLN +IE YQLT +D
Sbjct: 195 TDTTYGDKLNSIIEQYQLTIYD 216
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 1244/1468 (84%), Positives = 1351/1468 (91%), Gaps = 3/1468 (0%)

Query: 1    MSELFKKLMDQIEMPLEIKNSSVFSSADIIEVKVHSLSRLWEFHFSFPELLPIEVYRELQ 60
           MS+LF KLMDQIEMPL+++ SS FSSADIIEVKVHS+SRLWEFHF+F  +LPI  YREL
Sbjct: 1    MSDLFAKLMDQIEMPLDMRRSSAFSSADIIEVKVHSVSRLWEFHFAFAAVLPIATYRELH 60

Query: 61   TRLVNSFEKADIKATFDIRAETIDFSDDLLQDYYQQAFCEPLCNSASFKSSFSQLKVHYN 120
            RL+ +FE ADIK TFDI+A  +D+SDDLLQ YYQ+AF     CNSASFKSSFS+LKV Y
Sbjct: 61   DRLIRTFEAADIKVTFDIQAAQVDYSDDLLQAYYQEAFEHAPCNSASFKSSFSKLKVTYE 120

Query: 121  GSQMIISAPQFVNNNHFRQNHLPRLEQQFSLFGFGKLAIDMVSDEQMTQDLKSSFETNRE 180
             ++II+AP FVNN+HFR NHLP L +Q   FGFG L IDMVSD++MT+ L  +F ++R+
Sbjct: 121  DDKLIIAAPGFVNNDHFRNNHLPNLVKQLEAFGFGILTIDMVSDQEMTEHLTKNFVSSRQ 180
```

```
Query: 181  QLLEKANQEAMQALEAQKSLEDSAPPSEEVTPTQNYDFKERIKQRQAGFEKAEITPMIEV 240
             L++KA Q+    LEAQKSLE   PP EE TP   +D+KER +RQAGFEKA ITPMIE+
Sbjct: 181  ALVKKAVQDN---LEAQKSLEAMMPPVEEATPAPKFDYKERAAKRQAGFEKATITPMIEI 237

Query: 241  TTEENRIVFEGMVFSVERKTTRTGRHIINFKMTDYTSSFAMQKWAKDDEELKKYDMISKG 300
             TEENRIVFEGMVF VERKTTRTGRHIINFKMTDYTSSFA+QKWAKDDEEL+K+DMI+KG
Sbjct: 238  ETEENRIVFEGMVFDVERKTTRTGRHIINFKMTDYTSSFALQKWAKDDEELRKFDMIAKG 297

Query: 301  SWLRVRGNIENNNFTKSLTMNVQDIKEIVHHERKDLMPADQKRVEFHAHTNMSTMDALPT 360
             +WLRV+GNIE N FTKSLTMNVQ +KEIV HERKDLMP  QKRVE HAHTNMSTMDALPT
Sbjct: 298  AWLRVQGNIETNPFTKSLTMNVQQVKEIVRHERKDLMPEGQKRVELHAHTNMSTMDALPT 357

Query: 361  VESLIDTAAKWGHPAIAITDHANVQSFPHGYHRAKKAGIKAIFGLEANIVEDKVPISYNE 420
             VESLIDTAAKWGH AIAITDHANVQSFPHGYHRA+KAGIKAIFGLEANIVEDKVPISY
Sbjct: 358  VESLIDTAAKWGHKAIAITDHANVQSFPHGYHRARKAGIKAIFGLEANIVEDKVPISYEP 417

Query: 421  VDMNLHEATYVVFDVETTGLSAANNDLIQIAASKMFKGNIIEQFDEFIDPGHPLSAFTTE 480
             VDM+LHEATYVVFDVETTGLSA NNDLIQIAASKMFKGNI+EQFDEFIDPGHPLSAFTTE
Sbjct: 418  VDMDLHEATYVVFDVETTGLSAMNNDLIQIAASKMFKGNIVEQFDEFIDPGHPLSAFTTE 477

Query: 481  LTGITDNHVRGSKPILQVLQEFQNFCQGTVLVAHNATFDVGFMNANYERHNLPLITQPVI 540
             LTGITD H++G+KP++ VL+ FQ+FC+ ++LVAHNA+FDVGFMNANYERH+LP ITQPVI
Sbjct: 478  LTGITDKHLQGAKPLVTVLKAFQDFCKDSILVAHNASFDVGFMNANYERHDLPKITQPVI 537

Query: 541  DTLEFARNLYPEYKRHGLGPLTKRFQVALEHHHMANYDAEATGRLLFIFLKEARENRDVT 600
             DTLEFARNLYPEYKRHGLGPLTKRFQV+L+HHHMANYDAEATGRLLFIFLK+ARE   +
Sbjct: 538  DTLEFARNLYPEYKRHGLGPLTKRFQVSLDHHHMANYDAEATGRLLFIFLKDAREKHGIK 597

Query: 601  NLMELNTKLVAEDSYKKARIKHATIYVQNQVGLKNIFKLVSLSNVKYFEGVARIPRSVLD 660
             NL++LNT LVAEDSYKKARIKHATIYVQNQVGLKN+FKLVSLSN+KYFEGV RIPR+VLD
Sbjct: 598  NLLQLNTDLVAEDSYKKARIKHATIYVQNQVGLKNMFKLVSLSNIKYFEGVPRIPRTVLD 657

Query: 661  AHREGLLLGTACSDGEVFDALLSNGIDAAVTLAKYYDFIEVMPPAIYRPLVVRDLIKDEV 720
             AHREGLLLGTACSDGEVFDA+L+ GIDAAV LA+YYDFIE+MPPAIY+PLVVR+LIKD+
Sbjct: 658  AHREGLLLGTACSDGEVFDAVLTKGIDAAVDLARYYDFIEIMPPAIYQPLVVRELIKDQA 717

Query: 721  GIQQIIRDLIEVGRRLDKPVLATGNVHYIEPEDEIYREIIVRSLGQGAMINRTIGRGEDA 780
             GI+Q+IRDLIEVG+R  KPVLATGNVHY+EPE+EIYREIIVRSLGQGAMINRTIGRGE A
Sbjct: 718  GIEQVIRDLIEVGKRAKKPVLATGNVHYLEPEEEIYREIIVRSLGQGAMINRTIGRGEGA 777

Query: 781  QPAPLPKAHFRTTNEMLDEFAFLGKDLAYEIVVTNTNTFADRFEDVEVVKGDLYTPFVDR 840
             QPAPLPKAHFRTTNEMLDEFAFLGKDLAY++VV NT  FADR E+VEVVKGDLYTP++D+
Sbjct: 778  QPAPLPKAHFRTTNEMLDEFAFLGKDLAYQVVVQNTQDFADRIEEVEVVKGDLYTPYIDK 837

Query: 841  AEERVAELTYAKAFEIYGNPLPDIIDLRIEKELASILGNGFAVIYLASQMLVQRSNERGY 900
             AEE VAELTY KAFEIYGNPLPDIIDLRIEKEL SILGNGFAVIYLASQMLV RSNERGY
Sbjct: 838  AEETVAELTYQKAFEIYGNPLPDIIDLRIEKELTSILGNGFAVIYLASQMLVNRSNERGY 897

Query: 901  LVGSRGSVGSSFVATMIGITEVNPMPPHYVCPNCQHSEFITDGSCGSGYDLPNKNCPKCG 960
             LVGSRGSVGSSFVATMIGITEVNPMPPHYVCP+CQHSEFITDGS GSGYDLPNK CPKCG
Sbjct: 898  LVGSRGSVGSSFVATMIGITEVNPMPPHYVCPSCQHSEFITDGSVGSGYDLPNKPCPKCG 957

Query: 961  TLYKKDGQDIPFETFLGFDGDKVPDIDLNFSGDDQPSAHLDVRDIFGEEYAFRAGTVGTV 1020
             T Y+KDGQDIPFETFLGFDGDKVPDIDLNFSGDDQPSAHLDVRDIFG+EYAFRAGTVGTV
Sbjct: 958  TPYQKDGQDIPFETFLGFDGDKVPDIDLNFSGDDQPSAHLDVRDIFGDEYAFRAGTVGTV 1017

Query: 1021 AEKTAFGFVKGYERDYNKFYNDAEVERLATGAAGVKRSTGQHPGGIVVIPNYMDVYDFTP 1080
             AEKTA+GFVKGYERDY KFY DAEV+RLA GAAGVKR+TGQHPGGIVVIPNYMDVYDFTP
Sbjct: 1018 AEKTAYGFVKGYERDYGKFYRDAEVDRLAAGAAGVKRTTGQHPGGIVVIPNYMDVYDFTP 1077

Query: 1081 VQYPADDMTAAWQTTHFNFHDIDENVLKLDILGHDDPTMIRKLQDLSGIDPSNILPDDPD 1140
             VQYPADD+TA+WQTTHFNFHDIDENVLKLDILGHDDPTMIRKLQDLSGIDP I  DDP
Sbjct: 1078 VQYPADDVTASWQTTHFNFHDIDENVLKLDILGHDDPTMIRKLQDLSGIDPITIPADDPG 1137

Query: 1141 VMKLFSGTEVLGVTEEQIGTPTGMLGIPEFGTNFVRGMVNETHPTTFAELLQLSGLSHGT 1200
             VM LFSGTEVLGVT EQIGTPTGMLGIPEFGTNFVRGMVNETHPTTFAELLQLSGLSHGT
Sbjct: 1138 VMALFSGTEVLGVTPEQIGTPTGMLGIPEFGTNFVRGMVNETHPTTFAELLQLSGLSHGT 1197
```

```
Query: 1201 DVWLGNAQDLIKEGIATLSTVIGCRDDIMVYLMHAGLQPKMAFTIMERVRKGLWLKISED 1260
            DVWLGNAQDLIKEGIATL TVIGCRDDIMVYLMHAGL+PKMAFTIMERVRKGLWLKISE+
Sbjct: 1198 DVWLGNAQDLIKEGIATLKTVIGCRDDIMVYLMHAGLEPKMAFTIMERVRKGLWLKISEE 1257

Query: 1261 ERNGYIQAMRDNNVPDWYIESCGKIKYMFPKAHAAAAYVLMALRVAYFKVHYPIFYYCAYF 1320
            ERNGYI AMR+NNVPDWYIESCGKIKYMFPKAHAAAAYVLMALRVAYFKVH+PI YYCAYF
Sbjct: 1258 ERNGYIDAMRENNVPDWYIESCGKIKYMFPKAHAAAAYVLMALRVAYFKVHHPIMYYCAYF 1317

Query: 1321 SIRAKAFELRTMSAGLDAVKARMKDITEKRQRNEATNVENDLFTTLELVNEMLERGFKFG 1380
            SIRAKAFEL+TMS GLDAVKARM+DIT KR+ NEATNVENDLFTTLE+VNEMLERGFKFG
Sbjct: 1318 SIRAKAFELKTMSGGLDAVKARMEDITIKRKNNEATNVENDLFTTLEIVNEMLERGFKFG 1377

Query: 1381 KLDLYRSHATDFIIEEDTLIPPFVAMEGLGENVAKQIVRAREDGEFLSKTELRKRGGVSS 1440
            KLDLY+S A +F I+ DTLIPPF+A+EGLGENVAKQIV+AR++GEFLSK ELRKRGG SS
Sbjct: 1378 KLDLYKSDAIEFQIKGDTLIPPFIALEGLGENVAKQIVKARQEGEFLSKMELRKRGGASS 1437

Query: 1441 TLVEKFDEMGILGNLPEDNQLSLFDDFF 1468
            TLVEK DEMGILGN+PEDNQLSLFDDFF
Sbjct: 1438 TLVEKMDEMGILGNMPEDNQLSLFDDFF 1465
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 753

A DNA sequence (GBSx0800) was identified in *S.agalactiae* <SEQ ID 2315> which encodes the amino acid sequence <SEQ ID 2316>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1505(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10179> which encodes amino acid sequence <SEQ ID 10180> was also identified.

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

```
>GP:CAB13207 GB:Z99111 similar to transcriptional regulator (MarR
          family) [Bacillus subtilis]
 Identities = 49/124 (39%), Positives = 73/124 (58%)

Query: 18  VMRKAFRTIDGKVSESFKEFELTPTQFAVLDVLYAKGTMKIGELIENMLATSGNMTVVIK 77
            V  +AF+++         KE    PT+FAVL++LY +G  K+ ++   +L  SGN+T VI
Sbjct: 20  VFARAFKSVSEHSIRDSKEHGFNPTEFAVLELLYTRGPQKLQQIGSRLLLVSGNVTYVID 79

Query: 78  NMEKKGWVLRHSCPNDKRAFLVSLTTEGEEVIKKALPEHIKRVEDAFSVLTETEQEDLIN 137
             +E+ G+++R   P DKR+    LT +G E + K  P H  R+  AFS L+  EQ+ LI
Sbjct: 80  KLERNGFLVREQDPKDKRSVYAHLTDKGNEYLDKIYPIHALRIARAFSGLSPDEQDQLIV 139

Query: 138 LLKK 141
           LLKK
Sbjct: 140 LLKK 143
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2317> which encodes the amino acid sequence <SEQ ID 2318>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
             bacterial cytoplasm --- Certainty=0.0537(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 80/145 (55%), Positives = 111/145 (76%), Gaps = 1/145 (0%)

Query: 2   GDEMGNF-KNSAVKSMVVMRKAFRTIDGKVSESFKEFELTPTQFAVLDVLYAKGTMKIGE 60
           G++M +  KN+A+K+MVV RKA RT+D   ++ FK+ +LT TQF+VL+VLY KG M+I
Sbjct: 8   GNQMSHLDKNTALKAMVVFRKAQRTLDAFGADIFKKADLTATQFSVLEVLYTKGCMRINH 67

Query: 61  LIENMLATSGNMTVVIKNMEKKGWVLRHSCPNDKRAFLVSLTTEGEEVIKKALPEHIKRV 120
           LI+++LATSGNMTVV+ NME+ GW+ +     DKRA++V+LT +G  +I+  LP+H+ RV
Sbjct: 68  LIDSLLATSGNMTVVLNNMERNGWISKCKDKTDKRAYVVTLTDKGTRLIEAVLPKHVARV 127

Query: 121 EDAFSVLTETEQEDLINLLKKFKTL 145
           E+AF+VLTE EQ  LI LLKKFK L
Sbjct: 128 EEAFAVLTEKEQLCLIELLKKFKQL 152
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 754

A DNA sequence (GBSx0801) was identified in *S.agalactiae* <SEQ ID 2319> which encodes the amino acid sequence <SEQ ID 2320>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3742(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 GENPEPT database.

```
>GP:AAG05963 GB:AE004686 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 115/203 (56%), Positives = 143/203 (69%), Gaps = 7/203 (3%)

Query: 2   SFLEELKNRRSIYALGRNTEVSDEKIVEIIKEAVRQSPSAFNSQTSRVVILLNDEVTKFW 61
           +FL  +KNRR+IYAL +   VS EKIVE++KEAV  SPSAFNSQ+SRVV+L   E  +FW
Sbjct: 4   AFLSSIKNRRTIYALDKQLPVSQEKIVELVKEAVSHSPSAFNSQSSRVVVLFGAEHEQFW 63

Query: 62  DELVANDLVETMKVQGAPETAIAGTKEKLASFGASKGTVLFFEDQDVVKSLQEQFVLYAD 121
           +  +A D  E  K+   P  A A T+ KL SF A  GTVLFFEDQ VV+ LQEQF LYAD
Sbjct: 64  N--IAKD--ELKKI--VPADAFAATETKLNSFAAGAGTVLFFEDQTVVRQLQEQFALYAD 117

Query: 122 NFPVWSEQSTGIASVNTWTALSAELGLGGNLQHYNPVIDASVQAVYGVPASWKLRGQLNF 181
           NFPVWSEQ++G+A    WTAL AE  +G +LQHYNP++DA     + +P SWKLR Q+ F
Sbjct: 118 NFPVWSEQASGMAQFAVWTAL-AEHKVGASLQHYNPLVDAQTHKTWNLPESWKLRAQMPF 176

Query: 182 GSIEAETGEKEFMNDDDRFKVIG 204
           G+I A  GEK F+ + +RFKV G
Sbjct: 177 GAIAAPAGEKAFIAESERFKVFG 199
```

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

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

Example 755

A DNA sequence (GBSx0802) was identified in *S.agalactiae* <SEQ ID 2321> which encodes the amino acid sequence <SEQ ID 2322>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2730(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 GENPEPT database.

```
>GP:CAB62846 GB:AL035475 hypothetical protein [Plasmodium falciparum]
            (ver 2)
 Identities = 112/529 (21%), Positives = 217/529 (40%), Gaps = 67/529 (12%)

Query: 3    NKKHKLLKNIEEFKTITQKRLTERGKFPYDTVHSTFEIKDENFIMERLKSSGLSMGKP-- 60
            N K+  +K  +   ++ Q  + E+ KF  D  H   E + E FI E  +   +   K
Sbjct: 1063 NVKYNEMKGAKN-DSLNQNEIIEKEKF--DLQH---ENRSERFIEEEKQICIVDDKKNNI 1116

Query: 61   --VDYMGVNGIPIYTKTLSIVNKFAFENNSKDSSYSSNINISEDKIKENDQKILDLIVKS 118
              VD    +  P Y + L +       +N  +  YS+      DKI +N++   ++  K
Sbjct: 1117 MNVDEKRKSDHPSYERVLKMEG-----SNKNEEGYSNT-----DKILKNEKNEKNVNEKK 1166

Query: 119  GANNQNLTDEEKVIAFTKYIGEITNYDNEAYRARNVDTEYYRASDLFSVTERKLAMCVGY 178
            G N++   +E+K     K + E  + ++E       D       F          +C
Sbjct: 1167 GENDEKNENEKKEENDEKNVNEKKDENDEKNENEKKDENDNNNNSYFYNNSDTFELCTNS 1226

Query: 179  SVTAARAFNIMGIPSYVVSGKSPQGISHAAVRAYYNRSWHIIDITASTYWKNGNYKTTYS 238
             +      N + IPS     ++ +GI  +      N S   I+        KN N ++ YS
Sbjct: 1227 LIFINNKKNSILIPS-----ENEKGIIGSQKEEEQNISPVKINNKKKDLCKNIN-ESDYS 1280

Query: 239  DFIKEYCIDGYD--VYDPAKTNNRFK-VKYMESNEAFENWIHNNGSKSML-------FIN 288
            D      ++  +  +Y    +N++ + ++  + NE + +  + N S++ L       ++
Sbjct: 1281 DKQYSVLLNSIEKKIYKKCSSNSKIRGIEKKKINEDYVDLKNINCSRNTLEFFLTKKYLK 1340

Query: 289  ESAALKDKKPKDDFVPVTEKEKNELIDKYKKLLSQIPENTQNPGEKNIRDYLKNEYEEIL 348
             S  + ++    +   V EK+K +   K KKL  +I  N   P +  I + + + +EY  +
Sbjct: 1341 SSELIINEHDCQNINNVYEKKKKKEQAK-KKLNRKI--NVNIPNDSIIEENMSSEYNFVK 1397

Query: 349  KKDN----LFEHEHAE-------FKESLNLNESFYLQLKKEE-------MKPSDNLKKEE 390
            KK+N     FE + ++       F    N   + L   +E+       ++  +N K+ E
Sbjct: 1398 KKNNNCMVKFETKRSKSILSSEIFAVKKNKKRATNLMRSEEQFISSIGLVEKGENKKRIE 1457

Query: 391  KPRENSVKERETPAENNDFVSVTEKNNLIDKYKELLSKIPENTQNPGEKNIRN--YLEKE 448
            +  E  +KE+    + N+F     KNNL ++   L  K  EN    G  N     ++++
Sbjct: 1458 EKDEEYIKEK-IKNKKNEF-----KNNLTEQL--LFFKSAENINTSGSFNTEKIRHVKRT 1509

Query: 449  YEELLQKDKLFKHEYTEFTKSLNLNETFYSQLKEGEMKLSENPEKGETN 497
              ++   +    + ++   K L   E    ++ E + ++++N EKGE N
Sbjct: 1510 KRKVNLSNNFILNNFSNILKKLQRMEEDKIKMDEQKKEINKNNEKGEFN 1558
```

There is also homology to SEQ ID 598.

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

Example 756

A DNA sequence (GBSx0803) was identified in *S.agalactiae* <SEQ ID 2323> which encodes the amino acid sequence <SEQ ID 2324>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1243(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 757

A DNA sequence (GBSx0804) was identified in *S.agalactiae* <SEQ ID 2325> which encodes the amino acid sequence <SEQ ID 2326>. This protein is predicted to be 2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate al. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1057(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 GENPEPT database.

```
>GP:AAD35160 GB:AE001693 2-dehydro-3-deoxyphosphogluconate
          aldolase/4-hydroxy-2-oxoglutarate aldolase [Thermotoga maritima]
 Identities = 78/192 (40%), Positives = 118/192 (60%), Gaps = 6/192 (3%)

Query: 14  KIVAVIRGNSQEEAFQAAQACIKGGISAIEIAYTNSKASQVIEQLVTQYTNQEQVVVGAG 73
           KIVAV+R NS EEA + A A  +GG+  IEI +T   A  VI++L   +  ++  ++GAG
Sbjct: 11  KIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKEL--SFLKEKGAIIGAG 68

Query: 74  TVLDSETARMAILAGAKFIVSPAFNLQTAKLCNRYAIPYLPGCMTLSEVTTALEAGCEII 133
           TV   E  R A+ +GA+FIVSP  + + ++ C    + Y+PG MT +E+  A++ G  I+
Sbjct: 69  TVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTIL 128

Query: 134 KIFPGGTLGTSFISSLKAPLPQVQIMVTGGVNLTNAKDWFLSGVTAIGIGGEFNKLAALG 193
           K+FPG  +G  F+ ++K P P V+ + TGGVNL N  +WF +GV A+G+G     K    G
Sbjct: 129 KLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVK----G 184

Query: 194 EFDKITEMAKQY 205
             D++ E AK +
Sbjct: 185 TPDEVREKAKAF 196
```

There is also homology to SEQ ID 1252.

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

Example 758

A DNA sequence (GBSx0805) was identified in *S.agalactiae* <SEQ ID 2327> which encodes the amino acid sequence <SEQ ID 2328>. This protein is predicted to be 2-keto-3-deoxygluconate kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4213(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 GENPEPT database.

```
>GP:AAD35161 GB:AE001693 2-keto-3-deoxygluconate kinase [Thermotoga maritima]
 Identities = 94/329 (28%), Positives = 169/329 (50%), Gaps = 7/329 (2%)

Query: 3   KILFFGEPLIRITPKENDYFADSISTKLFYGGSEVNTARALQGFGQDTKLLSALPNNPIG 62
           K++ FGE ++R++P ++      + S  + YGG+E N A  L   G D   ++ LPNNP+G
Sbjct: 2   KVVTFGEIMLRLSPPDHKRIFQTDSFDVTYGGAEANVAAFLAQMGLDAYFVTKLPNNPLG 61

Query: 63  NSFLQFLKAQGIDTHSIQWVGERVGLYFLEDSFACRKGEVVYDRDHSSLHDFRINQIDFD 122
           ++    L+  G+ T  I   G R+G+YFLE   + R  +VVYDR HS++ + +    D++
Sbjct: 62  DAAAGHLRKFGVKTDYIARGGNRIGIYFLEIGASQRPSKVVYDRAHSAISEAKREDFDWE 121

Query: 123 QLFEGVSLFHFSGITLSLDESIQEITLLLLKEAKKREITISLDLNFRSKLISPKNAKILF 182
           ++ +G   FHFSGIT  L + +  I    LK A ++ +T+S DLN+R++L + + A+ +
Sbjct: 122 KILDGARWFHFSGITPPLGKELPLILEDALKVANEKGVTVSCDLNYRARLWTKEEAQKVM 181

Query: 183 SQFATFADICFG----IEPLMVDSQDTTFFNRDEATIEDVKERMISLINHFDFQVIFHTK 238
             F  + D+       IE ++  S +       +   E   +    ++F+ +  T
Sbjct: 182 IPFMEYVDVLIANEEDIEKVLGISVEGLDLKTGKLNREAYAKIAEEVTRKYNFKTVGITL 241

Query: 239 RLQDEWGRNHYQAYI-ANRKQEFVTSKEITTAVNQRIGSGDAFVAGALYQLLQHSDSKTV 297
           R       N++   +  N +  F    EI   +  R+G+GD+F   +Y  L   DS+
Sbjct: 242 RESISATVNYWSVMVFENGQPHFSNRYEI--HIVDRVGAGDSFAGALIYGSLMGFDSQKK 299

Query: 298 IDFAVASASLKCALEGDNMFETVTAVNKV 326
            +FA A++ LK  + GD +  ++  + K+
Sbjct: 300 AEFAAAASCLKHTIPGDFVVLSIEEIEKL 328
```

There is also homology to SEQ ID 1264.

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

Example 759

A DNA sequence (GBSx0806) was identified in *S.agalactiae* <SEQ ID 2329> which encodes the amino acid sequence <SEQ ID 2330>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22   Transmembrane   53 -  69 (  53 -  70)

----- Final Results -----
               bacterial membrane --- Certainty=0.1086(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 GENPEPT database.

```
>GP:AAD36157 GB:AE001768 sugar-phosphate isomerase [Thermotoga maritima]
 Identities = 41/125 (32%), Positives = 61/125 (48%), Gaps = 10/125 (8%)

Query: 1   MKIALINENSQASKNTIIYKELKAVSDEKGFEVFNYGMYGKEEESQLTYVQNGLLTAILL 60
           MKIA+ ++++       + +++K     KG EV ++G Y +E      Y +  ++ +IL
Sbjct: 1   MKIAIASDHAAFE----LKEKVKNYLLGKGIEVEDHGTYSEESVDYPDYAKK-VVQSILS 55
```

```
Query: 61  NSGAADFVITGCGTGIGAMLACNSFPGVVCGFAADPVDAYLFSQVNGGNALSLPFAKGFG 120
           N   ADF I  CGTG+G  +A N + G+       P  A L    N  N L LP   G
Sbjct: 56  NE--ADFGILLCGTGLGMSIAANRYRGIRAALCLFPDMARLARSHNNANILVLP---GRL 110

Query: 121 WGAEL 125
           GAEL
Sbjct: 111 IGAEL 115
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2331> which encodes the amino acid sequence <SEQ ID 2332>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2599(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 159/212 (75%), Positives = 186/212 (87%)

Query: 1   MKIALINENSQASKNTIIYKELKAVSDEKGFEVFNYGMYGKEEESQLTYVQNGLLTAILL 60
           MKIALINENSQA+KN IIY  L  V+D+ G++VFNYGMYG E ESQLTYVQNGLL +ILL
Sbjct: 1   MKIALINENSQAAKNGIIYDALTTVTDKHGYQVFNYGMYGTEGESQLTYVQNGLLASILL 60

Query: 61  NSGAADFVITGCGTGIGAMLACNSFPGVVCGFAADPVDAYLFSQVNGGNALSLPFAKGFG 120
            + AADFV+TGCGTG+GAMLA NSFPGV CGFA++P +AYLFSQ+NGGNALS+PFAKGFG
Sbjct: 61  TTKAADFVVTGCGTGVGAMLALNSFPGVTCGFASEPTEAYLFSQINGGNALSIPFAKGFG 120

Query: 121 WGAELNLRYLFERLFEDEKGGGYPKERAVPEQRNARILSEIKQITYRDLLSVLKEIDQDF 180
           WGAELNL  +FERLF +  GGGYPKERA+PEQRNARILS++K+ITYRDLL+++K+IDQDF
Sbjct: 121 WGAELNLTLIFERLFAEPMGGGYPKERAIPEQRNARILSDLKKITYRDLLAIVKDIDQDF 180

Query: 181 LKETISGEHFQEYFFANCQNQNIADYLKSVLD 212
           LKETISG HFQEYFFAN +   +  YLKSVL+
Sbjct: 181 LKETISGAHFQEYFFANAEPSELVTYLKSVLE 212
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 760

A DNA sequence (GBSx0807) was identified in *S.agalactiae* <SEQ ID 2333> which encodes the amino acid sequence <SEQ ID 2334>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane   10 -  26 (   8 -  26)

----- 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>
```

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

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

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

Example 761

A DNA sequence (GBSx0808) was identified in *S.agalactiae* <SEQ ID 2335> which encodes the amino acid sequence <SEQ ID 2336>. This protein is predicted to be gluconate 5-dehydrogenase (fabG). Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1117(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 GENPEPT database.

```
>GP:AAC77223 GB:AE000497 5-keto-D-gluconate 5-reductase [Escherichia
          coli K12]
 Identities = 116/260 (44%), Positives = 165/260 (62%), Gaps = 6/260 (2%)

Query: 6   LKDNFSLEGKVALITGASYGIGFSIATAFARAGATIVFNDIKQELVDKGISAYKKLGIKA 65
           + D FSL GK  LITG++ GIGF +AT   + GA I+ NDI  E  +  +    + GI+A
Sbjct: 1   MNDLFSLAGKNILITGSAQGIGFLLATGLGKYGAQIIINDITAERAELAVEKLHQEGIQA 60

Query: 66  HGYVCDVTDEDGINEMVDKISQDVGVIDILVNNAGIIKRTPMLEMSAADFRQVIDIDLNA 125
                +VT +  I+  V+ I +D+G ID+LVNNAGI +R P  E    ++  VI ++  A
Sbjct: 61  VAAPFNVTHKHEIDAAVEHIEKDIGPIDVLVNNAGIQRRHPFTEFPEQEWNDVIAVNQTA 120

Query: 126 PFIVSKAVLPGMIQKGHGKIINICSMMSELGRETVAAYAAAKGGLKMLTKNIASEYGSAN 185
            F+VS+AV   M+++  GK+INICSM SELGR+T+  YAA+KG +KMLT+ +  E    N
Sbjct: 121 VFLVSQAVTRHMVERKAGKVINICSMQSELGRDTITPYAASKGAVKMLTRGMCVELARHN 180

Query: 186 IQCNGIGPGYIATPQTAPLRERQDDGSRHPFDQFIIAKTPAARWGEAEDLGAPAIFLASD 245
           IQ NGI PGY  T  T  L E +       F  ++  +TPAARWG+ ++L   A+FL+S
Sbjct: 181 IQVNGIAPGYFKTEMTKALVEDE------AFTAWLCKRTPAARWGDPQELIGAAVFLSSK 234

Query: 246 ASNFINGHILYVDGGILAYI 265
           AS+F+NGH+L+VDGG+L  +
Sbjct: 235 ASDFVNGHLLFVDGGMLVAV 254
```

There is also homology to SEQ ID 1242:

```
Identities = 225/264 (85%), Positives = 246/264 (92%)

Query: 6   LKDNFSLEGKVALITGASYGIGFSIATAFARAGATIVFNDIKQELVDKGISAYKKLGIKA 65
           +++ FSL+GK+ALITGASYGIGF IA A+A+AGATIVFNDIKQELVDKG++AY++LGI+A
Sbjct: 1   MENMFSLQGKIALITGASYGIGFEIAKAYAQAGATIVFNDIKQELVDKGLAAYRELGIEA 60

Query: 66  HGYVCDVTDEDGINEMVDKISQDVGVIDILVNNAGIIKRTPMLEMSAADFRQVIDIDLNA 125
           HGYVCDVTDE GI +MV +I  +VG IDILVNNAGII+RTPMLEM+A DFRQVIDIDLNA
Sbjct: 61  HGYVCDVTDEAGIQQMVSQIEDEVGAIDILVNNAGIIRRTPMLEMAAEDFRQVIDIDLNA 120

Query: 126 PFIVSKAVLPGMIQKGHGKIINICSMMSELGRETVAAYAAAKGGLKMLTKNIASEYGSAN 185
           PFIVSKAVLP MI KGHGKIINICSMMSELGRETV+AYAAAKGGLKMLTKNIASE+G AN
Sbjct: 121 PFIVSKAVLPSMIAKGHGKIINICSMMSELGRETVSAYAAAKGGLKMLTKNIASEFGEAN 180

Query: 186 IQCNGIGPGYIATPQTAPLRERQDDGSRHPFDQFIIAKTPAARWGEAEDLGAPAIFLASD 245
           IQCNGIGPGYIATPQTAPLRERQ DGSRHPFDQFIIAKTPAARWG  EDL  PA+FLASD
Sbjct: 181 IQCNGIGPGYIATPQTAPLRERQADGSRHPFDQFIIAKTPAARWGTTEDLAGPAVFLASD 240

Query: 246 ASNFINGHILYVDGGILAYIGKQP 269
           ASNF+NGHILYVDGGILAYIGKQP
Sbjct: 241 ASNFVNGHILYVDGGILAYIGKQP 264
```

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

Example 762

A DNA sequence (GBSx0809) was identified in *S.agalactiae* <SEQ ID 2337> which encodes the amino acid sequence <SEQ ID 2338>. This protein is predicted to be mannose-specific phosphotransferase system component IIAB. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0886(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 GENPEPT database.

```
>GP:AAD46485 GB:AF130465 mannose-specific phosphotransferase system
          component IIAB [Streptococcus salivarius]
 Identities = 43/107 (40%), Positives = 61/107 (56%), Gaps = 3/107 (2%)

Query: 2   IKIIIVAHGNFPDGILSSLELIAGHQEYVVGINFIAGMSSNDVRVALQREVIDFK---EI 58
           I III +HG F +GI  S  +I G QE V  + F+     +D+       +  F    EI
Sbjct: 3   IGIIIASHGKFAEGIHQSGSMIFGDQEKVQVVTFMPSEGPDDLYAHFNDAIAQFDADDEI 62

Query: 59  LVLTDLLGGTPFNVSSALSVEYTDKKIKVLSGLNLSMLMEAVLSRTM 105
           LVL DL  G+PFN +S ++ E  D+KI +++GLNL ML++A   R M
Sbjct: 63  LVLADLWSGSPFNQASRIAGENPDRKIAIITGLNLPMLIQAYTERMM 109
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2339> which encodes the amino acid sequence <SEQ ID 2340>. 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>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF81086 GB:AF228498 AgaF [Escherichia coli]
 Identities = 48/127 (37%), Positives = 71/127 (55%), Gaps = 6/127 (4%)

Query: 1   MIAIIVMGHGHFASGIVSALELIAGKQEKVTAIDFTTEMTAADVQDQLSRALIP---EEE 57
           M++II+ GHG FASG+  A++ I G+Q +  AID     + A +  QL  A+      E+
Sbjct: 1   MLSIILTGHGGFASGMEKAMKQILGEQSQFIAIDVPETSSTALLTSQLEEAIAQLDCEDG 60

Query: 58  TLVLCDLLGGTPFKVAATLMESLPNTTCNVLSGLNLAMLIEASFARQTAASFDDLVSGLI 117
            + L DLLGGTPF+VA+TL     P   C V++G NL +L+E     R+  +  +  V  L
Sbjct: 61  IVFLTDLLGGTPFRVASTLAMQKPG--CEVITGTNLQLLLEMVLEREGLSGEEFRVQAL- 117

Query: 118 TCSKEGI 124
            C   G+
Sbjct: 118 ECGHRGL 124
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 73/146 (50%), Positives = 94/146 (64%), Gaps = 3/146 (2%)

Query: 1   MIKIIIVAHGNFPDGILSSLELIAGHQEYVVGINFIAGMSSNDVRVALQREVIDFKEILV 60
```

```
          MI II++ HG+F  GI+S+LELIAG QE V  I+F   M++ DV+  L R +I  +E LV
Sbjct: 1  MIAIIVMGHGHFASGIVSALELIAGKQEKVTAIDFTTEMTAADVQDQLSRALIPEEETLV 60

Query: 61  LTDLLGGTPFNVSSALSVEYTDKKIKVLSGLNLSMLMEAVLSRTMFEHVDDLVDKVITSS 120
           L DLLGGTPF V++ L     +    VLSGLNL+ML+EA  +R      DDLV  +IT S
Sbjct: 61  LCDLLGGTPFKVAATLMESLPNTTCNVLSGLNLAMLIEASFARQTAASFDDLVSGLITCS 120

Query: 121 HEGIVDFSTCLATQTAEATFE--GGI 144
            EGIVD+ T L+ Q   AT +  GGI
Sbjct: 121 KEGIVDWKT-LSQQEDGATDDELGGI 145
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 763

A DNA sequence (GBSx0811) was identified in *S.agalactiae* <SEQ ID 2341> which encodes the amino acid sequence <SEQ ID 2342>. This protein is predicted to be unsaturated glucuronyl hydrolase. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.11   Transmembrane  172 - 188 ( 172 - 188)

----- Final Results -----
               bacterial membrane --- Certainty=0.1044(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 GENPEPT database.

```
>GP:BAB05773 GB:AP001514 unsaturated glucuronyl hydrolase [Bacillus halodurans]
 Identities = 156/370 (42%), Positives = 219/370 (59%), Gaps = 3/370 (0%)

Query: 30  EEAIEKALKQLYINIDYFGEEYPTPATFNNIYKVMDNTEWTNGFWTGCLWLAYEYNQDKK 89
           ++A+    ++   NI  F   +P  +     Y++ +N EWTNGFW+G LWL YEY  D
Sbjct: 4   KQAMTDVAEKTLTNIKRFNGRFPHVSEDGEHYELNNNNEWTNGFWSGILWLCYEYTNDPA 63

Query: 90  LKNIAHKNVLSFLNRINNRIALDHHDLGFLYTPSCTAEYRINGDVKALEATIKAADKLME 149
            +  A   V SF  R+   + LDHHD+GFLY+ S  A++ I  D +A + TI+AAD LM+
Sbjct: 64  FRQAAASTVRSFQQRMEQNLELDHHDIGFLYSLSSKAQWIIERDERAKQLTIEAADVLMK 123

Query: 150 RYQEKGGFIQAWGELG-YKEHYRLIIDCLLNIQLLFFAYEQTGDEKYRQVAVNHFYASAN 208
           R++EK    QAWG  G      R+I+DCL+N+ LLF+A E TG+  YR+ A+ H   +
Sbjct: 124 RWREKIELFQAWGPEGDLSNGGRIIVDCLMNLPLLFWASEVTGNPDYREAAIIHADKTRR 183

Query: 209 NVVRDDSSAFHTFYFDPETGEPLKGVTRQGYSDESSWARGQAWGIYGIPLSYRKMKDYQQ 268
            +VR D S +HTFYF+ ETGE L+G T QGY D S+W+RGQAW IYG  ++YR   + +
Sbjct: 184 FIVRGDDSTYHTFYFNQETGEALRGGTHQGYEDGSTWSRGQAWAIYGFAIAYRYTGNERY 243

Query: 269 IILFKGMTNYFLNRLPEDKVSYWDLIFTDGSGQPRDTSATATAVCGIHEMLKYLPEVDPD 328
           +   K    YF+  LP D V+YWD          RD+SA+A A CGI E+L +L E DPD
Sbjct: 244 LETAKRTAKYFIENLPADYVAYWDFNAPITPDTKRDSSASAIASCGILELLSHLQETDPD 303

Query: 329 KETYKYAMHTMLRSLIEQYSNNELIAGRPLLLHGVYSWHSGKGVDEGNIWGDYYYLEALI 388
           K  ++ ++   + SL+E Y++ +   G  L+  G YS   G   D+  IWGDY+Y EAL+
Sbjct: 304 KAFFQQSVQKQMTSLVENYASEKDAQG--LIKRGSYSVRIGHAPDDYVIWGDYFYTEALM 361

Query: 389 RFYKDWELYW 398
           R  K    YW
Sbjct: 362 RLEKLRNGYW 371
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2343> which encodes the amino acid sequence <SEQ ID 2344>. Analysis of this protein sequence reveals the following:

```
     Possible site: 33
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane  173 - 189 ( 173 - 189)

----- 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>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 273/395 (69%), Positives = 336/395 (84%)

Query: 4   IKPVKVESIENPKRFLNSRLLTKIEVEEAIEKALKQLYINIDYFGEEYPTPATFNNIYKV 63
           +K + +E I+ P+RF      L++ ++ +A++ ALKQ+ +N+DYF E++PTPAT +N Y +
Sbjct: 5   LKTIALEPIKQPERFTKEDFLSQEDITQALDLALKQVRLNMDYFKEDFPTPATKDNQYAI 64

Query: 64  MDNTEWTNGFWTGCLWLAYEYNQDKKLKNIAHKNVLSFLNRINNRIALDHHDLGFLYTPS 123
           MDNTEWTN FWTGCLWLAYEY+ D  +K +A  N LSFL+R+    I LDHHDLGFLYTPS
Sbjct: 65  MDNTEWTNAFWTGCLWLAYEYSGDDAIKALAQANDLSFLDRVTRDIELDHHDLGFLYTPS 124

Query: 124 CTAEYRINGDVKALEATIKAADKLMERYQEKGGFIQAWGELGYKEHYRLIIDCLLNIQLL 183
           C AE+++     ++ EA +KAADKL++RYQ+KGGFIQAWGELG KE YRLIIDCLLNIQLL
Sbjct: 125 CMAEWKLLKTPESREAALKAADKLVQRYQDKGGFIQAWGELGKKEDYRLIIDCLLNIQLL 184

Query: 184 FFAYEQTGDEKYRQVAVNHFYASANNVVRDDSSAFHTFYFDPETGEPLKGVTRQGYSDES 243
           FFA ++TGD +YR +A+NHFYASAN+V+RDD+SA+HTFYFDPETG+P+KGVTRQGYSD+S
Sbjct: 185 FFASQETGDNRYRDMAINHFYASANHVIRDDASAYHTFYFDPETGDPVKGVTRQGYSDDS 244

Query: 244 SWARGQAWGIYGIPLSYRKMKDYQQIILFKGMTNYFLNRLPEDKVSYWDLIFTDGSGQPR 303
           +WARGQAWGIYGIPL+YR +K+ + I LFKGMT+YFLNRLP+D+VSYWDLIF DGS Q R
Sbjct: 245 AWARGQAWGIYGIPLTYRFLKEPELIQLFKGMTHYFLNRLPKDQVSYWDLIFGDGSEQSR 304

Query: 304 DTSATATAVCGIHEMLKYLPEVDPDKETYKYAMHTMLRSLIEQYSNNELIAGRPLLLHGV 363
           D+SATA AVCGIHEMLK LP+ DPDK+TY+ AMH+MLR+LI+ Y+N +L  G PLLLHGV
Sbjct: 305 DSSATAIAVCGIHEMLKTLPDHDPDKKTYEAAMHSMLRALIKDYANKDLKPGAPLLLHGV 364

Query: 364 YSWHSGKGVDEGNIWGDYYYLEALIRFYKDWELYW 398
           YSWHSGKGVDEGNIWGDYYYLEAL+RFYKDW  YW
Sbjct: 365 YSWHSGKGVDEGNIWGDYYYLEALLRFYKDWNPYW 399
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 764

A DNA sequence (GBSx0812) was identified in *S.agalactiae* <SEQ ID 2345> which encodes the amino acid sequence <SEQ ID 2346>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3035(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 GENPEPT database.

```
>GP:AAC44679 GB:U65015 PTS permease for mannose subunit IIIMan C
          terminal domain [Vibrio furnissii]
 Identities = 63/125 (50%), Positives = 89/125 (70%), Gaps = 1/125 (0%)

Query: 5   PNIVMTRVDERLIHGQ-GQLWVKFLSCNTVIVANDDVSKDHLQQTLMKTVVPESIALRFF 63
           PNIV++R+DERL+HGQ G  WV F   N V+VAND+V+ D +QQ LM+ V+ + IA+RF+
```

```
Sbjct: 2   PNIVLSRIDERLVHGQVGVQWVGFADANIVVVANDEVAADTIQQNLMEMVLADGIAIRFW 61

Query: 64  DIQKVIDIIHKANPAQTIFIIVKDLKDVYRLVAGGVPIKEINIGNIHNGEGKEQVSRSIF 123
           +QK ID IHKA+  Q I ++ K   D  RLV GGVPI  IN+GN+H  +GK Q+S+++
Sbjct: 62  TVQKTIDTIHKASDRQRILLVCKTPHDFRRLVEGGVPIAAINVGNMHYIDGKTQISKTVS 121

Query: 124 LGMKD 128
           +  +D
Sbjct: 122 VDAED 126
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2347> which encodes the amino acid sequence <SEQ ID 2348>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2511(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:

```
>GP:BAA84216 GB:AB019619 unsaturated glucuronyl hydrolase [Bacillus
           sp. GL1]
 Identities = 161/369 (43%), Positives = 220/369 (58%), Gaps = 1/369 (0%)

Query: 32  QALDLALKQVRLNMDYFKEDFPTPATKDNQYAIMDNTEWTNAFWTGCLWLAYEYSGDDAI 91
           QA+  AL     N+  F + FP  +   N+Y + DNT+WT+ FW+G LWL YEY+GD+
Sbjct: 4   QAIGDALGITARNLKKFGDRFPHVSDGSNKYVLNDNTDWTDGFWSGILWLCYEYTGDEQY 63

Query: 92  KALAQANDLSFLDRVTRDIELDHHDLGFLYTPSCMAEWKLLKTPESREAALKAADKLVQR 151
           +  A     SF +R+ R   LDHHD+GFLY+ S  A+W + K   +R+ AL AAD L++R
Sbjct: 64  REGAVRTVASFRERLDRFENLDHHDIGFLYSLSAKAQWIVEKDESARKLALDAADVLMRR 123

Query: 152 YQDKGGFIQAWGELGKKEDY-RLIIDCLLNIQLLFFASQETGDNRYRDMAINHFYASANH 210
           ++   G IQAWG  G  E+  R+IIDCLLN+ LL +A ++TGD  YR +A  H   S
Sbjct: 124 WRADAGIIQAWGPKGDPENGGRIIIDCLLNLPLLLWAGEQTGDPEYRRVAEAHALKSRRF 183

Query: 211 VIRDDASAYHTFYFDPETGDPVKGVTRQGYSDDSAWARGQAWGIYGIPLTYRFLKEPELI 270
           ++R D S+YHTFYFDPE G+ ++G T QG +D S W RGQAWGIYG  L  R+L   +L+
Sbjct: 184 LVRGDDSSYHTFYFDPENGNAIRGGTHQGNTDGSTWTRGQAWGIYGFALNSRYLGNADLL 243

Query: 271 QLFKGMTHYFLNRLPKDQVSYWDLIFGDGSEQSRDSSATAIAVCGIHEMLKTLPDHDPDK 330
           +  K M  +FL R+P+D V YWD           RDSSA+AI  CG+ E+   L + DP++
Sbjct: 244 ETAKRMARHFLARVPEDGVVYWDFEVPQEPSSYRDSSASAITACGLLEIASQLDESDPER 303

Query: 331 KTYEAAMHSMLRALIKDYANKDLKPGAPLLLHGVYSWHSGKGVDEGNIWGDYYYLEALLR 390
           + +  A  + + AL   YA +D      +  G Y   G  D+  IWGDYYYLEALLR
Sbjct: 304 QRFIDAAKTTVTALRDGYAERDDGEAEGFIRRGSYHVRGGISPDDYTIWGDYYYLEALLR 363

Query: 391 FYKDWNPYW 399
             +   YW
Sbjct: 364 LERGVTGYW 372
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 112/160 (70%), Positives = 132/160 (82%), Gaps = 1/160 (0%)

Query: 5   PNIVMTRVDERLIHGQGQLWVKFLSCNTVIVANDDVSKDHLQQTLMKTVVPESIALRFFD 64
           PNI+MTRVDERLIHGQGQLWVKFL+CNTVIVAND VS+D +QQ+LMKTV+P SIA+RFF
Sbjct: 4   PNIIMTRVDERLIHGQGQLWVKFLNCNTVIVANDAVSEDKIQQSLMKTVIPSSIAIRFFS 63

Query: 65  IQKVIDIIHKANPAQTIFIIVKDLKDVYRLVAGGVPIKEINIGNIHNGEGKEQVSRSIFL 124
           IQKVIDIIHKA+PAQ+IFI+VKDL+D   LV GGVPI EINIGNIH  + K  +++ I L
Sbjct: 64  IQKVIDIIHKASPAQSIFIVVKDLQDAKLLVEGGVPITEINIGNIHKTDDKVAITQFISL 123
```

```
Query: 125 GMKDKEIIRKLNQEYHIAFNTKTTPTGNDGAVEVNILDYI 164
           G  DK  IR L  ++H+ FNTKTTP GN  A +V+ILDYI
Sbjct: 124 GETDKSAIRCLAHDHHVVFNTKTTPAGN-SASDVDILDYI 162
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 765

A DNA sequence (GBSx0813) was identified in *S.agalactiae* <SEQ ID 2349> which encodes the amino acid sequence <SEQ ID 2350>. This protein is predicted to be AgaW (agaC). Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.95   Transmembrane  251 - 267 ( 244 - 269)
   INTEGRAL    Likelihood = -4.30   Transmembrane  213 - 229 ( 208 - 230)
   INTEGRAL    Likelihood = -2.71   Transmembrane  149 - 165 ( 148 - 165)
   INTEGRAL    Likelihood = -1.81   Transmembrane   31 -  47 (  31 -  49)
   INTEGRAL    Likelihood = -1.49   Transmembrane  173 - 189 ( 173 - 189)

----- Final Results -----
              bacterial membrane --- Certainty=0.3781(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 GENPEPT database.

```
>GP:AAF81084 GB:AF228498 AgaW [Escherichia coli]
 Identities = 93/295 (31%), Positives = 140/295 (46%), Gaps = 48/295 (16%)

Query: 1   MDISILQAVLIGLWTAFCFSGMLLGL-YTNRCIVLSLGVGVILGDIQTALAVGAISELAY 59
           M+IS+LQA  +G+        M  GL + +R +VL   VG++LGD+ T +  G   EL +
Sbjct: 1   MEISLLQAFALGIIAFIAGLDMFNGLTHMHRPVVLGPLVGLVLGDLHTGILTGGTLELVW 60

Query: 60  MGFGVGAGGTVPPNPIGPGIFGTLMAITTAGTKGKITPEAALALSTPIAVGIQFLQTATY 119
           MG    AG   PPN I   I GT  AITT      + P+ A+ ++ P AV +Q   T  +
Sbjct: 61  MGLAPLAGAQ-PPNVIIGTIVGTAFAITTG-----VKPDVAVGVAVPFAVAVQMGITFLF 114

Query: 120 TAFAGAPETAKK--------ALQAGNFRGFKIAANGT-IWAFAGLGFGLGVLGALSTQTL 170
           +  +G      +        AL A N+       N   + AF  + FG     A   +T+
Sbjct: 115 SVMSGVMSRCARMPRTPILAALNACNYLALLALGNFYFLCAFLPIYFG-----AEHAKTI 169

Query: 171 TDLFALIPPVLLNGLTLAGKMLPAIGFAMILSVMAKKELIPYILLGYVLAVYFGLPVLTP 230
            D+   +P  L++GL +AG ++PAIGFA++L +M K   IPY +LG+V A +  LPVL
Sbjct: 170 IDV---LPQRLIDGLGVAGGIMPAIGFAVLLKIMMKNVYIPYFILGFVAAAWLKLPVL-- 224

Query: 231 TANGDGVLTSVATNSVLGVPTIGVAIIATIFALLDIFRKPAAPTKETKTEGDNQD 285
                                +A  A   AL+D+ RK   PT+     + + +D
Sbjct: 225 ----------------------AIACPALAMALIDLLRKSPEPTQPAAQKEEFED 257
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2351> which encodes the amino acid sequence <SEQ ID 2352>. Analysis of this protein sequence reveals the following:

```
     Possible site: 52
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.37   Transmembrane  220 - 236 ( 214 - 241)
   INTEGRAL    Likelihood = -5.10   Transmembrane  146 - 162 ( 144 - 165)
   INTEGRAL    Likelihood = -1.59   Transmembrane  184 - 200 ( 184 - 202)

----- 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>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC44680 GB:U65015 PTS permease for mannose subunit IIPMan
          [Vibrio furnissii]
 Identities = 86/255 (33%), Positives = 137/255 (53%), Gaps = 11/255 (4%)

Query: 1   MDINLLQALLIGLWTAFCFSGMLLGI-YTNRCIILSFGVGIILGDLPTALSMGAISELAY 59
           M+I L QAL++GL        +  G+ + +R ++L   VG+ILGDL T + +G   EL +
Sbjct: 1   MEIGLFQALMLGLLAFLAGLDLFNGLTHFHRPVVLGPLVGLILGDLHTGILVGGTLELIW 60

Query: 60  MGFGVGAGGTVPPNPIGPGIFGTLMAITSAGKVTPEAALALSTPIAVAIQFLQTFAYTAF 119
           MG    AG   PPN I   I GT  AIT+   V P  A+ ++ P AVA+Q   T  ++A
Sbjct: 61  MGLAPLAGAQ-PPNVIIGTIVGTTFAITT--NVEPNVAVGVAVPFAVAVQMGITLLFSAM 117

Query: 120 AGAPETAKKQLQKGNIRGFK---FAANGTIWAFAFIGLGLGLLGALSMDTLLHLVDYIPP 176
           +       +  +  + RG +   + A   + +F F+   L +   L  D    +V  +P
Sbjct: 118 SAVMSKCDEYAKNADTRGIERVNYFALAVLGSFYFLCAFLPIY--LGADHAGAMVAALPK 175

Query: 177 VLLNGLTVAGKMLPAIGFAMILSVMAKKELIPFVLIGYVCAAYLQIPTIGIAIIGIIFAL 236
            L++GL VAG ++PAIGFA+++ +M K   IP+ ++G+V AA+LQ+P + I       A+
Sbjct: 176 ALIDGLGVAGGIMPAIGFAVLMKIMMKNAYIPYFILGFVAAAWLQLPILAIRCAATAMAI 235

Query: 237 NEFYNK--PKQVDAT 249
            +F  K  P  V+A+
Sbjct: 236 IDFMRKSEPTPVNAS 250
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 203/288 (70%), Positives = 225/288 (77%), Gaps = 28/288 (9%)

Query: 1   MDISILQAVLIGLWTAFCFSGMLLGLYTNRCIVLSLGVGVILGDIQTALAVGAISELAYM 60
           MDI++LQA+LIGLWTAFCFSGMLLG+YTNRCI+LS GVG+ILGD+ TAL++GAISELAYM
Sbjct: 1   MDINLLQALLIGLWTAFCFSGMLLGIYTNRCIILSFGVGIILGDLPTALSMGAISELAYM 60

Query: 61  GFGVGAGGTVPPNPIGPGIFGTLMAITTAGTKGKITPEAALALSTPIAVGIQFLQTATYT 120
           GFGVGAGGTVPPNPIGPGIFGTLMAIT+AG   K+TPEAALALSTPIAV IQFLQT  YT
Sbjct: 61  GFGVGAGGTVPPNPIGPGIFGTLMAITSAG---KVTPEAALALSTPIAVAIQFLQTFAYT 117

Query: 121 AFAGAPETAKKALQAGNFRGFKIAANGTIWAFAGLGFGLGVLGALSTQTLTDLFALIPPV 180
           AFAGAPETAKK LQ GN RGFK AANGTIWAFA +G GLG+LGALS  TL  L   IPPV
Sbjct: 118 AFAGAPETAKKQLQKGNIRGFKFAANGTIWAFAFIGLGLGLLGALSMDTLLHLVDYIPPV 177

Query: 181 LLNGLTLAGKMLPAIGFAMILSVMAKKELIPYILLGYVLAVYFGLPVLTPTANGDGVLTS 240
           LLNGLT+AGKMLPAIGFAMILSVMAKKELIP++L+GYV A Y
Sbjct: 178 LLNGLTVAGKMLPAIGFAMILSVMAKKELIPFVLIGYVCAAY---------------- 219

Query: 241 VATNSVLGVPTIGVAIIATIFALLDIFRKPAAPTKETKTEGDNQDDWI 288
                L +PTIG+AII  IFAL + + KP      T  +G  QDDWI
Sbjct: 220 ------LQIPTIGIAIIGIIFALNEFYNKP-KQVDATTVQGGQQDDWI 260
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 766

A DNA sequence (GBSx0814) was identified in *S.agalactiae* <SEQ ID 2353> which encodes the amino acid sequence <SEQ ID 2354>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2442(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 767

A DNA sequence (GBSx0815) was identified in *S.agalactiae* <SEQ ID 2355> which encodes the amino acid sequence <SEQ ID 2356>. This protein is predicted to be PTS permease for mannose subunit IIBMan. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.28   Transmembrane  278 - 294 ( 272 - 294)
   INTEGRAL    Likelihood = -3.45   Transmembrane  155 - 171 ( 155 - 174)
   INTEGRAL    Likelihood = -1.59   Transmembrane  250 - 266 ( 250 - 267)

----- Final Results -----
               bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8657> which encodes amino acid sequence <SEQ ID 8658> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:     -9.70
GvH: Signal Score (-7.5): -6.12
     Possible site: 19
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 3 value:  -8.28 threshold:  0.0
   INTEGRAL    Likelihood = -8.28   Transmembrane  254 - 270 ( 248 - 270)
   INTEGRAL    Likelihood = -3.45   Transmembrane  131 - 147 ( 131 - 150)
   INTEGRAL    Likelihood = -1.59   Transmembrane  226 - 242 ( 226 - 243)
   PERIPHERAL  Likelihood =  0.37     175
 modified ALOM score:   2.16

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4312(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 GENPEPT database.

```
>GP:AAA57943 GB:U18997 ORF_o290; Geneplot suggests frameshift
          linking to o267, not found [Escherichia coli]
 Identities = 101/278 (36%), Positives = 164/278 (58%), Gaps = 6/278 (2%)

Query: 17  LRQKETTKMTGSKKLAKSDYTKTALRAFYLQNGFNYSNYQGLGYANVIYPALKKYYGDDK 76
           ++ K+ T   GS+ ++K D T+   R+  LQ  FNY   Q  G+   + P LKK Y DDK
Sbjct: 19  VKMKKRTTAMGSE-ISKKDITRLGFRSSLLQASFNYERMQAGGFTWAMLPILKKIYKDDK 77

Query: 77  KALAGALEENVEFYNTNPHFLPFVTSLHLAMLDNERPEEEIRGIKMALMGPLAGIGDSLS 136
             L+ A+++N+EF NT+P+ + F+  L ++M +     + I+G+K+AL GP+AGIGD++
Sbjct: 78  PGLSAAMKDNLEFINTHPNLVGFLMGLLISMEEKGENRDTIKGLKVALFGPIAGIGDAIF 137

Query: 137 QFCLAPLFSTIAASLATDGLVMGPILFFVAMNTILTGIKLVTGMYGYRLGTSFIDKLSEQ 196
            F L P+ + I +S A+ G ++GPILFF A+  ++  +++     GY +G   IDK+ E
Sbjct: 138 WFTLLPIMAGICSSFASQGNLLGPILFF-AVYLLIFFLRVGWTHVGYSVGVKAIDKVREN 196
```

```
Query: 197 MSVISRAANIVGVTVISSLAATQVKLTIPYTFAPEKVTSTTQKIVTVQGMLDKIAPALLP 256
           +I+R+A I+G+TVI  L A+ V + +  +FA       T  +   Q   DK+ P +LP
Sbjct: 197 SQMIARSATILGITVIGGLIASYVHINVVTSFA----IDNTHSVALQQDFFDKVFPNILP 252

Query: 257 ALYTFLMFYLIKNKKWTTYKLVILTVIIGILGSWLGIL 294
            YT LM+Y ++ KK     L+ +T ++ I+ S  GIL
Sbjct: 253 MAYTLLMYYFLRVKKAHPVLLIGVTFVLSIVCSAFGIL 290
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2357> which encodes the amino acid sequence <SEQ ID 2358>. Analysis of this protein sequence reveals the following:

```
     Possible site: 45
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.49   Transmembrane  276 - 292 ( 270 - 292)
   INTEGRAL    Likelihood = -7.01   Transmembrane  151 - 167 ( 149 - 176)
   INTEGRAL    Likelihood = -3.03   Transmembrane  202 - 218 ( 202 - 220)
   INTEGRAL    Likelihood = -2.13   Transmembrane  249 - 265 ( 248 - 265)

----- 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>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAA57943 GB:U18997 ORF_o290; Geneplot suggests frameshift
           linking to o267, not found [Escherichia coli]
 Identities = 104/285 (36%), Positives = 162/285 (56%), Gaps = 7/285 (2%)

Query: 8   NKSMQQLSKEANKMTGSNKLTKKDYLKTALRAFFLQNGFNYNNYQGIGYANVIYPALKKH 67
           N+S   +  +       ++++KKD  +   R+  LQ  FNY   Q  G+   + P LKK
Sbjct: 13  NRSPLPVKMKKRTTAMGSEISKKDITRLGFRSSLLQASFNYERMQAGGFTWAMLPILKKI 72

Query: 68  FGNDKKGLYQALEDNCEFYNTNPHFLPFITSLHLVMLENNRPEEETRNIKMALMGPLAGI 127
           + +DK GL  A++DN EF NT+P+ + F+  L + M E     +  + +K+AL GP+AGI
Sbjct: 73  YKDDKPGLSAAMKDNLEFINTHPNLVGFLMGLLISMEEKGENRDTIKGLKVALFGPIAGI 132

Query: 128 GDSLSQFCLAPLFSTIAASLASDGLVLGPILFFLAMNIILTAIKIGSGLYGYKVGTSFID 187
           GD++  F L P+ + I +S AS G +LGPILFF A+ +++  +++G    GY VG   ID
Sbjct: 133 GDAIFWFTLLPIMAGICSSFASQGNLLGPILFF-AVYLLIFFLRVGWTHVGYSVGVKAID 191

Query: 188 KLSEQMAVVSRMANIVGVTVIAGLAATSVKITVPITFAAGKVDAANTAQKFVTIQGMLDK 247
           K+ E   +++R A I+G+TVI GL A+ V I V  +FA     +    Q F      DK
Sbjct: 192 KVRENSQMIARSATILGITVIGGLIASYVHINVVTSFAIDNTHSVALQQDF------FDK 245

Query: 248 IAPALLPALFTLLMYYLIKNKKWTTYKLVILTVIIGVIGSWLGIL 292
           + P +LP  +TLLMYY ++ KK     L+ +T ++ ++ S  GIL
Sbjct: 246 VFPNILPMAYTLLMYYFLRVKKAHPVLLIGVTFVLSIVCSAFGIL 290
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 224/288 (77%), Positives = 255/288 (87%), Gaps = 4/288 (1%)

Query: 12  HLLKKLRQ--KETTKMTGSKKLAKSDYTKTALRAFYLQNGFNYSNYQGLGYANVIYPALK 69
           +L K ++Q  KE  KMTGS KL K DY KTALRAF+LQNGFNY+NYQG+GYANVIYPALK
Sbjct: 6   NLNKSMQQLSKEANKMTGSNKLTKKDYLKTALRAFFLQNGFNYNNYQGIGYANVIYPALK 65

Query: 70  KYYGDDKKALAGALEENVEFYNTNPHFLPFVTSLHLAMLDNERPEEEIRGIKMALMGPLA 129
           K++G+DKK L  ALE+N EFYNTNPHFLPF+TSLHL ML+N RPEEE R IKMALMGPLA
Sbjct: 66  KHFGNDKKGLYQALEDNCEFYNTNPHFLPFITSLHLVMLENNRPEEETRNIKMALMGPLA 125

Query: 130 GIGDSLSQFCLAPLFSTIAASLATDGLVMGPILFFVAMNTILTGIKLVTGMYGYRLGTSF 189
           GIGDSLSQFCLAPLFSTIAASLA+DGLV+GPILFF+AMN ILT IK+ +G+YGY++GTSF
Sbjct: 126 GIGDSLSQFCLAPLFSTIAASLASDGLVLGPILFFLAMNIILTAIKIGSGLYGYKVGTSF 185

Query: 190 IDKLSEQMSVISRAANIVGVTVISSLAATQVKLTIPYTFAPEKV--TSTTQKIVTVQGML 247
```

```
           IDKLSEQM+V+SR ANIVGVTVI+ LAAT VK+T+P TFA  KV   +T QK VT+QGML
Sbjct: 186 IDKLSEQMAVVSRMANIVGVTVIAGLAATSVKITVPITFAAGKVDAANTAQKFVTIQGML 245

Query: 248 DKIAPALLPALYTFLMFYLIKNKKWTTYKLVILTVIIGILGSWLGILA 295
           DKIAPALLPAL+T LM+YLIKNKKWTTYKLVILTVIIG++GSWLGILA
Sbjct: 246 DKIAPALLPALFTLLMYYLIKNKKWTTYKLVILTVIIGVIGSWLGILA 293
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 768

A DNA sequence (GBSx0816) was identified in *S.agalactiae* <SEQ ID 2359> which encodes the amino acid sequence <SEQ ID 2360>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane  135 - 151 ( 135 - 151)

----- 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>
```

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

```
>GP:CAB01924 GB:Z79691 OrfA [Streptococcus pneumoniae]
 Identities = 76/206 (36%), Positives = 124/206 (59%), Gaps = 1/206 (0%)

Query: 428 SWTYNSYPKCDYCQLTSKDRYHLVEGQLHVQRASDIYYHKRWLLTLPQAITLVIDKVSCP 487
           SW Y  YP   +C    ++  H +EG          Y HKR +L L + + L++D + C
Sbjct: 2   SWEYEYYPHSLFCHHKEREGMHYIEGAYWSAEPDLPYLHKRKILMLVEDVWLLVDDIRCQ 61

Query: 488 GEHVLTNQYILDDQVIYENGFVNDLKLVSPTTFNLEDCLISKRYNQLTESHKLVKKIKFV 547
           G+H    Q+ILD  V Y++G +N L+L S   F+LED +IS +YN+L  S KL K+  F
Sbjct: 62  GQHEALTQFILDKDVTYQDGKINQLRLWSEVDFDLEDTIISPKYNELERSSKLTKRQFFE 121

Query: 548 DEVMDYTLIVDRNCQVKYVPLVQTNSHKELSNSIAFDIRSQDFHYLIGVLMDDIIFGDKL 607
           ++++DYT+I   + ++    + QT+  +E+ N++AF+++ +   LI +L +DI  G+KL
Sbjct: 122 NQMLDYTIIAHESFEIIRHSVYQTDD-REVENALAFEVKNDETDKLILLLSEDIRVGEKL 180

Query: 608 YLMQGIKCKGKVIVYDKNNGKMSRLK 633
            L+ G K +GK +VYDK N +M RL+
Sbjct: 181 CLVDGTKMRGKCLVYDKINERMIRLQ 206
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2361> which encodes the amino acid sequence <SEQ ID 2362>. Analysis of this protein sequence reveals the following:

```
    Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.55   Transmembrane  477 - 493 ( 477 - 493)

----- Final Results -----
               bacterial membrane --- Certainty=0.2020(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:

```
>GP:CAB01924 GB:Z79691 OrfA [Streptococcus pneumoniae]
 Identities = 75/207 (36%), Positives = 125/207 (60%), Gaps = 2/207 (0%)

Query: 434 SWAYLSYPKSNYCHLRQNGHVYFIEGSYQTQFSDRNNYQHDRQILILPPGIFLIIDTIQA 493
           SW Y  YP S +CH ++    +++IEG+Y +   D   Y H R+IL+L   ++L++D I+
Sbjct: 2   SWEYEYYPHSLFCHHKEREGMHYIEGAYWSAEPDLP-YLHKRKILMLVEDVWLLVDDIRC 60
```

```
Query: 494 QGNHCLVSQFILDNHLDVKTDHLSDLRLISDCPFTIEETILSKKYNQYLTSHKLIKRKPF 553
           QG H  ++QFILD  +  +   ++ LRL S+  F +E+TI+S KYN+   S KL KR+ F
Sbjct: 61  QGQHEALTQFILDKDVTYQDGKINQLRLWSEVDFDLEDTIISPKYNELERSSKLTKRQFF 120

Query: 554 KDKGCTSTLLVPDDTKVTPLTPLQTGKRNPIETALSWHLKGKQFDYSICVLQEDLIKGEK 613
           +++    T++  +  ++   +  QT  R  +E AL++ +K  + D  I +L ED+  GEK
Sbjct: 121 ENQMLDYTIIAHESFEIIRHSVYQTDDRE-VENALAFEVKNDETDKLILLLSEDIRVGEK 179

Query: 614 LVLLNSHKIRGKVVVVINHITNEIIRLK 640
           L L++  K+RGK +V + I   +IRL+
Sbjct: 180 LCLVDGTKMRGKCLVYDKINERMIRLQ 206
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 282/631 (44%), Positives = 414/631 (64%), Gaps = 2/631 (0%)

Query: 6   YNKFKD-FDREFCQKYIKTYQSNAYQEMKASVNLMMRNTFVFNDNWDMEPCSKAYCLDPL 64
           + +FK+  + +FC+ Y+  YQ+++Y + K   +L++ NTF+F DNWDMEPC   Y LDP+
Sbjct: 11  FARFKETVNPDFCRNYLLDYQTDSYADQKRIADLLLTNTFLFEDNWDMEPCHIPYHLDPI 70

Query: 65  EWDKPVTDDPEWLYMLNRQTYLFKFLVVYIVEGDKSYLRQMKYFMYHWIDCQFTLKPEGA 124
            W + V DDPEW +MLNRQTYL K ++VY+VE D+ YL   K F+ +WI+    L P+G
Sbjct: 71  TWQEAVIDDPEWNFMLNRQTYLQKLILVYLVERDERYLLTAKGFILNWIESAIPLDPKGL 130

Query: 125 VSRTIDTGIRCMSWLKVLIFLDYFGLITETKKIKLLTSLREQITYMRDYYREKDSLSNWG 184
            +RT+DTGIRC +W+K LI+L+ F  +T+ ++  +L SL +Q+ ++    Y +K SLSNWG
Sbjct: 131 ATRTLDTGIRCFAWVKCLIYLNLFNALTKQEESLILASLEKQLQFLHANYLDKYSLSNWG 190

Query: 185 ILQTTAILACLYYYEDELNLPEIQSFAEEELLLQIKLQILDDGSQYEQSIMYHVEVLKSL 244
           ILQTTAIL    Y+  +L++    +FA +EL  QI LQIL+DGSQ+EQS MYHVEVLK+L
Sbjct: 191 ILQTTAILLADAYFGSDLDIAAATAFARKELTQQIALQILEDGSQFEQSTMYHVEVLKAL 250

Query: 245 MELVILAPKYYLPLEETIEKMVTYLIAMTGPDYCQLAIGDSDVTDTRDILTLATLVLKSS 304
           +EL  L P Y   L  T+  M  YL+ MTGPD+ Q+ +GDSDVTDTRDILTLA  +L+
Sbjct: 251 LELTALVPDYLPQLRPTLLAMSDYLLKMTGPDHKQIPLGDSDVTDTRDILTLAATILEEP 310

Query: 305 KTKSFSFDNVNLETLLLFGKPSIYLFEEIPRATIGESAYLFPDSGHVCLRDDRRYIFFKN 364
             K+ +F  +++++LLL G+  ++ FE++P  T+   A+ F  SGH+ +  + Y+FFKN
Sbjct: 311 HLKAAAFPTLDIDSLLLLGEKGVHTFEQLPVQTLPTFAHHFEHSGHITINQENYYLFFKN 370

Query: 365 GPFGSAHTHSDNNSVCLYDKKKPIFIDAGRYTYKEEQLRYDFKRSTSHSTCTLDGQPLEM 424
           GP GS+HTHSD NS+CLY K +P+F DAGRYTYKEE LRY  K ++ HST  L+ Q  E
Sbjct: 371 GPIGSSHTHSDQNSLCLYYKGQPLFCDAGRYTYKEEPLRYALKSASHHSTAFLEEQLPEQ 430

Query: 425 IKDSWTYNSYPKCDYCQLTSKDRYHLVEGQLHVQRAS-DIYYHKRWLLTLPQAITLVIDK 483
           I  SW Y SYPK +YC L      + +EG    Q +  + Y H R +L LP  I L+ID
Sbjct: 431 IDSSWAYLSYPKSNYCHLRQNGHVYFIEGSYQTQFSDRNNYQHDRQILILPPGIFLIIDT 490

Query: 484 VSCPGEHVLTNQYILDDQVIYENGFVNDLKLVSPTTFNLEDCLISKRYNQLTESHKLVKK 543
           +  G H L +Q+ILD+ +  +   ++DL+L+S   F +E+ ++SK+YNQ   SHKL+K+
Sbjct: 491 IQAQGNHCLVSQFILDNHLDVKTDHLSDLRLISDCPFTIEETILSKKYNQYLTSHKLIKR 550

Query: 544 IKFVDEVMDYTLIVDRNCQVKYVPLVQTNSHKELSNSIAFDIRSQDFHYLIGVLMDDIIF 603
             F D+    TL+V  + +V  +  +QT     +  ++++ ++ F Y I VL +D+I
Sbjct: 551 KPFKDKGCTSTLLVPDDTKVTPLTPLQTGKRNPIETALSWHLKGKQFDYSICVLQEDLIK 610

Query: 604 GDKLYLMQGIKCKGKVIVYDKNNGKMSRLKN 634
           G+KL L+   K +GKV+V +    ++ RLK+
Sbjct: 611 GEKLVLLNSHKIRGKVVVVINHITNEIIRLKH 641
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 769

A DNA sequence (GBSx0817) was identified in *S.agalactiae* <SEQ ID 2363> which encodes the amino acid sequence <SEQ ID 2364>. This protein is predicted to be RegR (kdgR). Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2545(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 GENPEPT database.

```
>GP:CAB01925 GB:Z79691 RegR [Streptococcus pneumoniae]
 Identities = 222/333 (66%), Positives = 279/333 (83%)

Query: 1   MSKKMTINDIAQLSKTSKTTVSFFLNQKFEKMSDETRQRIQEVIDETGYRPSTIARSLNS 60
           M KK+TI DIA++++TSKTTVSF+LN K+EKMS ETR++I++VI ET Y+PS +ARSLNS
Sbjct: 1   MEKKLTIKDIAEMAQTSKTTVSFYLNGKYEKMSQETREKIEKVIHETNYKPSIVARSLNS 60

Query: 61  KKTKLLGVLIGDITNTFSNQIVKGIEHITKQKGYQIIVGNSNYDAKSEEDYIENMLNLGV 120
           K+TKL+GVLIGDITN+FSNQIVKGIE I  Q GYQ+++GNSNY  +SE+ YIE+ML LGV
Sbjct: 61  KRTKLIGVLIGDITNSFSNQIVKGIEDIASQNGYQVMIGNSNYSQESEDRYIESMLLLGV 120

Query: 121 DGFIIQPTSNFRKYSRILKEKKKPMVFFDSQLYEHKTSWVKANNYDAVYDMTQECLNRGY 180
           DGFIIQPTSNFRKYSRI+ EKKK MVFFDSQLYEH+TSWVK NNYDAVYDMTQ C+ +GY
Sbjct: 121 DGFIIQPTSNFRKYSRIIDEKKKKMVFFDSQLYEHRTSWVKTNNYDAVYDMTQSCIEKGY 180

Query: 181 KKFIMITADTSLLSTRIERASGFMDALKDNGFGYDTLVIEDDDHSKSDIEDFLKAVVPDK 240
           + F++ITADTS LSTRIERASGF+DAL D     + +L IED    +    I++FL+  +
Sbjct: 181 EYFLLITADTSRLSTRIERASGFVDALTDANMRHASLTIEDKHTNLEQIKEFLQKEIDPD 240

Query: 241 EETLVFAPNCWALPMVFTAMKNLNFDMPRVGLVGFDNIEWTDFSSPKVSTIVQPAYEEGE 300
           E+TLVF PNCWALP+VFT +K LN+++P+VGL+GFDN EWT FSSP VST+VQP++EEG+
Sbjct: 241 EKTLVFIPNCWALPLVFTVIKELNYNLPQVGLIGFDNTEWTCFSSPSVSTLVQPSFEEGQ 300

Query: 301 QVAQILINRIEGDDSVDNQQIVDCQMFWKESTF 333
           Q  +ILI++IEG +  + QQ++DC + WKESTF
Sbjct: 301 QATKILIDQIEGRNQEERQQVLDCSVNWKESTF 333
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2365> which encodes the amino acid sequence <SEQ ID 2366>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2928(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 214/333 (64%), Positives = 266/333 (79%), Gaps = 2/333 (0%)

Query: 1   MSKKMTINDIAQLSKTSKTTVSFFLNQKFEKMSDETRQRIQEVIDETGYRPSTIARSLNS 60
           M +K+TI DIA+L+KTSKTTVSF+LN +F+KMS+ET+ RI E I  T Y+PS  ARSLN+
Sbjct: 13  MQRKVTIKDIAELAKTSKTTVSFYLNGRFDKMSEETKNRISESIKATNYKPSIAARSLNA 72

Query: 61  KKTKLLGVLIGDITNTFSNQIVKGIEHITKQKGYQIIVGNSNYDAKSEEDYIENMLNLGV 120
           K TKL+GV+IGDITN+FSNQIVKGIE    ++ GYQII+GNSNYD   E++ IE MLNLGV
Sbjct: 73  KSTKLIGVVIGDITNSFSNQIVKGIESKAQEFGYQIIIGNSNYDPSREDELIEKMLNLGV 132
```

```
Query: 121 DGFIIQPTSNFRKYSRILKEKKKPMVFFDSQLYEHKTSWVKANNYDAVYDMTQECLNRGY 180
           DGFIIQPTSNFRKYSRI+  KKK +VFFDSQLYEH+T+WVK NNYDAVYD  Q+C+++GY
Sbjct: 133 DGFIIQPTSNFRKYSRIIDIKKKKVVFFDSQLYEHRTNWVKTNNYDAVYDTIQQCIDKGY 192

Query: 181 KKFIMITADTSLLSTRIERASGFMDALKDNGFGYDTLVIEDDDHSKSDIEDFLKAVVPDK 240
           + FIMIT + +LLSTRIERASGF+D L+ N   +  ++I+++  S   I  FL+  +  K
Sbjct: 193 EHFIMITGNPNLLSTRIERASGFIDVLEANHLTHQEMIIDENQTSSEAIAQFLQGSLTKK 252

Query: 241 EETLVFAPNCWALPMVFTAMKNLNFDMPRVGLVGFDNIEWTDFSSPKVSTIVQPAYEEGE 300
             +LVF PNCWALP VFTAMK+L F++P +GLVGFDNIEWT FSSP ++TI+QPAYEEGE
Sbjct: 253 --SLVFVPNCWALPKVFTAMKSLKFNIPEIGLVGFDNIEWTKFSSPTLTTIIQPAYEEGE 310

Query: 301 QVAQILINRIEGDDSVDNQQIVDCQMFWKESTF 333
           Q  +ILI+ IEG      QQI DCQ+ W+ESTF
Sbjct: 311 QATKILIDDIEGHSQEAKQQIFDCQVNWQESTF 343
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 770

A DNA sequence (GBSx0818) was identified in *S.agalactiae* <SEQ ID 2367> which encodes the amino acid sequence <SEQ ID 2368>. This protein is predicted to be polypeptide defromylase (def-1). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2339(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 GENPEPT database.

```
>GP:CAC15392 GB:AJ278785 polypeptide deformylase [Streptococcus pneumoniae]
 Identities = 169/204 (82%), Positives = 192/204 (93%), Gaps = 1/204 (0%)

Query: 1   MSAIDKLVKASHLIDMNDIIREGNPTLRKVAEEVTFPLSEKEEILGEKMMQFLKHSQDPI 60
           MSAI+++ KA+HLIDMNDIIREGNPTLR +AEEVTFPLS++E ILGEKMMQFLKHSQDP+
Sbjct: 1   MSAIERITKAAHLIDMNDIIREGNPTLRAIAEEVTFPLSDQEIILGEKMMQFLKHSQDPV 60

Query: 61  MAEKLGLRGGVGLAAPQLDISKRIIAVLVPNVEDAQGNPPKEAYSLQEVMYNPKVVSHSV 120
           MAEK+GLRGGVGLAAPQLDISKRIIAVLVPN+ + +G  P+EAY L+ +MYNPK+VSHSV
Sbjct: 61  MAEKMGLRGGVGLAAPQLDISKRIIAVLVPNIVE-EGETPQEAYDLEAIMYNPKIVSHSV 119

Query: 121 QDAALSDGEGCLSVDREVPGYVVRHARVTIEYFDKTGEKHRLKLKGYNSIVVQHEIDHID 180
           QDAAL +GEGCLSVDR VPGYVVRHARVT++YFDK GEKHR+KLKGYNSIVVQHEIDHI+
Sbjct: 120 QDAALGEGEGCLSVDRNVPGYVVRHARVTVDYFDKDGEKHRIKLKGYNSIVVQHEIDHIN 179

Query: 181 GIMFYDRINEKNPFAVKEGLLILE 204
           GIMFYDRINEK+PFAVK+GLLILE
Sbjct: 180 GIMFYDRINEKDPFAVKDGLLILE 203
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2369> which encodes the amino acid sequence <SEQ ID 2370>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1745(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 160/204 (78%), Positives = 186/204 (90%)

Query: 1   MSAIDKLVKASHLIDMNDIIREGNPTLRKVAEEVTFPLSEKEEILGEKMMQFLKHSQDPI 60
           MSA DKL+K SHLI M+DIIREGNPTLR VA+EV+ PL +++ +LGEKMMQFLKHSQDP+
Sbjct: 1   MSAQDKLIKPSHLITMDDIIREGNPTLRAVAKEVSLPLCDEDILLGEKMMQFLKHSQDPV 60

Query: 61  MAEKLGLRGGVGLAAPQLDISKRIIAVLVPNVEDAQGNPPKEAYSLQEVMYNPKVVSHSV 120
           MAEKLGLR GVGLAAPQ+D+SKRIIAVLVPN+ D +GNPPKEAYS QEV+YNPK+VSHSV
Sbjct: 61  MAEKLGLRAGVGLAAPQIDVSKRIIAVLVPNLPDKEGNPPKEAYSWQEVLYNPKIVSHSV 120

Query: 121 QDAALSDGEGCLSVDREVPGYVVRHARVTIEYFDKTGEKHRLKLKGYNSIVVQHEIDHID 180
           QDAALSDGEGCLSVDR V GYVVRHARVT++Y+DK G++HR+KLKGYN+IVVQHEIDHI+
Sbjct: 121 QDAALSDGEGCLSVDRVVEGYVVRHARVTVDYYDKEGQQHRIKLKGYNAIVVQHEIDHIN 180

Query: 181 GIMFYDRINEKNPFAVKEGLLILE 204
           G++FYDRIN KNPF  KE LLIL+
Sbjct: 181 GVLFYDRINAKNPFETKEELLILD 204
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 771

A DNA sequence (GBSx0819) was identified in *S.agalactiae* <SEQ ID 2371> which encodes the amino acid sequence <SEQ ID 2372>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3620(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10177> which encodes amino acid sequence <SEQ ID 10178> was also identified.

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

```
>GP:AAC75224 GB:AE000305 putative transcriptional regulator
          [Escherichia coli K12]
 Identities = 58/191 (30%), Positives = 98/191 (50%)

Query: 37  DLQVITLTAGQSVCKQGEQLEYLHYIVKGRFKIVRRLFNGKEHILDIKTKPTLIGDIELL 96
           D ++    A   + ++G+Q  +L Y+ +GR ++   L NG+  ++D   P  IG+IEL+
Sbjct: 17  DTRLFHFLARDYIVQEGQQPSWLFYLTRGRARLYATLANGRVSLIDFFAAPCFIGEIELI 76

Query: 97  TNRQIVSSVIALEDLTVIQLSLKGRKEKLLTDATFLLKLSQELAQAFHDQNIKASTNLGY 156
                +V A+E+   + L +K  +  LL D  FL KL   L+   +  + + N  +
Sbjct: 77  DKDHEPRAVQAIEECWCLALPMKHYRPLLLNDTLFLRKLCVTLSHKNYRNIVSLTQNQSF 136

Query: 157 TVKELLASHILAIEEQGYFQLELSSLADSFGVSYRHLLRVIHDMVKEGLIQKEKPKYFIK 216
            +   LA+ IL  +E   + + +  A+  GVSYRHLL V+   + +GL+ K K  Y IK
Sbjct: 137 PLVNRLAAFILLSQEGDLYHEKHTQAAEYLGVSYRHLLYVLAQFIHDGLLIKSKKGYLIK 196

Query: 217 NRFALESLNIQ 227
           NR  L  L ++
Sbjct: 197 NRKQLSGLALE 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2373> which encodes the amino acid sequence <SEQ ID 2374>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3809(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 23/63 (36%), Positives = 35/63 (55%), Gaps = 1/63 (1%)

Query: 146 QNIKASTNLGYTVKELLASHILAIEEQGYFQLELSSLADSFGVSYRHLLRVIHDMVKEGL 205
           QN+    N+ YTVKE  AS+ L  +      L L+ LA+ FG S RHL  V+    + + +
Sbjct: 3   QNV-CQQNITYTVKERFASYTLEAQANQEVHLNLTLLANRFGTSDRHLKHVLKQPIFQRI 61

Query: 206 IQK 208
           I++
Sbjct: 62  IER 64
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 772

A DNA sequence (GBSx0820) was identified in *S.agalactiae* <SEQ ID 2375> which encodes the amino acid sequence <SEQ ID 2376>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.24   Transmembrane  163 - 179 ( 159 - 185)
   INTEGRAL    Likelihood = -8.49   Transmembrane  204 - 220 ( 201 - 226)
   INTEGRAL    Likelihood = -7.80   Transmembrane  272 - 288 ( 269 - 296)
   INTEGRAL    Likelihood = -6.00   Transmembrane  333 - 349 ( 331 - 352)
   INTEGRAL    Likelihood = -5.41   Transmembrane   75 -  91 (  73 -  92)
   INTEGRAL    Likelihood = -4.94   Transmembrane  245 - 261 ( 240 - 262)
   INTEGRAL    Likelihood = -4.41   Transmembrane  362 - 378 ( 359 - 380)
   INTEGRAL    Likelihood = -4.14   Transmembrane   96 - 112 (  95 - 113)
   INTEGRAL    Likelihood = -2.44   Transmembrane  141 - 157 ( 141 - 158)
   INTEGRAL    Likelihood = -1.81   Transmembrane  302 - 318 ( 301 - 320)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 8659> which encodes amino acid sequence <SEQ ID 8660> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:     -3.52
GvH: Signal Score (-7.5): 0.340001
     Possible site: 25
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 11 value:  -9.24 threshold:  0.0
   INTEGRAL    Likelihood = -9.24   Transmembrane  134 - 150 ( 130 - 156)
   INTEGRAL    Likelihood = -8.60   Transmembrane   17 -  33 (  13 -  37)
   INTEGRAL    Likelihood = -8.49   Transmembrane  175 - 191 ( 172 - 197)
   INTEGRAL    Likelihood = -7.80   Transmembrane  243 - 259 ( 240 - 267)
   INTEGRAL    Likelihood = -6.00   Transmembrane  304 - 320 ( 302 - 323)
   INTEGRAL    Likelihood = -5.41   Transmembrane   46 -  62 (  44 -  63)
   INTEGRAL    Likelihood = -4.94   Transmembrane  216 - 232 ( 211 - 233)
   INTEGRAL    Likelihood = -4.41   Transmembrane  333 - 349 ( 330 - 351)
   INTEGRAL    Likelihood = -4.14   Transmembrane   67 -  83 (  66 -  84)
   INTEGRAL    Likelihood = -2.44   Transmembrane  112 - 128 ( 112 - 129)
```

```
  INTEGRAL    Likelihood = -1.81   Transmembrane  273 - 289 ( 272 - 291)
  PERIPHERAL  Likelihood =  3.45     193
modified ALOM score:   2.35

*** Reasoning Step: 3

----- 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>
```

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

```
>GP:CAB50057 GB:AJ248286 TRANSPORT PROTEIN, permease [Pyrococcus abyssi]
 Identities = 94/382 (24%), Positives = 173/382 (44%), Gaps = 30/382 (7%)

Query: 5   MEKLSLLSL-SLILLSTFSTSPALPQMISYY-RDKGLPSPQVELLFSIPSMAIIFILLIT 62
           MEKL +L L SL  +  +S   A+P +     +D G+ + ++ LL +   +    I +
Sbjct: 1   MEKLIILILISLGWIFNYSHRMAVPSLAPIIMKDLGINNAEIGLLMTSLLLPYSLIQVPA 60

Query: 63  PWLSKKLSEKHMIIFGLLLTALGGGLPVVSQNYLLVFVSRLLLGSGIGFINTRAISVISE 122
            ++  K+  K ++   +L  +L   L V++++Y  +   R L G   G     A ++ISE
Sbjct: 61  GYIGDKIGRKKLLTISILGYSLSSALIVLTRDYWDLVTVRALYGFFAGLYYAPATALISE 120

Query: 123 YYQGKERRKLLGLRGSFEVLGNA---GLTAL--VGLLLTFGWSKSFMIYFLALPILVLYL 177
            ++ ++   L      F ++G A   G+T L  V + LT  W  +F++  +   I+ + L
Sbjct: 121 LFRERKGSALFFMVGPAIGSGITPLIVVPVALTLSWRYAFLVLSIMSSIVGILL 175

Query: 178 VFAPKKVVKDTNDKIKTKGQKIPKADLTYIVALAILAGFVITINTGINLRIPLLVVEFGL 237
           + A K       + IK +G K       ++++LA   G         +   + LV   G+
Sbjct: 176 MVAIK------GEPIKVEGVKFKIPRGVFLLSLANFLGLGAFFAM-LTFLVSYLVSR-GV 227

Query: 238 GTPAQASLVLSAMMLMGIIAGMSFGQLIAMFHKQLIPICLVLFS-LTLLGVGLPSNLMVL 296
           G   +ASL+ S + L+GI+  +  G L     K  +  L S LT L + +PS L ++
Sbjct: 228 GME-KASLMFSMLSLVGILGSIIAGFLYDHLGKVSVLLAYALNSLLTFLVIVIPSPLFLI 286

Query: 297 TISAMASGFLYSL--MVTAVFSLVADRVEYSLVGSATTLVLVF-CNIGGASAAILLSCFD 353
            +  +    LYS+  ++TA  S  A R    +V     +V  F   IG       L+
Sbjct: 287 PLGLV----LYSVGGIMTAYTSEKASRENLGVVMGFVNMVGFFGATIGPYIVGFLIDRLG 342

Query: 354 HLLGQINAVFYVYAILSLAVGM 375
           + L  + +V   Y + ++ +G+
Sbjct: 343 YSLALL-SVPLAYLVSAVIIGL 363
```

There is also homology to SEQ ID 2378.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 773

A DNA sequence (GBSx0821) was identified in *S.agalactiae* <SEQ ID 2379> which encodes the amino acid sequence <SEQ ID 2380>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38   Transmembrane  171 - 187 ( 171 - 187)

----- 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>
```

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

```
>GP:CAB61731 GB:AL133220 putative oxidoreductase. [Streptomyces
          coelicolor A3(2)]
 Identities = 101/327 (30%), Positives = 169/327 (50%), Gaps = 12/327 (3%)

Query: 8   WATLGTGVIANEL-AQALEARGQKLYSVANRTYDKGLEFATKYGIQKVYDHIDQVFEDPE 66
           W  L TG +A    A  ++    ++ +VA+RT      FA ++GI + Y   + +  D +
Sbjct: 11  WGILATGGMAARFTADLVDLPDAEVVAVASRTEASAKTFAERFGIPRAYGGWETLARDED 70

Query: 67  VDIIYISTPHNTHISFLRKALANGKHVLCEKSITLNSTELKEAIDLAETNHVVLAEAMTI 126
           VD++Y++TPH+ H +     L  G++VLCEK  TLN+ E  E + LA  N V L EAM +
Sbjct: 71  VDVVYVATPHSAHRTAAGLCLEAGRNVLCEKPFTLNAREAAELVALARENGVFLMEAMWM 130

Query: 127 FHMPIYRQLKTLVDSGKLGPLKMIQMNFGSYKEYDMTNRFFSRDLAGGALLDIGVYALSC 186
           +  P+ R+LK LV  G +G ++ +Q +FG    +   +R       GGALLD+GVY +S
Sbjct: 131 YCNPLVRRLKELVADGAIGEVRSLQADFGLAGPFPAAHRLRDPAQGGGALLDLGVYPVSF 190

Query: 187 IRWFMSEAPHNITSQVTFAPTGVDEQVGILLTNPANEMATVSLSLHAKQPKRATIAYDKG 246
            +  + E P ++ ++   +  GVD Q G LL+   + +A++  S+    P  A+I   +G
Sbjct: 191 AQLLLGE-PTDVAARAVLSEEGVDLQTGALLSYGNDALASIHCSITGGTPNSASITGSEG 249

Query: 247 YIEL---FEYPRGQKAVITYTEDGHQDIL--EAGKTENALQYEVADMEEAV-SGKTNH-- 298
           I++   F +P     V+  T    Q+     A     +L++E  ++  A+ +G+T
Sbjct: 250 RIDVPNGFFFP--DHFVLHRTGRDPQEFRADPADGPRESLRHEAEEVMRALRAGETESPL 307

Query: 299 MYLNYTKDVMDIMTQLRQEWGFTYPEE 325
           + L+ T  VM  +  +R   G  YP E
Sbjct: 308 VPLDGTLAVMRTLDAIRDRVGVRYPGE 334
```

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

Example 774

A DNA sequence (GBSx0822) was identified in *S.agalactiae* <SEQ ID 2381> which encodes the amino acid sequence <SEQ ID 2382>. This protein is predicted to be oligopeptidase. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2881(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 GENPEPT database.

```
>GP:CAC14579 GB:AJ249396 oligopeptidase [Streptococcus thermophilus]
 Identities = 504/631 (79%), Positives = 563/631 (88%)

Query: 1   MIKYQDDFYQAVNGEWAKTAVIPDDKPRTGGFSDLADDIEALMLSTTDKWLADENKPSDT 60
           M + QDDFY A+NGEW KTAVIPDDKP TGGFSDLAD+IE LML TTD+WLA EN P +
Sbjct: 1   MTRLQDDFYHAINGEWEKTAVIPDDKPCTGGFSDLADEIEDLMLETTDQWLAGENVPDNA 60

Query: 61  ILNHFIAFHKMTADYQKREEVGVSPVLPLIEEYKGLQSFSEFASKVAEYELEGKPNEFPF 120
           IL +FI FH+MTADY +RE VG+ PV PLIEEYK L SFSEFASK+AEYE+ GKPNEFPF
Sbjct: 61  ILQNFIKFHRMTADYDRREAVGIEPVKPLIEEYKKLSSFSEFASKIAEYEMSGKPNEFPF 120

Query: 121 GVAPDFMNAQLNVLWAEAPGIILPDTTYYSEDNEKGKELLAFWRKSQEDLLPLFGLSEQE 180
            V+PDFMNAQLNVLWA+APGIILPDTTYY+EDNEKGKELL  WR+ QE+LL  +G + +E
Sbjct: 121 SVSPDFMNAQLNVLWADAPGIILPDTTYYTEDNEKGKELLEIWREMQEELLGKYGFTAEE 180

Query: 181 IKDILDKVLALDAKLAQYVLSREESSEYVKLYHPYNWEDFTKLAPELPLDAIFQKILGQK 240
           IKD+LDKV+ LDAKLA+YVLS EESSEYV+LYHPY+W DFTKLAPELPLD+IF +ILGQ
Sbjct: 181 IKDLLDKVIDLDAKLAKYVLSHEESSEYVELYHPYDWADFTKLAPELPLDSIFTEILGQV 240
```

```
Query: 241 PDKVIVPEERFWTEFASDYYSESNWELLKADLILSAANAYNAYLTDDIRIKSGVYSRALS 300
           PDKVIV EE FWTEFA++YYSE+NWELLKA L++ A  ++NAYLTD++R+ SG YSRALS
Sbjct: 241 PDKVIVSEESFWTEFAAEYYSEANWELLKAVLLIDATTSWNAYLTDELRVLSGKYSRALS 300

Query: 301 GTPQAMDKKKAAYYLASGPYNQALGLWYAGEKFSPEAKADVEHKIATMIDVYKSRLEKAD 360
           GTPQAMDKKKAA+YLA GPYNQALGLWYAGEKFSPEAKADVE K+ATMIDVYKSRL+ AD
Sbjct: 301 GTPQAMDKKKAAFYLAQGPYNQALGLWYAGEKFSPEAKADVEAKVATMIDVYKSRLQTAD 360

Query: 361 WLAQSTREKAIMKLNVITPHIGYPEKLPETYTKKIIDPKLSLVENATNLDKISIAYGWSK 420
           WLA  TREKAI KLNVITPHIGYPEKLPETY KKIID  LSLVENA  L +ISIA+ WSK
Sbjct: 361 WLAPETREKAITKLNVITPHIGYPEKLPETYDKKIIDENLSLVENAQKLVEISIAHSWSK 420

Query: 421 WNKPVDRSEWHMPAHMVNAYYDPQQNQIVFPAAILQEPFYALEQSSSANYGGIGAVIAHE 480
           WNKPVDRSEWHMPAHMVNAYYDPQQNQIVFPAAILQ PFY + QSSSANYGGIGAVIAHE
Sbjct: 421 WNKPVDRSEWHMPAHMVNAYYDPQQNQIVFPAAILQAPFYDIAQSSSANYGGIGAVIAHE 480

Query: 481 ISHAFDTNGASFDEHGSLNNWWTDEDFEAFKKLTDKVVEQFDGLESYGAKVNGKLTVSEN 540
           ISHAFDTNGASFDE+GSL NWWT++D+ AFK+ TDK+V+QF+GL+SYGAKVNGKLTVSEN
Sbjct: 481 ISHAFDTNGASFDENGSLKNWWTEDDYAAFKERTDKIVDQFEGLDSYGAKVNGKLTVSEN 540

Query: 541 VADLGGVACALEAAQRESDFSARDFFINFATIWRMKARDEYMQMLASVDVHAPAQWRTNI 600
           VADLGGVACALEAA+R+ DFS R+FFINFATIWR KAR+EYMQMLASVDVHAPA+WRTN+
Sbjct: 541 VADLGGVACALEAAKRDEDFSVREFFINFATIWRTKAREEYMQMLASVDVHAPAKWRTNV 600

Query: 601 TVTNFEEFHKEFDVKDGDNMWRPVEKRVIIW 631
            VTNF+EFHKEFDVK+GD MWR  E RVIIW
Sbjct: 601 IVTNFDEFHKEFDVKEGDGMWRAPEDRVIIW 631
```

Endopeptidases are often exposed antigens.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2383> which encodes the amino acid sequence <SEQ ID 2384>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2622(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 504/631 (79%), Positives = 564/631 (88%)

Query: 1   MIKYQDDFYQAVNGEWAKTAVIPDDKPRTGGFSDLADDIEALMLSTTDKWLADENKPSDT 60
           M  YQDDFYQAVNG+WA+TAVIPDDKPRTGGFSDLAD+IEALML TTD WLA EN P D
Sbjct: 1   MTTYQDDFYQAVNGKWAETAVIPDDKPRTGGFSDLADEIEALMLDTTDAWLAGENIPDDA 60

Query: 61  ILNHFIAFHKMTADYQKREEVGVSPVLPLIEEYKGLQSFSEFASKVAEYELEGKPNEFPF 120
           IL +F+ FH++ ADY KR+EVGVSP+LPLIEEY+ L+SFSEF + +A+YEL G PNEFPF
Sbjct: 61  ILKNFVKFHRLVADYAKRDEVGVSPILPLIEEYQSLKSFSEFVANIAKYELAGLPNEFPF 120

Query: 121 GVAPDFMNAQLNVLWAEAPGIILPDTTYYSEDNEKGKELLAFWRKSQEDLLPLFGLSEQE 180
            VAPDFMNAQLNVLWAEAP I+LPDTTYY E NEK +EL   WR+SQE LLP FG S +E
Sbjct: 121 SVAPDFMNAQLNVLWAEAPSILLPDTTYYEEGNEKAEELRGIWRQSQEKLLPQFGFSTEE 180

Query: 181 IKDILDKVLALDAKLAQYVLSREESSEYVKLYHPYNWEDFTKLAPELPLDAIFQKILGQK 240
           IKD+LDKV+ LD +LA+YVLSREE SEY KLYHPY W DF KLAPELPLD+IF+KILGQ
Sbjct: 181 IKDLLDKVIELDKQLAKYVLSREEGSEYAKLYHPYVWADFKKLAPELPLDSIFEKILGQV 240

Query: 241 PDKVIVPEERFWTEFASDYYSESNWELLKADLILSAANAYNAYLTDDIRIKSGVYSRALS 300
           PDKVIVPEERFWTEFA+ YYSE+NW+LLKA+LI+ AANAYNAYLTDDIR++SG YSRALS
Sbjct: 241 PDKVIVPEERFWTEFAATYYSEANWDLLKANLIVDAANAYNAYLTDDIRVESGAYSRALS 300

Query: 301 GTPQAMDKKKAAYYLASGPYNQALGLWYAGEKFSPEAKADVEHKIATMIDVYKSRLEKAD 360
```

```
            GTPQAMDK+KAA+YLA GP++QALGLWYAG+KFSPEAKADVE K+A MI+VYKSRLE AD
Sbjct: 301 GTPQAMDKQKAAFYLAQGPFSQALGLWYAGQKFSPEAKADVESKVARMIEVYKSRLETAD 360

Query: 361 WLAQSTREKAIMKLNVITPHIGYPEKLPETYTKKIIDPKLSLVENATNLDKISIAYGWSK 420
            WLA +TREKAI KLNVITPHIGYPEKLPETY KK+ID  LSLVENA NL KI+IA+ WSK
Sbjct: 361 WLAPATREKAITKLNVITPHIGYPEKLPETYAKKVIDESLSLVENAQNLAKITIAHTWSK 420

Query: 421 WNKPVDRSEWHMPAHMVNAYYDPQQNQIVFPAAILQEPFYALEQSSSANYGGIGAVIAHE 480
            WNKPVDRSEWHMPAH+VNAYYD QQNQIVFPAAILQEPFY+L+QSSSANYGGIGAVIAHE
Sbjct: 421 WNKPVDRSEWHMPAHLVNAYYDLQQNQIVFPAAILQEPFYSLDQSSSANYGGIGAVIAHE 480

Query: 481 ISHAFDTNGASFDEHGSLNNWWTDEDFEAFKKLTDKVVEQFDGLESYGAKVNGKLTVSEN 540
            ISHAFDTNGASFDEHGSLN+WWT ED+ AFK+ TDK+V QFDGLES+GAKVNGKLTVSEN
Sbjct: 481 ISHAFDTNGASFDEHGSLNDWWTQEDYAAFKERTDKIVAQFDGLESHGAKVNGKLTVSEN 540

Query: 541 VADLGGVACALEAAQRESDFSARDFFINFATIWRMKARDEYMQMLASVDVHAPAQWRTNI 600
            VADLGGVACALEAAQ E DFSARDFFINFATIWRMKAR+EYMQMLAS+DVHAP + RTN+
Sbjct: 541 VADLGGVACALEAAQSEEDFSARDFFINFATIWRMKAREEYMQMLASIDVHAPGELRTNV 600

Query: 601 TVTNFEEFHKEFDVKDGDNMWRPVEKRVIIW 631
            T+TNF+ FH+ FD+K+GD MWR  + RVIIW
Sbjct: 601 TLTNFDAFHETFDIKEGDAMWRAPKDRVIIW 631
```

SEQ ID 2382 (GBS193) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 3; MW 73kDa).

The GBS193-His fusion product was purified (Figure 196, lane 5) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 253). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 775

A DNA sequence (GBSx0823) was identified in *S.agalactiae* <SEQ ID 2385> which encodes the amino acid sequence <SEQ ID 2386>. This protein is predicted to be immunity protein (mccF-1). Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1627(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9433> which encodes amino acid sequence <SEQ ID 9434> was also identified.

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

```
>GP:AAB84435 GB:AF027868 YocD [Bacillus subtilis]
 Identities = 114/270 (42%), Positives = 170/270 (62%), Gaps = 4/270 (1%)

Query: 1   MSFSKHYLENDILYSASITSRVEDLHEAFADPSVDAILATIGGFNSNELLPYLDYDLISK 60
            ++ ++H  E +   S+SI SRV DLH AF DP V AIL T+GGFNSN+LL YLDY+ I +
Sbjct: 43  VTIAEHANECNEFDSSSIESRVHDLHAAFFDPGVKAILTTLGGFNSNQLLRYLDYEKIKR 102

Query: 61  NPKIICGYSDSTAFLNAIFAKAKIQTYMGPAYSSFKMKEGQPYQTQAWLT-AMTENHYEL 119
            +PKI+CGYSD TA  NAI+ K  + TY GP +S+F MK+G  Y  + +L+   +++ +E+
Sbjct: 103 HPKILCGYSDITALCNAIYQKTGLVTYSGPHFSTFAMKKGLDYTEEYFLSCCASDDPFEI 162
```

```
Query: 120 WPSEEWSSDPWYDPSKPRQFFPTEWK-IYNHGKASGTIIGGNLSTFGLLRGTPYAPKIER 178
           PS EWS D W+    + R+F+P      +   G A GT+IGGNL T  LL+GT Y P+ E
Sbjct: 163 HPSSEWSDDRWFLDQENRRFYPNNGPVVIQEGYAEGTLIGGNLCTLNLLQGTEYFPETEH 222

Query: 179 YVLLIEEAEESNFYEFDRNLAAI--LQAYPHPQAILMGRFPKECGMTPQVFEYILSKHAI 236
           +LLIE+    S+ + FDR+L ++  L A+ H +AIL+GRF K    ++  + + ++
Sbjct: 223 TILLIEDDYMSDIHMFDRDLQSLIHLPAFSHVKAILIGRFQKASNVSIDLVKAMIETKKE 282

Query: 237 FKEIPVIYDMDFAHTQPLLTVTIGAELSVD 266
              IP+I +++  HT P+ T  IG    ++
Sbjct: 283 LSGIPIIANINAGHTSPIATFPIGGTCRIE 312
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2387> which encodes the amino acid sequence <SEQ ID 2388>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1162(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 75/252 (29%), Positives = 125/252 (48%), Gaps = 22/252 (8%)

Query: 34  VDAILATIGGFNSNELLPYLDYDLISKNPKIICGYSDSTAFLNAIFAKAKIQTYMGPAYS 93
           VD I+ +IGG+NSN +L Y+DYDL  +   I  GYSD+TA   A++ K    TY+  +
Sbjct: 1   VDVIMTSIGGYNSNSVLKYIDYDLFKQKFPIFIGYSDTTALALALYKKTGCITYLSQSVI 60

Query: 94  SFKMKEGQP----------YQTQAWLTAMTENHYELWPSEEWSSDPWYDPSKPRQFFPTE 143
           S    E +P          +  Q+    +     ++W ++EW +  W    + ++    E
Sbjct: 61  S-NFGEFEPFNELNYFYFDFMLQSKCETLMVQIPDVW-TDEWIN--WETYERTKKTNKNE 116

Query: 144 WKIYNHGKASGTIIGGNLSTFGLLRGTPYAPKIERYVLLIEEAEESNFYEFDRNLA--AI 201
           W I+N G+ +GT+IGGNL T   + GT Y PKI    +L+ E   ++     RN    A+
Sbjct: 117 WIIFNKGEFNGTLIGGNLDTIVGIIGTEYMPKITEDTILLLEDVYTDLGRLYRNFTTLAL 176

Query: 202 LQAYPHPQAILMGRFPKECGMTPQVFEYILSKHAIFKEIPVIYDMDFAHTQPLLTVTIGA 261
              +     +++ +F +  G    V   I+++    ++IP++ + D  HT P   + IG
Sbjct: 177 HGIFDKIGGLIISKF-ETIGENSDVINDIINEFVGHRKIPILLNFDCGHTHPSCLMPIGG 235

Query: 262 ELSVDTTTLSLS 273
           ++     TLSLS
Sbjct: 236 KI-----TLSLS 242
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 776

A DNA sequence (GBSx0824) was identified in *S.agalactiae* <SEQ ID 2389> which encodes the amino acid sequence <SEQ ID 2390>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3112(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 777

A DNA sequence (GBSx0825) was identified in *S.agalactiae* <SEQ ID 2391> which encodes the amino acid sequence <SEQ ID 2392>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.6171(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10175> which encodes amino acid sequence <SEQ ID 10176> was also identified.

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

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 778

A DNA sequence (GBSx0826) was identified in *S.agalactiae* <SEQ ID 2393> which encodes the amino acid sequence <SEQ ID 2394>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.19   Transmembrane   83 -  99 (  80 - 113)
   INTEGRAL    Likelihood = -9.71   Transmembrane    4 -  20 (   1 -  24)
   INTEGRAL    Likelihood = -9.45   Transmembrane  315 - 331 ( 307 - 337)
   INTEGRAL    Likelihood = -8.33   Transmembrane  186 - 202 ( 180 - 210)
   INTEGRAL    Likelihood = -7.75   Transmembrane  233 - 249 ( 227 - 255)
   INTEGRAL    Likelihood = -3.98   Transmembrane  390 - 406 ( 382 - 407)
   INTEGRAL    Likelihood = -3.61   Transmembrane   27 -  43 (  27 -  45)
   INTEGRAL    Likelihood = -3.29   Transmembrane  107 - 123 ( 105 - 125)
   INTEGRAL    Likelihood = -1.75   Transmembrane  273 - 289 ( 273 - 290)

----- Final Results -----
             bacterial membrane --- Certainty=0.5076(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 GENPEPT database.

```
>GP:CAB15347 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 174/524 (33%), Positives = 275/524 (52%), Gaps = 13/524 (2%)

Query: 1   MEETILIVSFLLFLILSNVINRIFPKLPLPFIQLVFGILSGLVFHKSQVHIDPELFLAFV 60
           M+  ++++  L  + +SN++NR  P +P+P IQ+  GIL+           ++ ELF
Sbjct: 1   MDIFLVVLVLLTIIAISNIVNRFIPFIPVPLIQVALGILAASFPQGLHFELNTELFFVLF 60

Query: 61  IAPLNFREGQESDIGSFIKYRAIILYLILPTVFLTAIVVGYVAGHLLPVSLPLAACFALG 120
           IAPL F +G+ +         RA IL L L  VF T IV GY    ++P ++PLAA F L
Sbjct: 61  IAPLLFNDGKRTPRAELWNLRAPILLLALGLVFATVIVGGYTIHWMIP-AIPLAAAFGLA 119
```

```
Query: 121 AALGPTDAVAFISIAKRFQFPKRAENILKLEGLLNDASGLVSFQFALTALVTGYFSLAKA 180
           A L PTD VA  +++ R + PK    +L+ EGL+NDASGLV+F+FA+ A VTG FSLA+A
Sbjct: 120 AILSPTDVVAVSALSGRVKMPKGILRLLEGEGLMNDASGLVAFKFAIAAAVTGAFSLAQA 179

Query: 181 SLKLALAIMGGFLIGLLFAFLMRLCLTVLEKFDAADVTGALLLELTLPFVAYFVADLLGF 240
           ++       +GG L G++ +FL+       L +    DVT  +L+++  PFV Y  A+ +G
Sbjct: 180 AVSFVFISLGGLLCGVVISFLIIRFRLFLRRLGMQDVTMHMLIQILTPFVIYLAAEEIGV 239

Query: 241 SAIIAVVVAGVMQANRLKKVTLFDAQVDRVTSVIWETLNFILNGLVFLIFGRELTRIIGP 300
           S I+AVV  G+  A    ++      ++  V+S  W  + FILNGLVF+I G ++  +I
Sbjct: 240 SGILAVVAGGITHAVEQDRLESTMIKLQIVSSSTWNIILFILNGLVFVILGTQIPDVISV 299

Query: 301 LLTSNAYSNFDLISIVVLVTCTLFLVRFLAVSCFY--AWRSFKYHKSFKKYWREIQLLTF 358
           +    A SN  +I  ++++T TL L+RFL V  F+   W   K    +K   R   L++
Sbjct: 300 IFNDTAISNMKVIGYILVITFTLMLLRFLWVLFFWNGKWFFNKDQNIYKPGLRSTLLISI 359

Query: 359 SGVKGSVSIATILLLPKHSVIGE--LGYSLILFTVGAVTLMSFLTGLLVLPKLAPPLQVK 416
           SGV+G+V++A    +P     G       +LILF    V L + +   +VLP L    +
Sbjct: 360 SGVRGAVTLAGSFSIPYFLEDGTPFPERNLILFLAAGVILCTLVIATVVLPILTEKEEED 419

Query: 417 DD-----YLIRLSILTKVLSVLEEDGKSSENQASFYAVIDNYNSRIRHLILEQ--ESSDI 469
           ++         R  ++   L  ++ED   +   AS  AVI  YN ++++L  +Q   S+ I
Sbjct: 420 EERNKKLLTARRKLIKTALQTIKEDMNETNKTASL-AVIAEYNEKMKNLRFQQYTSSNRI 478

Query: 470 KKDLAELQLMMLSIESDGLEAAYRYGNISIKEYRIYQRYLKYLE 513
           KK  +++   + E + L     G+I  +   + Q     LE
Sbjct: 479 KKHERKVRAQGVKAEQEALMKMLERGDIPEETANVLQERFNELE 522
```

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

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

Example 779

A DNA sequence (GBSx0827) was identified in *S.agalactiae* <SEQ ID 2395> which encodes the amino acid sequence <SEQ ID 2396>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3494(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 780

A DNA sequence (GBSx0828) was identified in *S.agalactiae* <SEQ ID 2397> which encodes the amino acid sequence <SEQ ID 2398>. This protein is predicted to be integrase (phage-relatedpr). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
bacterial cytoplasm --- Certainty=0.5094(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10173> which encodes amino acid sequence <SEQ ID 10174> was also identified.

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

```
>GP:AAF12706 GB:AF066865 integrase [bacteriophage TPW22]
 Identities = 171/353 (48%), Positives = 253/353 (71%), Gaps = 1/353 (0%)

Query: 21  MASYRKRENGLWEYRISYKTIDGKYKRKEKGGFKTKKLAQAAAIEIEKKLTQNILTNDEV 80
           MA++RKR    W++R+SYK  +G+YK+ EKGG+KTKK A+AAA E +K+L  +   ++++
Sbjct: 1   MANFRKRGK-TWQFRLSYKDNNGEYKKFEKGGYKTKKEAEAAADEAKKRLNNHSEFDNDI 59

Query: 81  TLYDFVKTWSEVYKRPYVKDKTWETYSKNFKHIKNYFQELKVKDITPLYYQKKLNEFGEK 140
           +LYDF + W++VYK+P+V + TW TY +    I  Y ++  + +ITP +YQ  LN+
Sbjct: 60  SLYDFFEKWAKVYKKPHVTEATWRTYKRTLNLIDKYIKDKPIAEITPTFYQAVLNKMSLL 119

Query: 141 YAQETLEKFHYQIKGAMKVAVREQVVTFNFAEGAKVKSQVEPKNEEEDFLEEREYKALLA 200
           Y QE+L+KF++QIK AMK+AV E+V++ NFA+  K KS++  +  EE +L   EY  LLA
Sbjct: 120 YRQESLDKFYFQIKSAMKIAVHEKVISENFADFTKAKSKLAARPVEEKYLHADEYLKLLA 179

Query: 201 LTRENIQYVSYFTLYLLAVTGLRFSEAMGLTWSDIDFKNGILDINKSFDYSNTQDFADLK 260
           +  E ++Y SYF  YL AVTG+RF+E +GLTWS +DF    + I +++DYS T +FA+ K
Sbjct: 180 IAEEKMEYTSYFACYLTAVTGMRFAELLGLTWSHVDFDKKEISIQRTWDYSITNNFAETK 239

Query: 261 NESSKRKVPIDSNTIDILREYKKNHWQANIKNRVCFGVSNSACNKLIKKIVGRKVRNHSL 320
           NESSKRK+PI S TI +L++YKK +W  N  +RV + +SN+  NK IK I GRKV  HSL
Sbjct: 240 NESSKRKIPISSKTIKLLKKYKKEYWHENKYDRVIYNLSNNGLNKTIKVIAGRKVHPHSL 299

Query: 321 RHTYASFLILNGVDIVTISKLLGHESPDITLKVYTHQMEALAERNFEKIKNIF 373
           RH++AS+LI  G+D++T+SKLLGHE+ ++TLKVY HQ++ + + N + I+ IF
Sbjct: 300 RHSFASYLIYKGIDLLTVSKLLGHENLNVTLKVYAHQLKEMEQENNDVIRKIF 352
```

There is also homology to SEQ ID 578.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 781

A DNA sequence (GBSx0829) was identified in *S.agalactiae* <SEQ ID 2399> which encodes the amino acid sequence <SEQ ID 2400>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3377(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 782

A DNA sequence (GBSx0830) was identified in *S.agalactiae* <SEQ ID 2401> which encodes the amino acid sequence <SEQ ID 2402>. This protein is predicted to be homology to cI-like repressor. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0827(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 GENPEPT database.

```
>GP:AAD44097 GB:AF115103 orf122 gp [Streptococcus thermophilus
          bacteriophage Sfi21]
 Identities = 57/125 (45%), Positives = 77/125 (61%), Gaps = 5/125 (4%)

Query: 3   MKLDQLCKEFGVELCLFDASDWHSSGFYNPITKVLGVDVNLSEQEQKQVALHELQHKNHF 62
           M   +L ++FGV LC F +S W   GF +P+ +V+ ++ +L  + + +V LHEL H  H
Sbjct: 1   MNESELLEQFGVSLCEFSSSQWTRDGFLDPVNRVVYINRDLPTERRLKVLLHELGHLEHD 60

Query: 63  PYQYQLFRERCELDANRNMIHHLLKEELEIAEDHTQFNYLVFMEKYKLKTIADEAMIKEE 122
           P QY+  RE+ E  ANRNMIH LLK      E+   FNY+ FMEKY L TI DE  +K E
Sbjct: 61  PKQYERLREKYEAQANRNMIHELLKN-----ENLDNFNYVHFMEKYNLTTICDETFVKNE 115

Query: 123 YLNLV 127
           YL L+
Sbjct: 116 YLKLI 120
```

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

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

Example 783

A DNA sequence (GBSx0831) was identified in *S.agalactiae* <SEQ ID 2403> which encodes the amino acid sequence <SEQ ID 2404>. This protein is predicted to be EpsR protein. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4692(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 GENPEPT database.

```
>GP:AAF12710 GB:AF066865 repressor protein [bacteriophage TPW22]
 Identities = 36/101 (35%), Positives = 62/101 (60%), Gaps = 7/101 (6%)

Query: 4   LIDRIRELSNKKGMSLNDLEDTLGYSRNSLYSLNE-NSKMGKPKEIAQYFNVSLDYLLGL 62
           L ++I+EL+++K +S+  +E+ LG++  ++     + N  + K K++A+YFNVS+D+LLGL
Sbjct: 3   LYEKIKELASQKNVSIRQVEEKLGFANGTIRQWGKKNPGINKVKDVAKYFNVSVDFLLGL 62

Query: 63  TDNPRIAS--DETAIIDGQVVDLREAAAHTMLFDGKPLDED 101
            DN R     D    +D   V+  E  +     FDGKPL ++
Sbjct: 63  DDNQRKKEPVDLADFVDDNKVNWDEWVS----FDGKPLSDE 99
```

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

Example 784

A DNA sequence (GBSx0832) was identified in *S.agalactiae* <SEQ ID 2405> which encodes the amino acid sequence <SEQ ID 2406>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4079(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 785

A DNA sequence (GBSx0833) was identified in *S.agalactiae* <SEQ ID 2407> which encodes the amino acid sequence <SEQ ID 2408>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

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

A related GBS nucleic acid sequence <SEQ ID 10171> which encodes amino acid sequence <SEQ ID 10172> was also identified.

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

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 786

A DNA sequence (GBSx0834) was identified in *S.agalactiae* <SEQ ID 2409> which encodes the amino acid sequence <SEQ ID 2410>. This protein is predicted to be a replication initiation protein Rep (RC). Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3335(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 787

A DNA sequence (GBSx0835) was identified in *S.agalactiae* <SEQ ID 2411> which encodes the amino acid sequence <SEQ ID 2412>. This protein is predicted to be antirepressor. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3380(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 GENPEPT database.

```
>GP:BAA97816 GB:AB044554 antirepressor [Staphylococcus aureus
           prophage phiPV83]
 Identities = 70/153 (45%), Positives = 93/153 (60%), Gaps = 15/153 (9%)

Query: 3   EIFVFHGQEVRTVTINNEPWFVGKDVADILGYSKSRNAIALHVDEDDALKQGITDNLGRM 62
           + F F    VRTV I NEP+FVGKD+A+ILGY+++ NAI  HVD +D L    + + G+
Sbjct: 5   QTFNFKELPVRTVEIENEPYFVGKDIAEILGYARTDNAIRNHVDSEDKLTHQFSAS-GQN 63

Query: 63  QETIIINESGLYSLIL----SSKLPQVKE----FKRWVTSEVLPQIRQQGAYVPENLSDE 114
           +  IIINESGLYSLI      SK  +++E    FKRWVTS+VLP IR+ G Y  +N+ ++
Sbjct: 64  RNMIIINESGLYSLIFDASKQSKNEKIRETARKFKRWVTSDVLPAIRKHGIYATDNVIEQ 123

Query: 115 A------FIALFTGQKKLKEHQLALAQDVDYLK 141
                 I + T  KK KE  L L Q V+  K
Sbjct: 124 TLKDPDYIITVLTEYKKEKEQNLVLQQQVEVNK 156
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2413> which encodes the amino acid sequence <SEQ ID 2414>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4609(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 54/142 (38%), Positives = 73/142 (51%), Gaps = 7/142 (4%)

Query: 11  EVRTVTINNEPWFVGKDVADILGYSKSRNAIALHVDEDDALKQGITDNLGRMQETIIINE 70
           EVRT TINN+ +F   D   IL  S  R  I   +++D      I D+LGR Q+   INE
Sbjct: 13  EVRTATINNQIYFNLNDCCQILELSNPRKTIE-RLNKDGVTTSDIIDSLGRTQQANFINE 71

Query: 71  SGLYSLILSSKLPQVKEFKRWVTSEVLPQIRQQGAYVPENLSDEA------FIALFTGQK 124
           S  Y L+  S+ P+ ++F  WVTSEVLP IR+ GAY+ E   ++A       I L    K
Sbjct: 72  SNFYKLVFQSRKPEAEKFADWVTSEVLPSIRKHGAYMTEQTLEQALTSPDFLIRLANELK 131

Query: 125 KLKEHQLALAQDVDYLKNEQPI 146
           + KE    L  +   L  E  +
```

```
Sbjct: 132 EEKERSRQLEAEKSILSVENMV 153
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 788

A DNA sequence (GBSx0836) was identified in *S.agalactiae* <SEQ ID 2415> which encodes the amino acid sequence <SEQ ID 2416>. This protein is predicted to be e11. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3281(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 GENPEPT database.

```
>GP:AAC27227 GB:AF009630 e11 [bacteriophage bIL170]
 Identities = 66/161 (40%), Positives = 93/161 (56%), Gaps = 13/161 (8%)

Query: 15  YQVSNLGRVRSIGRTVNAKQRTRKTKGRILKQSL-SSGYAIVTLSVNGLRKSIRVHRLVA 73
           Y+VSNLG+VR+I               GRILK  +  +GY +  L  N  +K++ +HR++A
Sbjct: 16  YEVSNLGKVRNI------------KSGRILKPWIVPNGYLMHQLCENNKKKNLFLHRIIA 63

Query: 74  EAFIPNPINKRTINHIDENKLNNRVDNLEWATDKENANHGNRTTKSSLGRCKPVEQFTLE 133
            AFI NP  K  +NHIDENKLNN ++NLEW T KEN  HG R  + +     K V Q  L
Sbjct: 64  TAFIDNPEEKPQVNHIDENKLNNDLNNLEWCTVKENNIHGTRMKRIAEKHFKKVIQLDLN 123

Query: 134 GEFINTFDSIKSASMKTGISSQRITATAMGHQKQTHGYKWR 174
               +N F+S+  A  +TG+S + I++   G +K    +KWR
Sbjct: 124 DNVLNEFESMVQAEQETGVSRRNISSCCNGKRKSAGRFKWR 164
```

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

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

Example 789

A DNA sequence (GBSx0837) was identified in *S.agalactiae* <SEQ ID 2417> which encodes the amino acid sequence <SEQ ID 2418>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2357(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10169> which encodes amino acid sequence <SEQ ID 10170> was also identified.

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

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 790

A DNA sequence (GBSx0838) was identified in *S.agalactiae* <SEQ ID 2419> which encodes the amino acid sequence <SEQ ID 2420>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.47   Transmembrane   21 -  37 (  19 -  38)

----- Final Results -----
               bacterial membrane --- Certainty=0.3187(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 791

A DNA sequence (GBSx0839) was identified in *S.agalactiae* <SEQ ID 2421> which encodes the amino acid sequence <SEQ ID 2422>. This protein is predicted to be DNA polymerase III delta prime subunit (dnaB). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0544(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 GENPEPT database.

```
AAF98347 AF280763 DNA polymerase III delta prime subunit [Streptococcus pyogenes]
 Identities = 284/444 (63%), Positives = 357/444 (79%), Gaps = 4/444 (0%)

Query: 3   ELKVLPHDIQAEQSVLGSIFIKPEKMIEVAEYLKPNDFYRPAHKILFKAMVSLADRGEAI 62
           EL+V P D+ AEQSVLGSIFI P+K+I V E++ P+DFY+ AHKI+F+AM++L+DR +AI
Sbjct: 8   ELRVQPQDLLAEQSVLGSIFISPDKLIAVREFISPDDFYKYAHKIIFRAMITLSDRNDAI 67

Query: 63  DIVTIKSTLESTDELGMVGGISYIAEIVNAVPTSSHAEHYAKIVAKKAQLRSIIDNLSDS 122
           D  TI++ L+  D+L  +GG+SYI E+VN+VPTS++AE+YAKIVA+KA LR II  L++S
Sbjct: 68  DATTIRTILDDQDDLQSIGGLSYIVELVNSVPTSANAEYYAKIVAEKAMLRDIIARLTES 127

Query: 123 IGNAYDEDMDIDEIIAKAERSLIEVSQASNKSSFRPIHDVLLENHSKIEERSNNTSQITG 182
           +  AYDE +  +E+IA  ER+LIE+++ SN+S FR I DVL  N+  +E RS  TS +TG
Sbjct: 128 VNLAYDEILKPEEVIAGVERALIELNEHSNRSGFRKISDVLKVNYEALEARSKQTSNVTG 187

Query: 183 IETGFYDFDKLITGLHEDQLIVLAARPAMGKTALALNIAQNVATKSNKAVAVFSLEMGAE 242
           + TGF D DK+ TGLH DQL++LAARPA+GKTA  LNIAQNV TK  K VA+FSLEMGAE
Sbjct: 188 LPTGFRDLDKITTGLHPDQLVILAARPAVGKTAFVLNIAQNVGTKQKKTVAIFSLEMGAE 247

Query: 243 SLVERMLSAEGTIINHHIRTGNLTVNEWQRLIYAQGQLAEAPIFIDDTAGVKITDIRARA 302
           SLV+RML+AEG + +H +RTG LT  +W  +  AQG LAEAPI+IDDT G+KIT+IRAR+
Sbjct: 248 SLVDRMLAAEGMVDSHSLRTGQLTDQDWNNVTIAQGALAEAPIYIDDTPGIKITEIRARS 307
```

```
Query: 303 RRLSQETD-GLGLIVIDYLQLIQGSRSDNRQQEVSEISRQLKIIAKELKVPVIALSQLSR 361
           R+LSQE D GLGLIVIDYLQLI G++ +NRQQEVS+ISRQLKI+AKELKVPVIALSQLSR
Sbjct: 308 RKLSQEVDGGLGLIVIDYLQLITGTKPENRQQEVSDISRQLKILAKELKVPVIALSQLSR 367

Query: 362 GVEQRNDKRPIMSDLRESGSIEQDADIVAFLYRDAYYQ---DKKEGQPENDITELIIRKN 418
           GVEQR DKRP++SD+RESGSIEQDADIVAFLYRD YY+   D  E   E++  E+I+ KN
Sbjct: 368 GVEQRQDKRPVLSDIRESGSIEQDADIVAFLYRDDYYRKECDDAEEAVEDNTIEVILEKN 427

Query: 419 RHGNLGTVKLYFHKEYTKFSSVEE 442
           R G  GTVKL F KEY KFSS+ +
Sbjct: 428 RAGARGTVKLMFQKEYNKFSSIAQ 451
```

There is also homology to SEQ ID 2424:

```
 Identities = 284/444 (63%), Positives = 357/444 (79%), Gaps = 4/444 (0%)

Query: 3   ELKVLPHDIQAEQSVLGSIFIKPEKMIEVAEYLKPNDFYRPAHKILFKAMVSLADRGEAI 62
           EL+V P D+ AEQSVLGSIFI P+K+I V E++ P+DFY+ AHKI+F+AM++L+DR +AI
Sbjct: 11  ELRVQPQDLLAEQSVLGSIFISPDKLIAVREFISPDDFYKYAHKIIFRAMITLSDRNDAI 70

Query: 63  DIVTIKSTLESTDELGMVGGISYIAEIVNAVPTSSHAEHYAKIVAKKAQLRSIIDNLSDS 122
           D  TI++ L+  D+L  +GG+SYI E+VN+VPTS++AE+YAKIVA+KA LR II  L++S
Sbjct: 71  DATTIRTILDDQDDLQSIGGLSYIVELVNSVPTSANAEYYAKIVAEKAMLRDIIARLTES 130

Query: 123 IGNAYDEDMDIDEIIAKAERSLIEVSQASNKSSFRPIHDVLLENHSKIEERSNNTSQITG 182
           +  AYDE +  +E+IA  ER+LIE+++ SN+S FR I DVL  N+  +E RS  TS +TG
Sbjct: 131 VNLAYDEILKPEEVIAGVERALIELNEHSNRSGFRKISDVLKVNYEALEARSKQTSNVTG 190

Query: 183 IETGFYDFDKLITGLHEDQLIVLAARPAMGKTALALNIAQNVATKSNKAVAVFSLEMGAE 242
           + TGF D DK+ TGLH DQL++LAARPA+GKTA  LNIAQNV TK  K VA+FSLEMGAE
Sbjct: 191 LPTGFRDLDKITTGLHPDQLVILAARPAVGKTAFVLNIAQNVGTKQKKTVAIFSLEMGAE 250

Query: 243 SLVERMLSAEGTIINHHIRTGNLTVNEWQRLIYAQGQLAEAPIFIDDTAGVKITDIRARA 302
           SLV+RML+AEG + +H +RTG LT  +W  +  AQG LAEAPI+IDDT G+KIT+IRAR+
Sbjct: 251 SLVDRMLAAEGMVDSHSLRTGQLTDQDWNNVTIAQGALAEAPIYIDDTPGIKITEIRARS 310

Query: 303 RRLSQETD-GLGLIVIDYLQLIQGSRSDNRQQEVSEISRQLKIIAKELKVPVIALSQLSR 361
           R+LSQE D GLGLIVIDYLQLI G++ +NRQQEVS+ISRQLKI+AKELKVPVIALSQLSR
Sbjct: 311 RKLSQEVDGGLGLIVIDYLQLITGTKPENRQQEVSDISRQLKILAKELKVPVIALSQLSR 370

Query: 362 GVEQRNDKRPIMSDLRESGSIEQDADIVAFLYRDAYYQ---DKKEGQPENDITELIIRKN 418
           GVEQR DKRP++SD+RESGSIEQDADIVAFLYRD YY+   D  E   E++  E+I+ KN
Sbjct: 371 GVEQRQDKRPVLSDIRESGSIEQDADIVAFLYRDDYYRKECDDAEEAVEDNTIEVILEKN 430

Query: 419 RHGNLGTVKLYFHKEYTKFSSVEE 442
           R G  GTVKL F KEY KFSS+ +
Sbjct: 431 RAGARGTVKLMFQKEYNKFSSIAQ 454
```

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

Example 792

A DNA sequence (GBSx0840) was identified in *S.agalactiae* <SEQ ID 2425> which encodes the amino acid sequence <SEQ ID 2426>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2146(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10167> which encodes amino acid sequence <SEQ ID 10168> was also identified.

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

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

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

Example 793

A DNA sequence (GBSx0841) was identified in *S.agalactiae* <SEQ ID 2427> which encodes the amino acid sequence <SEQ ID 2428>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2774(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 794

A DNA sequence (GBSx0842) was identified in *S.agalactiae* <SEQ ID 2429> which encodes the amino acid sequence <SEQ ID 2430>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.91   Transmembrane   63 -  79 (  62 -  79)

----- Final Results -----
             bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8661> which encodes amino acid sequence <SEQ ID 8662> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:    -11.31
GvH: Signal Score (-7.5): -1.86
     Possible site: 28
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -1.91 threshold:  0.0
   INTEGRAL    Likelihood = -1.91   Transmembrane   61 -  77 (  60 -  77)
   PERIPHERAL  Likelihood =  9.92       19
 modified ALOM score:   0.88

*** Reasoning Step: 3

----- Final Results -----
             bacterial membrane --- Certainty=0.1765(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 GENPEPT database.

```
>GP:AAB18686 GB:U38906 ORF11 [Bacteriophage r1t]
 Identities = 101/249 (40%), Positives = 157/249 (62%), Gaps = 21/249 (8%)

Query: 3   MAQRRMFSRKITETDRFLEMPLSSQALYFHLNMGADDEGFIDKAKTIQRTIGASDDDMKL 62
            MAQRRM  ++  +T +FL +PL +QALYFHL + ADD+G ++ A  + R +GA++D + L
Sbjct: 1   MAQRRMIDKRTIQTQKFLRLPLETQALYFHLMLNADDDGVVE-AFPVVRMVGAAEDSLGL 59

Query: 63  LIAKGFLIPFDSGVV-VIRHWRIHNYIQSDRFQSTLYQSEKAQLEYDKSKTASLKPIGNC 121
            L+ K F+ P +  +V  I  ++  N I+ DR++++ Y    AQL  ++      ++P  N
Sbjct: 60  LVVKQFIKPLNEEMVYFIIDFKEQNTIKKDRYKASKY----AQLLTNEEFGTEMEPKRNQ 115

Query: 122 IQNVSKMETQVRLSKGSLDKDSLTTYPTVSDNEEEDIPYKEIISYLNEKANRNYRPNIQK 181
            +    K     RL K  LDK++      +S   ++ IPY EI+ YLN+K  R++R N++
Sbjct: 116 LGTSDKN----RLDKNRLDKNN-----NMSGKPDDVIPYSEILEYLNKKTGRSFR-NVEA 165

Query: 182 NKTLIKARWSEGFRLDDFKHVIDTTVKDWSGTKY-----EKYLRPETLFGSKFEGYLNQA 236
            NK LIKARW+EG++L+DFK V+D  V +WSG  +     E YL+P+TLF +KF+ YLNQ
Sbjct: 166 NKKLIKARWNEGYKLEDFKTVVDNMVSNWSGKMFNGVPAENYLQPKTLFSNKFDSYLNQV 225

Query: 237 PRIKTETID 245
            PRI+ + I+
Sbjct: 226 PRIEQKEIN 234
```

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

SEQ ID 8662 (GBS344) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 12; MW 30.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 3; MW 59kDa).

The GBS344-GST fusion product was purified (Figure 213, (lane 3; Figure 226, lanes 4-6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 271), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 795

A DNA sequence (GBSx0843) was identified in *S.agalactiae* <SEQ ID 2431> which encodes the amino acid sequence <SEQ ID 2432>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2549(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 GENPEPT database:

```
>GP:AAG31329 GB:AF182207 ORF 272 [Bacteriophage mv4]
 Identities = 70/241 (29%), Positives = 125/241 (51%), Gaps = 30/241 (12%)

Query: 12  VLEETCEVHGCQLWLTKVPIKGRLEELKQCPECTKAAINIFENKLNSQSKINSKLADTYA 71
            VLE+ C  HG  L +T    +G  E++  CP+C   A+ + + + + +++  S +A
Sbjct: 16  VLEQKCSKHGLNL-ITYKNHEG--EQVTCCPQCQAEALEVLQERFDQKAR-QSIIARK-- 69
```

```
Query: 72  VFERDSLVSDKLRAKSLENYE---------IKDEIDQHAINYAKRMEQFYRQDRTGNAII 122
           F  +SL + K+   + + +E          IK ++   A+ +A +    +       A++
Sbjct: 70  -FRENSLANSKMWKCTFDTFEAQPGSAEELIKGQVRNAAVAFATKPVAHH-------AVL 121

Query: 123 TGPSGVGKSHLTYGLAKFMNEQFKAYESPKSVLFISLVSLFTKIKESFKVDNGY-RQADM 181
            G  G GKSHL    A  M ++   +   K++ FI++  LF+KIK SF   + Y  +
Sbjct: 122 YGQPGAGKSHL----AMAMMQEIHKHRPTKTMAFINISRLFSKIKNSFDDPSEYWTKEKA 177

Query: 182 IELLTRVDYLFLDDLGKESRKGDS--QNNEWTHQILYEILDNRSNTIINTNLSSKEIKALY 240
           +E++  VD L +DDLG ES  G +   +  +W   ++Y++L+N+   II TNLS +E+K +Y
Sbjct: 178 LEIMRGVDLLCIDDIGTESSMGRTGQEATKWAQDVIYDVLENQDRIIITTNLSERELKRVY 238
```

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

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

Example 796

A DNA sequence (GBSx0844) was identified in *S.agalactiae* <SEQ ID 2433> which encodes the amino acid sequence <SEQ ID 2434>. This protein is predicted to be methyl transferase. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1241(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10165> which encodes amino acid sequence <SEQ ID 10166> was also identified.

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

```
>GP:AAC98421 GB:L29323 methyl transferase [Streptococcus pneumoniae]
 Identities = 262/474 (55%), Positives = 313/474 (65%), Gaps = 71/474 (14%)

Query: 2   MKFLDLFAGIGGFRLGMEQAGHECIGFCEINKFARASYKVIHDTEGEIELHDITRVSD-E 60
           M+F+DLF+GIGGFRLGME  GHECIGFCEI+KFAR SYK I  TEGEIE HDI  VSD E
Sbjct: 1   MRFIDLFSGIGGFRLGMESVGHECIGFCEIDKFARESYKSIFQTEGEIEFHDIRDVSDDE 60

Query: 61  FIRGIGSVDVICGGFPCQAFSIAGNRRGFEDTRGTLFFEIARFASILRPKYLFLENVKGL 120
           F +  G VDVICGGFPCQAFSIAG R GFEDTRGTLFFEIAR A  ++P++LFLENVKGL
Sbjct: 61  FKKLRGKVDVICGGFPCQAFSIAGRRLGFEDTRGTLFFEIARAAKQIQPRFLFLENVKGL 120

Query: 121 LNHEGGATFETIIRTLDELGYNVEWQIFNSKNFGVPQNRERVFIIGHLRGEGTRPIFPFE 180
           LNH+ G TF TI+ TLDELG++VEWQ+ NSK+FGVPQNRERVFIIGH R  GTR  FPF
Sbjct: 121 LNHDKGRTFFTILTTLDELGFDVEWQMLNSKDFGVPQNRERVFIIGHSRKRGTRLGFPFR 180

Query: 181 SSITENYPIHTRKIGNVNPSGNGMNGEVYDSEGLSPTLTTNKGEGVKIAVN--------- 231
                P   + +GN+NPS +GM+G+VY SEGL+PTL   KGEG KIA+
Sbjct: 181 REGQATNPETLKILGNLNPSKSGMSGKVYYSEGLAPTLVRGKGEGFKIAIPCMTPDRLDK 240

Query: 232 -------------------------VVGRLPGKFEMPNRVYDPDGLAPTIRTMQGGGLE 265
                                  VVG LP  F+    RVY  +GL+PT+ TMQGG
Sbjct: 241 RQNGRRFKDNQEPMFTLNTQDRHGIVVVGDLPTSFKETGRVYGSEGLSPTLTTMQGGDKI 300

Query: 266 PKIIQRGRGYNQGGEYEISPTVTCNSWQENNLLKIKEATKKGYSEAEAGDSVNLSHPNSE 325
           PKI+                    +    LK++EATKKGY++AE GDS+NL  P+S+
Sbjct: 301 PKILIP--------------------EPIQFLKVREATKKGYAQAEIGDSINLERPSSQ 339

Query: 326 TRRGRVGKGIANTLLTGEEQGVVV--YDLYNRRKKDIVGTLTASGHNGNTTTGTFGISNG 383
```

```
            RRGRVGKGIANTL T  + GVVV  Y+  +++   + G L              G
Sbjct: 340 HRRGRVGKGIANTLTTSGQMGVVVASYEGEDKQVYQVAGVLID------------GQFYR 387

Query: 384 FRIRKLTPRECWRLQGFPDWAFDKASQVNSNSQLYKQAGNSVTVNVIAAIARRL 437
            RIR++TP+EC+RLQGFPDWAF+ A +V+SNSQLYKQAGNSVTV VIAAIA++L
Sbjct: 388 LRIRRITPKECFRLQGFPDWAFEAARKVSSNSQLYKQAGNSVTVPVIAAIAKKL 441
```

There is also homology to SEQ ID 2436:

```
Identities = 53/75 (70%), Positives = 62/75 (82%), Gaps = 1/75 (1%)

Query: 2  MKFLDLFAGIGGFRLGMEQAGHECIGFCEINKFARASYKVIHDTEGEIELHDITRVSDEF 61
          MKFLDLFAGIGGFRLG+    HECIGFCEI+KFAR SYK I++TEGEIE HDI +V+D+
Sbjct: 4  MKFLDLFAGIGGFRLGLINQCHECIGFCEIDKFARQSYKAIYETEGEIEFHDIRQVTDQD 63

Query: 62 IRGI-GSVDVICGGF 75
           R + G VD+ICGGF
Sbjct: 64 FRQLRGQVDIICGGF 78
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 797

A DNA sequence (GBSx0845) was identified in *S.agalactiae* <SEQ ID 2437> which encodes the amino acid sequence <SEQ ID 2438>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2585(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 798

A DNA sequence (GBSx0846) was identified in *S.agalactiae* <SEQ ID 2439> which encodes the amino acid sequence <SEQ ID 2440>. This protein is predicted to be arpR protein. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5070(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 GENPEPT database.

```
>GP:AAB09197 GB:U24159 orf12 [Bacteriophage HP1]
 Identities = 34/69 (49%), Positives = 47/69 (67%), Gaps = 1/69 (1%)

Query: 1  MTKTMTLEEKVEQWFIDRNLHE-ANPVKQFQKLIEETGELYSGIAKGKSEIIRDSLGDMQ 59
          M     L + +EQW  DRNL E + P KQF KL+EE GEL SG+AK K ++I+DS+GD
```

```
Sbjct: 1  MADLQQLIKNIEQWAEDRNLVEDSTPQKQFIKLMEEFGELCSGVAKNKPDVIKDSIGDCF 60

Query: 60 VVLIGIEQQ 68
          VV++ + +Q
Sbjct: 61 VVMVILAKQ 69
```

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

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

Example 799

A DNA sequence (GBSx0847) was identified in *S.agalactiae* <SEQ ID 2441> which encodes the amino acid sequence <SEQ ID 2442>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.10   Transmembrane   13 -  29 (  10 -  36)

----- Final Results -----
               bacterial membrane --- Certainty=0.3039(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 GENPEPT database.

```
>GP:AAD21919 GB:AF085222 unknown [Streptococcus thermophilus
          bacteriophage DT1]
 Identities = 31/67 (46%), Positives = 49/67 (72%), Gaps = 1/67 (1%)

Query: 42  HQEADRVIIYVADNAGAEMFGKITDKEIIEGRHTVTAGAYGKFLVTEEQYNEITVGDDIP 101
           ++  + ++++ ADN   E+ GK+T K ++   +T+  GAYGKFLV++EQY+ + VGD+IP
Sbjct: 34  NRPVEAIVVHKADNF-VELHGKVTGKSMVGKLYTIDCGAYGKFLVSKEQYDSVQVGDEIP 92

Query: 102 DYLKGRG 108
            YLKGRG
Sbjct: 93  SYLKGRG 99
```

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

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

Example 800

A DNA sequence (GBSx0848) was identified in *S.agalactiae* <SEQ ID 2443> which encodes the amino acid sequence <SEQ ID 2444>. This protein is predicted to be gene 17 protein. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5428(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 GENPEPT database.

```
>GP:CAA24397 GB:V01146 gene 1.7 [Bacteriophage T7]
 Identities = 30/72 (41%), Positives = 40/72 (54%)
```

```
Query: 47  DNVNYPSHYQGKYGLESIDVLRNFMTPEMLKGFYLGNALKYQLRYRKKNGLEDLKKARKN 106
           + V  PSHY      +E+I+V+    MT E   KG+  GN LKY+LR   KK+ L   L+K
Sbjct: 120 EGVTKPSHYMLFDDIEAIEVIARSMTVEQFKGYCFGNILKYRLRAGKKSELAYLEKDLAK 179

Query: 107 LDWLIEEMEKEK 118
            D+  E  EK K
Sbjct: 180 ADFYKELFEKHK 191
```

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

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

Example 801

A DNA sequence (GBSx0849) was identified in *S.agalactiae* <SEQ ID 2445> which encodes the amino acid sequence <SEQ ID 2446>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1375(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 802

A DNA sequence (GBSx0850) was identified in *S.agalactiae* <SEQ ID 2447> which encodes the amino acid sequence <SEQ ID 2448>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0087(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10163> which encodes amino acid sequence <SEQ ID 10164> was also identified.

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

```
>GP:AAF26608 GB:AF145054 ORF9 [Streptococcus thermophilus
          bacteriophage 7201]
 Identities = 99/148 (66%), Positives = 116/148 (77%), Gaps = 10/148 (6%)

Query: 5   MINNVVLIGRLTRDVELRYTPSNIANATFNLAVNRNFKNAAGDREADFINCVMWRQQAEN 64
            MINN VL+GRLT+D E +YT SNIA A+F+LAVNRNFK+A G+READFINCV+WRQQAEN
Sbjct: 1   MINNTVLVGRLTKDPEFKYTGSNIAVASFSLAVNRNFKDANGEREADFINCVIWRQQAEN 60

Query: 65  LANWTKKGMLIGITGRIQTRSYENQQGQRIYVTEVVADSFQILEKR----DNSTNQASMD 120
            LANW KKG LIGITGRIQTRSYENQQGQR+YVTEVVA++FQ+LE R        + N +
Sbjct: 61  LANWAKKGALIGITGRIQTRSYENQQGQRVYVTEVVAENFQMLESRAAREGGNANNSYSQ 120
```

```
Query: 121 DQLP------PSFGNSQPMDISDDDLPF 142
           Q+P         + N QP+DIS DDLPF
Sbjct: 121 QQVPNFARKNTEYSNKQPLDISSDDLPF 148
```

There is also homology to SEQ ID 1492.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 803

A DNA sequence (GBSx0851) was identified in *S.agalactiae* <SEQ ID 2449> which encodes the amino acid sequence <SEQ ID 2450>. This protein is predicted to be puff C4B protein. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1203(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10161> which encodes amino acid sequence <SEQ ID 10162> was also identified.

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

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 804

A DNA sequence (GBSx0852) was identified in *S.agalactiae* <SEQ ID 2451> which encodes the amino acid sequence <SEQ ID 2452>. This protein is predicted to be F5M15.19. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.34   Transmembrane    7 -  23 (   6 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.1935(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 805

A DNA sequence (GBSx0853) was identified in *S.agalactiae* <SEQ ID 2453> which encodes the amino acid sequence <SEQ ID 2454>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4398(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10159> which encodes amino acid sequence <SEQ ID 10160> was also identified.

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

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

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

Example 806

A DNA sequence (GBSx0855) was identified in *S.agalactiae* <SEQ ID 2455> which encodes the amino acid sequence <SEQ ID 2456>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2992(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 807

A DNA sequence (GBSx0856) was identified in *S.agalactiae* <SEQ ID 2457> which encodes the amino acid sequence <SEQ ID 2458>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4639(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 GENPEPT database.

```
>GP:BAB07758 GB:AP001520 unknown conserved protein [Bacillus halodurans]
 Identities = 65/184 (35%), Positives = 102/184 (55%), Gaps = 6/184 (3%)

Query: 1   MNIVEPLRDKDDIQAMKDYLSSWNEKYYMLFLLGINTGFRVGDILKLKVKDVQGWHIKVR 60
```

```
            M  V P RD D IQA+K  L    + + Y+LF +GINTG R+  +L LK+KDV
Sbjct: 1    MEYVVPFRDVDQIQAIKRSLKKKSPRDYLLFTIGINTGLRISQLLALKIKDVYDGQKPKD 60

Query: 61   EQKTGKYKSIKMTRPLKNELR---EFVKDKELHEYLFQSRVGKNKALSYKTVYWFLKRAA 117
              +    + + +   +K  L+    F++ +E H  LF S    ++ ++ +  Y  +K+AA
Sbjct: 61   YLQLESGEIVYLNDQVKKALQFYAHFIEFQEQH-CLFAS-TNPDQPMTRQHAYRIIKQAA 118

Query: 118  EDLGI-DNVGTHTMRKTFGYHYYKKYKNVADLMSLFNHSSPAVTLIYICVRQDELDTKMS 176
             +G+ D +GTHT+RKTFGYH Y++   ++ L   FNH +PA TL YI + ++E
Sbjct: 119  LQVGLTDQIGTHTLRKTFGYHAYRQGVALSLLQQRFNHQTPAQTLRYIDIAKNEQTIPRI 178

Query: 177  NFSL 180
            N +L
Sbjct: 179  NVNL 182
```

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

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

Example 808

A DNA sequence (GBSx0857) was identified in *S.agalactiae* <SEQ ID 2459> which encodes the amino acid sequence <SEQ ID 2460>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3582(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 809

A DNA sequence (GBSx0858) was identified in *S.agalactiae* <SEQ ID 2461> which encodes the amino acid sequence <SEQ ID 2462>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2732(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 810

A DNA sequence (GBSx0859) was identified in *S.agalactiae* <SEQ ID 2463> which encodes the amino acid sequence <SEQ ID 2464>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1720(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 811

A DNA sequence (GBSx0860) was identified in *S.agalactiae* <SEQ ID 2465> which encodes the amino acid sequence <SEQ ID 2466>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2619(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10157> which encodes amino acid sequence <SEQ ID 10158> was also identified.

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

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

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 812

A DNA sequence (GBSx0861) was identified in *S.agalactiae* <SEQ ID 2467> which encodes the amino acid sequence <SEQ ID 2468>. This protein is predicted to be terminase large subunit. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2753(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 GENPEPT database.

```
>GP:AAC27181 GB:AF009630 putative terminase subunit [bacteriophage
          bIL170]
```

```
 Identities = 147/531 (27%), Positives = 261/531 (48%), Gaps = 26/531 (4%)

Query: 19  IRICKLTMKSIRRVERYKEQYLFKQEEADKRIEFIEEECSNTKGLAGKLRLALPQKVWLE 78
           I + K   K+I++  R  ++Y+++ +   + IE+IE+    T G   K++L   QK W E
Sbjct: 16  IELNKYMRKTIQKQIRIHKKYIYRYDRVTQAIEWIEDNFYLTTGNLMKIKLHPTQKYWYE 75

Query: 79  TTWGFYHTVEVTKTNPDTLEEYTDYEERRLIHEVPIIVPRGTGKTTLGSAIAEVGQIIDG 138
              G+           D ++E     +  LI+E+ + + RG+GK++L +       I+ G
Sbjct: 76  LMLGY----------DMVDEKG--VQVNLINEIFLNLGRGSGKSSLMATRVLNWMILGG 122

Query: 139 EWGADIQLLAYSREQAGYLFNASRAMLSNEESLLHYMREADILRSTKQGILYETTNSLMS 198
           ++G +  ++AY   QA ++F+  R      ++L  Y  E  I +STKQG+ +    +
Sbjct: 123 QYGGESLVIAYDNTQARHVFDQVRNQTEASDTLRVY-NENKIFKSTKQGLEFTAFKTTFK 181

Query: 199 IKTSDYESLDGTNAHYNIFDEVHTYDDDFIKVVNDGSSRKRKNWITWYISTNGTKRDKLF 258
            +T+D    G N+  NIFDEVHTY +D  + VN GS +K+ NW + YI++ G KRD L+
Sbjct: 182 KQTNDTLRAQGGNSSLNIFDEVHTYGEDITESVNKGSRQKQDNWQSIYITSGGLKRDGLY 241

Query: 259 DKYYNIWVDILDDKIINDSVMPWIYQLDDVSEIHDPDMWQKAMPLLGITTEKETIARDIE 318
           DK    +   +++  ND     +Y L++  ++ D   W  A+PL+G   +  +  + E
Sbjct: 242 DKLVERFKS--EEEFYNDRSFGLLYMLENHEQVKDKKNWTMALPLIGDVPKWSGVIEEYE 299

Query: 319 MSKNDPAQQAELMAKTFNLPVNNYLAYFSNEECKGWSDKFDESLFVGDDERNARCVIGID 378
           +++ DPA Q + +A    LP+ +   YF+ ++ K    +F+ S+F         R  +GID
Sbjct: 300 LAQGDPALQNKFLAFNMGLPMQDTAYYFTPQDTK--LTEFNLSVF-----NKNRTYVGID 352

Query: 379 LSDVNDICSISFMVVRGEERHYLNKKFMPRHTIETLPKELRDKYTEWELSGMLHVHELDY 438
           LS + D+ ++SF+     + +     F  R   E L  E ++ +TE+   G L + + +Y
Sbjct: 353 LSLIGDLTAVSFVCELEGKTYSHTLTFSVRSQYEQLDTEQQELWTEFVDRGELILLDTEY 412

Query: 439 NDQAYIFEELRQFMSDNRILPVAVGYDRYNARELIRLFNDYYGDICHDIPQTVK---SLS 495
            +   + F S        +GYD      L  L   Y+ D   D  + ++   S++
Sbjct: 413 INVNDLIPYINDFRSKTGCRLRKIGYDPARYEILKGLIERYFFDKDGDNQRAIRQGFSMN 472

Query: 496 NPLKVYKEKAKMGKIIFDDPVATWNHANVRVKIDANNNIFPNKEKAKEKID 546
           + +K+ K K    K+I +  V  W   N  VKI  + +   K+  K+KID
Sbjct: 473 DYIKLLKSKLVENKLIHNQKVMQWALNNTAVKIGQSGDYMYTKKLEKDKID 523
```

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

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

Example 813

A DNA sequence (GBSx0862) was identified in *S.agalactiae* <SEQ ID 2469> which encodes the amino acid sequence <SEQ ID 2470>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3319(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 GENPEPT database.

```
>GP:AAB41469 GB:L35061 orfL4 [Bacteriophage phi-41]
 Identities = 86/374 (22%), Positives = 166/374 (43%), Gaps = 38/374 (10%)

Query: 12  FARIFRPNNRKSTRTYLQRSISYWRRNSIYLDNIYNKISTDTAQLRFKHVKITRNPGGVD 71
           F+R    N+ +    +   ++ Y    S ++ NI+NKI+ +  ++ F HVK  ++  G D
Sbjct: 10  FSRGKLNNDTQRVTAWQNEAVEY---TSAFVTNIHNKIANEITKVEFNHVKYKKSDVGSD 66

Query: 72  SMVWYEHSDLAEVLTVSPNPLEVPVVFWSNVTRAMLRDGVAVVVPRW--KNGRLVEIWLA 129
```

```
            +++    SDL EVL  S       + FW  V + +L      + P +  K G LV++  A
Sbjct: 67   TLISMAGSDLDEVLNWSSKGERNSMEFWQKVIKKLLTTRYIDLYPIFDRKTGDLVDLLFA 126

Query: 130 KKTVTWTAESVELMLDDVAVELPLTDVWVFENPKLNVTAQLNQITELIDINLNALTEKLS 189
                + E +  ++    +                        N+ T ++D  L  +  KL
Sbjct: 127 DNKKEYKPEELVRLISPFYI--------------------NEDTSILDNALAGIQTKLE 165

Query: 190 DGNSSLRGFLKLPT---KAADEHLKQQARDRVDSMLDLAKNGGIAYLEQGEEFQELSKDY 246
             G   ++G LK+        D+  K +A   + +M +++   G+   +   E  EL KDY
Sbjct: 166 QGK--MKGLLKINAFIDTDNDQEFKDKAMLTIKNMQEMSNYNGLTPTDNKTEIVELKKDY 223

Query: 247 STASKEELEFLKSQLYNAHGINEKLFTCDYTEEQYRAYYSSVMKLYQRVYSEEINRKYFT 306
            S  +K+E++ +KS+L   + +NE +     ++EQ   +Y+S +         +E+  K  +
Sbjct: 224 SVLNKDEIDLIKSELLTGYFMNENILLGTASQEQQIYFYNSTIIPLLIQLEKELTYKLIS 283

Query: 307 KTAR--TQGN----KLLVFFDMADMISFKDLVEGGFKSKYAGLMNSNEFRETYLGLPGYE 360
              R   +GN     +++V    +     + K+L++   ++     N+      +G    E
Sbjct: 284 TNRRRVVKGNLYYERIIVDNQLFKFATLKELIDLYHENINGPIFTQNQLL-VKMGEQPIE 342

Query: 361 GGEVFETNLNAVRI 374
            GG+V+  NLNAV +
Sbjct: 343 GGDVYIANLNAVAV 356
```

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

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

Example 814

A DNA sequence (GBSx0863) was identified in *S.agalactiae* <SEQ ID 2471> which encodes the amino acid sequence <SEQ ID 2472>. This protein is predicted to be a prohead protease. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3496(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 GENPEPT database.

```
>GP:AAF31089 GB:AF069529 protease [Bacteriophage HK97]
 Identities = 52/142 (36%), Positives = 73/142 (50%), Gaps = 11/142 (7%)

Query: 21  FEAYASTYDNTDREGDVMAKGCFDNTLKSKA-VVPMCLNHDR-NCVIGKHE-LSVDEKGL 77
           FE YAS ++NTD +GD++  G F N L ++   V M  NH      +GK + L+ DEKGL
Sbjct: 26  FEGYASVFNNTDSDGDIILPGAFKNALANQTRKVAMFFNHKTWELPVGKWDSLAEDEKGL 85

Query: 78  RTRSTFNLSDPEAKKTYDLMKMGALDSLSIGFFI--KDYEPIDAKQPYGGWIFKEVE-IF 134
             R          A      M+ G ++ +S+GF +   DY  I       G IFK ++ +
Sbjct: 86  YVRGQLTPGHSGAADLKAAMQHGTVEGMSVGFSVAKDDYTIIPT-----GRIFKNIQALR 140

Query: 135 EISVVTVPANPQATVDNIKEFD 156
           EISV T PAN QA +  +K  D
Sbjct: 141 EISVCTFPANEQAGIAAMKSVD 162
```

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

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