A related GBS nucleic acid sequence <SEQ ID 9843> which encodes amino acid sequence <SEQ ID 9844> was also identified.

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

```
>GP:AAC65759 GB:AE001250 conserved hypothetical protein [Treponema
          pallidum]
 Identities = 103/317 (32%), Positives = 163/317 (50%), Gaps = 15/317 (4%)

Query: 7   QLSHSLRLMGTTIDIQINSKNAQKQIR----EVIELLELYKNRFSANDFNSELMAINNNA 62
           + S +  ++GT   +++ SK    ++      EV  LL+  +   SAN  +S L A+N  A
Sbjct: 31  EYSRAELVIGTLCRVRVYSKRPAAEVHAALEEVFTLLQQQEMVLSANRDDSALAALNAQA 90

Query: 63  GIKPIQVHPDLFELITIGKEHSLARPSNLNIAIGPLVQTWRIGFSDAKLPSPSEISEAMI 122
           G  P+ V   L+ L+      +       N A+G  V+ W IGF  A +P P  + EA+
Sbjct: 91  GSAPVVVDRSLYALLERALFFAEKSGGAFNPALGAKVKLWNIGFDRAAVPDPDALKEALT 150

Query: 123 LSDPTHILLDSN-----KQSVFLNQIGMKIDLGALAKGYIADKIMTYLKNEMIDSAIINL 177
             D   + L +        +V L Q GM++DLGA+AKG++ADKI+  L    +DSA+++L
Sbjct: 151 RCDFRQVHLRAGVSVGAPHTVQLAQAGMQLDLGAIAKGFLADKIVQLLTAHALDSALVDL 210

Query: 178 GGNV----LVHGDNPNRSEGY--WVIGIQHPKKKRGKNIGTVKIKNQSVVTSGTYERRLI 231
           GGN+    L +GD  + +      W +GI+ P      K     V +++ SVVTSG YER
Sbjct: 211 GGNIFALGLKYGDVRSAAAQRLEWNVGIRDPHGTGQKPALVVSVRDCSVVTSGAYERFFE 270

Query: 232 IDDKEYHHIFDRQTGYPIQTEMASISIVSKQSVDCEIWTTRLFGLSIKEALDILNAVSYI 291
            D   YHHI D  TG+P  T++ S+SI + +S D +   T  F L  +++  +L     +
Sbjct: 271 RDGVRYHHIIDPVTGFPAHTDVDSVSIFAPRSTDADALATACFVLGYEKSCALLREFPGV 330

Query: 292 EGIIITKDDRIYLSDGL 308
           + + I  D R+  S G+
Sbjct: 331 DALFIFPDKRVRASAGI 347
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6585> which encodes the amino acid sequence <SEQ ID 6586>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1020(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 = 182/310 (58%), Positives = 232/310 (74%)

Query: 8   LSHSLRLMGTTIDIQINSKNAQKQIREVIELLELYKNRFSANDFNSELMAINNNAGIKPI 67
           ++  L+LMGT IDIQI S  A +Q+  VI+LL  YKNRFSAND NSELMAIN  AG+KP+
Sbjct: 3   VTQQLKLMGTVIDIQIESDKACQQLSRVIDLLYTYKNRFSANDSNSELMAINQAAGVKPV 62

Query: 68  QVHPDLFELITIGKEHSLARPSNLNIAIGPLVQTWRIGFSDAKLPSPSEISEAMILSDPT 127
            VH DLF LI IGK HSL+ PSNLNIAIGPLVQ WRIGF DA++PS + IS+ + L+DP
Sbjct: 63  SVHSDLFNLIQIGKAHSLSTPSNLNIAIGPLVQAWRIGFEDARVPSHNLISQQLALTDPR 122

Query: 128 HILLDSNKQSVFLNQIGMKIDLGALAKGYIADKIMTYLKNEMIDSAIINLGGNVLVHGDN 187
            +L+D  KQ+VFL Q+GM +DLGALAKGYI DKIM YL  + IDSA+INLGGNV VHG N
Sbjct: 123 QVLIDDKKQTVFLQQVGMALDLGALAKGYITDKIMAYLIEDGIDSALINLGGNVRVHGPN 182

Query: 188 PNRSEGYWVIGIQHPKKKRGKNIGTVKIKNQSVVTSGTYERRLIIDDKEYHHIFDRQTGY 247
           P   +  + IGIQ P  KRG+++G +K+ N SVVTSG YER+      K+YHHI DRQTGY
Sbjct: 183 PKSPDKTFRIGIQKPDAKRGQHLGVIKVNNHSVVTSGIYERQFTSKGKQYHHILDRQTGY 242

Query: 248 PIQTEMASISIVSKQSVDCEIWTTRLFGLSIKEALDILNAVSYIEGIIITKDDRIYLSDG 307
```

```
           PI+T+M S++I++  S  C+IWTTRLFGL     + +LN    IEG+++T+   + +S+G
Sbjct: 243 PIETDMLSLTIMAPSSFYCDIWTTRLFGLDSSMIITLLNTFDNIEGLLVTRKHHVLMSNG 302

Query: 308 LKHHFQLFYH 317
           L+H+FQ +YH
Sbjct: 303 LRHYFQPYYH 312
```

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

Example 2132

A DNA sequence (GBSx2248) was identified in *S.agalactiae* <SEQ ID 6587> which encodes the amino acid sequence <SEQ ID 6588>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0966(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:AAG18632 GB:AY007504 unknown [Streptococcus mitis]
 Identities = 92/160 (57%), Positives = 119/160 (73%), Gaps = 1/160 (0%)

Query: 1   MKLIGIVGTNSNKSTNRQLLQYMQQHFADKAEIELIEVKDLPLFNKPADKNVPQVILDIA 60
           MKL+ IVGTNSN+STNR+LL++MQ+HF+DKA+IE++E+K LP FN+P D+  P  +   +
Sbjct: 1   MKLVAIVGTNSNRSTNRKLLKFMQKHFSDKADIEVLEIKQLPAFNEPEDEQAPAEVQAFS 60

Query: 61  AKIEETDGVIIGTPEYDHSIPSALMSVLAWLSYGIYPLLNKPVMITGASYGTLGSSRAQL 120
            KI   DGVII TPEYDH+IP+ L S L W++Y    L+NKP MI GAS G LG+SRAQ
Sbjct: 61  EKILAADGVIISTPEYDHTIPAPLASALEWIAYTSRALINKPTMIVGASLGLLGTSRAQA 120

Query: 121 QLRQILNAPELKASVLP-DEFLLSHSLQAFDKDGNLHDIE 159
           LRQIL+APELKA V+P  EF L HS Q  D + +L++ E
Sbjct: 121 HLRQILDAPELKARVMPGTEFFLGHSEQVLDDECHLNNPE 160
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6589> which encodes the amino acid sequence <SEQ ID 6590>. 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 homology with the following sequences in the databases:

```
>GP:CAB62679 GB:AL133422 putative secreted protein. [Streptomyces
          coelicolor A3(2)]
 Identities = 68/192 (35%), Positives = 94/192 (48%), Gaps = 25/192 (13%)

Query: 4   ILFIVGSLREGSFNHQLAAQAQK-ALEHQAVVSYLNWKDVPVLNQDIEANAPLPVVDA-- 60
           IL +VGSLR GS N QLA  A + A E   V +    ++P  N+DI+    +P   A
Sbjct: 5   ILALVGSLRAGSHNRQLAEAAVRFAPEGAEVQLFEGLAEIPFYNEDIDVEGSVPAAAAKL 64

Query: 61  RQAVQSADAIWIFTPVYNFSIPGSVKNLLDWLSRALDLSDPTGPSAIGGKVVTVSSVANG 120
           R+A Q A A  +F+P YN +IP  +KN +DWLSR      P G  A  GK V V   A G
Sbjct: 65  REAAQGAQAFLLFSPEYNGTIPAVLKNAIDWLSR------PYGAGAFTGKPVAVVGTAFG 118
```

```
Query: 121 GHDQVFDQFKA----------LLPFIRTSVAGEFTK-ATVNP--DAWGTGRLEISKETKA 167
           +  V+ Q +A          ++  I+ S+ G  T+ A  +P  DA    +L    E  A
Sbjct: 119 QYGGVWAQDEARKAVGIAGGKVIEDIKLSIPGSVTRFAETHPADDAEVAAQL---TEVVA 175

Query: 168 NLLSQAEALLAA 179
             L   A+  +AA
Sbjct: 176 RLHGHADEAIAA 187
```

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

```
 Identities = 28/90 (31%), Positives = 49/90 (54%)

Query: 3  LIGIVGTNSNKSTNRQLLQYMQQHFADKAEIELIEVKDLPLFNKPADKNVPQVILDIAAK 62
          ++ IVG+    S N QL    Q+    +A +  +  KD+P+ N+  + N P  ++D
Sbjct: 4  ILFIVGSLREGSFNHQLAAQAQKALEHQAVVSYLNWKDVPVLNQDIEANAPLPVVDARQA 63

Query: 63 IEETDGVIIGTPEYDHSIPSALMSVLAWLS 92
          ++  D + I TP Y+ SIP ++ ++L WLS
Sbjct: 64 VQSADAIWIFTPVYNFSIPGSVKNLLDWLS 93
```

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

Example 2133

A DNA sequence (GBSx2249) was identified in *S.agalactiae* <SEQ ID 6591> which encodes the amino acid sequence <SEQ ID 6592>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1160(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 2134

A DNA sequence (GBSx2250) was identified in *S.agalactiae* <SEQ ID 6593> which encodes the amino acid sequence <SEQ ID 6594>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2132(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:AAG18632 GB:AY007504 unknown [Streptococcus mitis]
 Identities = 80/162 (49%), Positives = 112/162 (68%)

Query: 1   MKFVGIVGSNAEQSYNRMLLEFIRKNFKTKFELEVLEIDDIPMFNQDQNWEESFQLRLLN 60
```

```
           MK V IVG+N+ +S NR LL+F++K+F  K ++EVLEI  +P FN+ ++ +   +++  +
Sbjct: 1   MKLVAIVGTNSNRSTNRKLLKFMQKHFSDKADIEVLEIKQLPAFNEPEDEQAPAEVQAFS 60

Query: 61  NKITRADGVIIATPEHNHTITAALKSVLEWLSFAVHPLENKPVMIVGASYYDQGTSRAQI 120
            KI  ADGVII+TPE++HTI A L S LEW+++    L NKP MIVGAS    GTSRAQ
Sbjct: 61  EKILAADGVIISTPEYDHTIPAPLASALEWIAYTSRALINKPTMIVGASLGLLGTSRAQA 120

Query: 121 HLRKILDAPGVNAYTLPGNEFLLGKAKEAFDDNGNIINPGTV 162
           HLR+ILDAP + A  +PG EF LG +++  DD  ++ NP  V
Sbjct: 121 HLRQILDAPELKARVMPGTEFFLGHSEQVLDDECHLNNPEKV 162
```

There is also homology to SEQ ID 6596.

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

Example 2135

A DNA sequence (GBSx2251) was identified in *S.agalactiae* <SEQ ID 6597> which encodes the amino acid sequence <SEQ ID 6598>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.32   Transmembrane   13 -  29 (  11 -  29)

----- 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 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 2136

A DNA sequence (GBSx2252) was identified in *S.agalactiae* <SEQ ID 6599> which encodes the amino acid sequence <SEQ ID 6600>. This protein is predicted to be potential nitrite transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.92   Transmembrane   61 -  77 (  54 -  82)
   INTEGRAL    Likelihood = -5.57   Transmembrane  106 - 122 ( 103 - 126)
   INTEGRAL    Likelihood = -5.15   Transmembrane  160 - 176 ( 159 - 177)
   INTEGRAL    Likelihood = -4.09   Transmembrane  180 - 196 ( 179 - 199)
   INTEGRAL    Likelihood = -1.01   Transmembrane  233 - 249 ( 233 - 249)

----- Final Results -----
               bacterial membrane --- Certainty=0.4970(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:CAB15832 GB:Z99123 alternate gene name: ipa-48r~similar to
          nitrite transporter [Bacillus subtilis]
 Identities = 82/253 (32%), Positives = 119/253 (46%), Gaps = 10/253 (3%)

Query: 6   EKIAYNCAKKEALYKESLGRYALRSMLAGAYLTMSTAAGIVAADTIGK-ISPALSGFVF- 63
```

```
           +K+     KK+ ++  S  RY LRS+LA  ++      GI AA   G     A S F F
Sbjct: 7   QKVEQYALKKQNIFASSKIRYVLRSILASIFIGF----GITAASKTGSYFFMADSPFAFP 62

Query: 64  --AFIFSFGLIYVLIFNGELATSNMLYLTAGAYNKNISWKKAMTILIYCTFFNLVGACIL 121
             A  F   ++ +    G+L T N  Y T  A  K ISW+  + + +     NL+GA +
Sbjct: 63  AAAVTFGAAILMIAYGGGDLFTGNTFYFTYTALRKKISWRDTLYLWMSSYAGNLIGAILF 122

Query: 122 AWLFNQSYSFQHLTNDSFLGHVVAKKLGKPSSGAFLEGIIANMFVNLAILAYMLLKEESA 181
           A L + +  F+  +  SFL H+   K+  P+S  F  G++ N  V LA    M LK E A
Sbjct: 123 AILISATGLFEEPSVHSFLIHLAEHKMEPPASELFFRGMLCNWLVCLAFFIPMSLKGEGA 182

Query: 182 KMTVILSAIFMFVFLSNEHLIANFASFMLAAFSHIEHIKGFTLLNIIRQWTLVFFGNWIG 241
           K+  ++  +F F     EH IAN  +F ++    IEH    TL+  +R    V  GN
Sbjct: 183 KLFTMMLFVFCFFISGFEHSIANMCTFAISLL--IEHPDTVTLMGAVRNLIPVTLGNLTA 240

Query: 242 GGVFIGLAYAWLN 254
    .      G V +G  Y  LN
Sbjct: 241 GIVMMGWMYYTLN 253
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6601> which encodes the amino acid sequence <SEQ ID 6602>. Analysis of this protein sequence reveals the following:

```
     Possible site: 32
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.77   Transmembrane  142 - 158 ( 139 - 171)
   INTEGRAL    Likelihood = -9.34   Transmembrane   95 - 111 (  89 - 119)
   INTEGRAL    Likelihood = -2.02   Transmembrane   61 -  77 (  61 -  79)
   INTEGRAL    Likelihood = -1.12   Transmembrane  261 - 277 ( 261 - 279)
   INTEGRAL    Likelihood = -0.53   Transmembrane  191 - 207 ( 191 - 207)

----- Final Results -----
               bacterial membrane --- Certainty=0.4906(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:AAB80864 GB:U93874 formate dehydrogenase [Bacillus subtilis]
 Identities = 133/258 (51%), Positives = 181/258 (69%)

Query: 36  KTPEQILEATIHIGEHKVTKTFLAKSILGFIGGAMISLGYLLYVRIAASGLETFGAFSSI 95
           + P++I EA I  G  K+     +  +LGF+GGA I+LGYLL +R+     + +G+ SS+
Sbjct: 4   RKPDEIAEAAIEAGMKKIKLPLPSLLVLGFLGGAFIALGYLLDIRVIGDLPKEWGSLSSL 63

Query: 96  VGACAFPIGLIIILMAGGELITGNMMAVSAALLAKKIKFSELAKNWLIITLFNVIGAVFV 155
           +GA  FP+GLI++++AG ELITGNMM+V+ AL ++KI   ELA NW I+T+ N+IGA+FV
Sbjct: 64  IGAAVFPVGLILVVLAGAELITGNMMSVAMALFSRKISVKELAINWGIVTIMNLIGALFV 123

Query: 156 AFVFGHFLGLTSAGIFKEEVIEVAHAKIAASPLQALVSGIGCNWFVGLALWLCYGANDAA 215
           A+ FGH +GLT  G + E+ I VA  K+  S  + L+S IGCNW V LA+WL +GA DAA
Sbjct: 124 AYFFGHLVGLTETGPYLEKTIAVAQGKLDMSFGKVLISAIGCNWLVCLAVWLSFGAQDAA 183

Query: 216 GKFLGTWFPVMTFVALGFQHSVANAFVIPAAIFEGGATWLDFVTNFIFVYSGNIIGGAIF 275
           GK LG WFP+M FVA+GFQH VAN FVIPAAIF G  TW  F+ N I  + GN+IGGA+F
Sbjct: 184 GKILGIWFPIMAFVAIGFQHVVANMFVIPAAIFAGSFTWGQFIGNIIPAFIGNVIGGAVF 243

Query: 276 VSFLYFKVYYHPQKSKTQ 293
           V  +YF  Y+   +S+ +
Sbjct: 244 VGLIYFIAYHKKDRSRKE 261
```

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

```
 Identities = 69/240 (28%), Positives = 101/240 (41%), Gaps = 18/240 (7%)

Query: 15  KEALYKESLGRYALRSMLAGAYLTMSTAAGIVAADTIGKISPALSGFVFAFIFSFGLIYV 74
           K L K LG      +  G  L +  AA     +T G    A S  V A  F  GLI +
Sbjct: 55  KTFLAKSILGFIGGAMISLGYLLYVRIAAS--GLETFG----AFSSIVGACAFPIGLIII 108
```

```
Query: 75  LIFNGELATSNMLYLTAGAYNKNISWKKAMTILIYCTFFNLVGACILAWLFNQSYSFQHL 134
           L+  GEL T NM+ ++A    K I + +     +  T FN++GA  +A++F     F  L
Sbjct: 109 LMAGGELITGNMMAVSAALLAKKIKFSELAKNWLIITLFNVIGAVFVAFVFGH---FLGL 165

Query: 135 TNDSFLGHVVAK----KLGKPSSGAFLEGIIANMFVNLAILAYMLLKEESAKMTVILSAI 190
           T+       V +    K+      A + GI  N FV LA+       + + K       +
Sbjct: 166 TSAGIFKEEVIEVAHAKIAASPLQALVSGIGCNWFVGLALWLCYGANDAAGKFLGTWFPV 225

Query: 191 FMFVFLSNEHLIANFASFMLAAFSHIEHIKGFTLLNIIRQWTLVFFGNWIGGGVFIGLAY 250
             FV L  +H +AN      A F       G T L+ +  + V+ GN IGG +F+   Y
Sbjct: 226 MTFVALGFQHSVANAFVIPAAIFE-----GGATWLDFVTNFIFVYSGNIIGGAIFVSFLY 280
```

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

Example 2137

A DNA sequence (GBSx2253) was identified in *S.agalactiae* <SEQ ID 6603> which encodes the amino acid sequence <SEQ ID 6604>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1342(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 2138

A DNA sequence (GBSx2254) was identified in *S.agalactiae* <SEQ ID 6605> which encodes the amino acid sequence <SEQ ID 6606>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22   Transmembrane   44 -  60 (  44 -  60)

----- 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 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 2139

A DNA sequence (GBSx2255) was identified in *S.agalactiae* <SEQ ID 6607> which encodes the amino acid sequence <SEQ ID 6608>. This protein is predicted to be xanthine permease (pbuX). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.91   Transmembrane  160 - 176 ( 156 - 188)
   INTEGRAL    Likelihood = -6.48   Transmembrane  184 - 200 ( 179 - 211)
   INTEGRAL    Likelihood = -6.21   Transmembrane  101 - 117 (  96 - 121)
   INTEGRAL    Likelihood = -4.04   Transmembrane  309 - 325 ( 306 - 332)
   INTEGRAL    Likelihood = -3.98   Transmembrane  334 - 350 ( 331 - 353)
   INTEGRAL    Likelihood = -3.88   Transmembrane  400 - 416 ( 396 - 420)
   INTEGRAL    Likelihood = -3.45   Transmembrane   19 -  35 (  18 -  38)
   INTEGRAL    Likelihood = -2.81   Transmembrane  127 - 143 ( 127 - 146)
   INTEGRAL    Likelihood = -2.71   Transmembrane  228 - 244 ( 227 - 249)
   INTEGRAL    Likelihood = -2.02   Transmembrane   47 -  63 (  47 -  63)
   INTEGRAL    Likelihood = -1.97   Transmembrane   75 -  91 (  73 -  92)
   INTEGRAL    Likelihood = -0.85   Transmembrane  368 - 384 ( 368 - 384)

----- Final Results -----
               bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB14123 GB:Z99115 xanthine permease [Bacillus subtilis]
 Identities = 213/412 (51%), Positives = 292/412 (70%), Gaps = 5/412 (1%)

Query: 14  LGLQHLLAMYAGSILVPIMIASALGYNAKQLTYLIATDIFMCGIATLLQLRLSKHFGVGL 73
           LG+QH+LAMYAG+I+VP+++  A+G   +QLTYL++ DIFMCG+ATLLQ+  ++ FG+GL
Sbjct: 11  LGIQHVLAMYAGAIVVPLIVGKAMGLTVEQLTYLVSIDIFMCGVATLLQVWSNRFFGIGL 70

Query: 74  PVVLGCAFQSVAPLSIIGAQQGSGYMFGALIASGIYVVLVAGIFSKVANFFPPIVTGSVI 133
           PVVLGC F +V+P+  IG++ G   ++G++IASGI V+L++  F K+ +FFPP+VTGSV+
Sbjct: 71  PVVLGCTFTAVSPMIAIGSEYGVSTVYGSIIASGILVILISFFFGKLVSFFPPVVTGSVV 130

Query: 134 TTIGLTLIPVAMGNMGD---NAKEPSLQSLTLSLVTIGVVLLINIFAKGFLKSISILIGL 190
           T IG+TL+PVAM NM      +A    L +L L+   + +++L+  F KGF+KS+SILIG+
Sbjct: 131 TIIGITLMPVAMNNMAGGEGSADFGDLSNLALAFTVLSIIVLLYRFTKGFIKSVSILIGI 190

Query: 191 ISGTILAAFMGLVDASVVADAPLVHIPKPFYFGAPRFEFTSILMMCIIATVSMVESTGVY 250
           + GT +A FMG V     V+DA +V + +PFYFGAP F     I+ M I+A VS+VESTGVY
Sbjct: 191 LIGTFIAYFMGKVQFDNVSDAAVVQMIQPFYFGAPSFHAAPIITMSIVAIVSLVESTGVY 250

Query: 251 LALSDITNDKLDSKRLRNGYRSEGLAVLLGGLFNTFPYTGFSQNVGLVQISGIRTRKPIY 310
            AL D+TN +L    L  GYR+EGLAVLLGG+FN FPYT FSQNVGLVQ++GI+    I
Sbjct: 251 FALGDLTNRRLTEIDLSKGYRAEGLAVLLGGIFNAFPYTAFSQNVGLVQLTGIKKNAVIV 310

Query: 311 FTALFLVILGLLPKFGAMAQMIPSPVLGGAMLVLFGMVALQGMKMLNQVDFEHNEHNFII 370
           T + L+  GL PK  A   +IPS VLGGAM+ +FGMV   G+KML+++DF   E N +I
Sbjct: 311 VTGVILMAFGLFPKIAAFTTIIPSAVLGGAMVAMFGMVIAYGIKMLSRIDFAKQE-NLLI 369

Query: 371 AAVSIAAGVGFNGT-NLFISLPNTLQMFLTNGIVISTLTAVVLNIILNGLPK 421
           A S+  G+G      ++F  LP+ L +  TNGIV  + TAVVLNI+ N   K
Sbjct: 370 VACSVGLGLGVTVVPDIFKQLPSALTLLTTNGIVAGSFTAVVLNIVYNVFSK 421
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6609> which encodes the amino acid sequence <SEQ ID 6610>. Analysis of this protein sequence reveals the following:

```
     Possible site: 29
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.32   Transmembrane  160 - 176 ( 158 - 181)
   INTEGRAL    Likelihood = -6.37   Transmembrane  103 - 119 (  98 - 124)
```

```
INTEGRAL    Likelihood = -5.84   Transmembrane  130 - 146 ( 126 - 152)
INTEGRAL    Likelihood = -5.68   Transmembrane  187 - 203 ( 182 - 207)
INTEGRAL    Likelihood = -3.98   Transmembrane  337 - 353 ( 334 - 356)
INTEGRAL    Likelihood = -3.82   Transmembrane  232 - 248 ( 225 - 252)
INTEGRAL    Likelihood = -3.35   Transmembrane  403 - 419 ( 399 - 421)
INTEGRAL    Likelihood = -2.50   Transmembrane   22 -  38 (  21 -  41)
INTEGRAL    Likelihood = -2.07   Transmembrane  312 - 328 ( 312 - 328)
INTEGRAL    Likelihood = -1.97   Transmembrane   78 -  94 (  76 -  95)

----- 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 databases:

```
>GP:CAB15234 GB:Z99120 similar to purine permease [Bacillus subtilis]
 Identities = 216/421 (51%), Positives = 302/421 (71%), Gaps = 5/421 (1%)

Query: 6   KQEHSHSQSAVLGLQHVLSMYAGSILVPIMIAGALGYSARELTYLISTDIFMCGVATFLQ 65
           K++H+  Q  +LGLQH+L+MYAG+ILVP+++  A+G +A +LTYLI+ D+FMCG AT LQ
Sbjct: 2   KEQHNALQLMMLGLQHMLAMYAGAILVPLIVGAAIGLNAGQLTYLIAIDLFMCGAATLLQ 61

Query: 66  LKLTKHTGVGLPVVLGCAFQSVAPLSIIGAQQGSGAMFGALIASGIYVILVAGIFSKIAR 125
           L   ++ G+GLPVVLGC F +V P+  IG+  G  A++GA+IA+G+ V+L AG F K+ R
Sbjct: 62  LWRNRYFGIGLPVVLGCTFTAVGPMISIGSTYGVPAIYGAIIAAGLIVVLAAGFFGKLVR 121

Query: 126 FFPPIVTGSVITVIGLSLVGVAMGNM--GDNVKE-PTAQSMMLSLLTIVIILLVQKFTKG 182
           FFPP+VTGSV+ +IG+SL+  AM N+  G+  KE  +  +++L       ILL+  F KG
Sbjct: 122 FFPPVVTGSVVMIIGISLIPTAMNNLAGGEGSKEFGSLDNVLLGFGVTAFILLLFYFFKG 181

Query: 183 FVKSISILLIGLVAGTLVSAMMGLVDTTPVVEASWIHVPTPFYFGMPTFEITSIVMMCIIA 242
           F++SI+IL+GL+AGT  +  MG VD + V+EASW+HVP+ FYFG PTFE+ ++V M ++A
Sbjct: 182 FIRSIAILLGLIAGTAAAYFMGKVDFSEVLEASWLHVPSLFYFGPPTFELPAVVTMLLVA 241

Query: 243 TVSMVESTGVYLALSDLTNDQLDEKRLRNGYRSEGIAVFLGGLFNTFPYTGFSQNVGLVQ 302
            VS+VESTGVY AL+D+TN +L EK L  GYR+EG+A+ LGGLFN FPYT FSQNVG+VQ
Sbjct: 242 IVSLVESTGVYFALADITNRRLSEKDLEKGYRAEGLAILLGGLFNAFPYTAFSQNVGIVQ 301

Query: 303 ISGIKTRRPIYYAAGILVVIGLLPKFRAMAQMIPSPVLGGAMLVLFGMVALQGMQMLNRV 362
           +S +K+   I      ILV IGL+PK  A+  +IP+PVLGGAM+V+FGMV   G++ML+ V
Sbjct: 302 LSKMKSVNVIAITGIILVAIGLVPKAAALTTVIPTPVLGGAMIVMFGMVISYGIKMLSSV 361

Query: 363 DFQKNEYNFIIAAVSISAGLGFNGT-NLFASLPETAQMFLTNGIVIATLTSVVLNLVLNGK 422
           D   ++ N +I A S+S GLG      LF+SL   A +   +GIVI +LT++ L+     K
Sbjct: 362 DLD-SQGNLLIIASSVSLGLGATTVPALFSSLSGAASVLAGSGIVIGSLTAIALHAFFQTK 421
```

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

```
 Identities = 328/416 (78%), Positives = 380/416 (90%)

Query: 7   SNSQAALLGLQHLLAMYAGSILVPIMIASALGYNAKQLTYLIATDIFMCGIATLLQLRLS 66
           S+SQ+A+LGLQH+L+MYAGSILVPIMIA ALGY+A++LTYLI+TDIFMCG+AT LQL+L+
Sbjct: 10  SHSQSAVLGLQHVLSMYAGSILVPIMIAGALGYSARELTYLISTDIFMCGVATFLQLKLT 69

Query: 67  KHFGVGLPVVLGCAFQSVAPLSIIGAQQGSGYMFGALIASGIYVVLVAGIFSKVANFFPP 126
           KH GVGLPVVLGCAFQSVAPLSIIGAQQGSG MFGALIASGIYV+LVAGIFSK+A FFPP
Sbjct: 70  KHTGVGLPVVLGCAFQSVAPLSIIGAQQGSGAMFGALIASGIYVILVAGIFSKIARFFPP 129

Query: 127 IVTGSVITTIGLTLIPVAMGNMGDNAKEPSLQSLTLSLVTIGVVLLINIFAKGFLKSISI 186
           IVTGSVIT IGL+L+ VAMGNMGDN KEP+ QS+ LSL+TI ++LL+  F KGF+KSISI
Sbjct: 130 IVTGSVITVIGLSLVGVAMGNMGDNVKEPTAQSMMLSLLTIVIILLVQKFTKGFVKSISI 189

Query: 187 LIGLISGTILAAFMGLVDASVVADAPLVHIPKPFYFGAPRFEFTSILMMCIIATVSMVES 246
           LIGL++GT+++A MGLVD + V +A  +H+P PFYFG P FE TSI+MMCIIATVSMVES
Sbjct: 190 LIGLVAGTLVSAMMGLVDTTPVVEASWIHVPTPFYFGMPTFEITSIVMMCIIATVSMVES 249

Query: 247 TGVYLALSDITNDKLDSKRLRNGYRSEGLAVLLGGLFNTFPYTGFSQNVGLVQISGIRTR 306
```

```
           TGVYLALSD+TND+LD KRLRNGYRSEG+AV LGGLFNTFPYTGFSQNVGLVQISGI+TR
Sbjct: 250 TGVYLALSDLTNDQLDEKRLRNGYRSEGIAVFLGGLFNTFPYTGFSQNVGLVQISGIKTR 309

Query: 307 KPIYFTALFLVILGLLPKFGAMAQMIPSPVLGGAMLVLFGMVALQGMKMLNQVDFEHNEH 366
           +PIY+ A  LV++GLLPKF AMAQMIPSPVLGGAMLVLFGMVALQGM+MLN+VDF+ NE+
Sbjct: 310 RPIYYAAGILVVIGLLPKFRAMAQMIPSPVLGGAMLVLFGMVALQGMQMLNRVDFQKNEY 369

Query: 367 NFIIAAVSIAAGVGFNGTNLFISLPNTLQMFLTNGIVISTLTAVVLNIILNGLPKK 422
           NFIIAAVSI+AG+GFNGTNLF SLP T QMFLTNGIVI+TLT+VVLN++LNG  K+
Sbjct: 370 NFIIAAVSISAGLGFNGTNLFASLPETAQMFLTNGIVIATLTSVVLNLVLNGKDKQ 425
```

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

Example 2140

A DNA sequence (GBSx2256) was identified in *S.agalactiae* <SEQ ID 6611> which encodes the amino acid sequence <SEQ ID 6612>. This protein is predicted to be xanthine phosphoribosyltransferase (xpt). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1921(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:CAA13587 GB:AJ233894 xanthine phosphoribosyltransferase
           [Streptococcus pneumoniae]
 Identities = 133/162 (82%), Positives = 144/162 (88%)

Query: 16  GENILKVDSFLTHQVDFELMQEIGKVFADKYKEAGITKVVTIEASGIAPAVYAAQALGVP 75
           G+NILKVDSFLTHQVDF LM+EIGKVFA+K+  AGITKVVTIEASGIAPA++ A+AL VP
Sbjct: 1   GDNILKVDSFLTHQVDFSLMREIGKVFAEKFASAGITKVVTIEASGIAPALFTAEALNVP 60

Query: 76  MIFAKKAKNITMTEGILTAEVYSFTKQVTSQVSIVSRFLSNDDTVLIIDDFLANGQAAKG 135
           MIFAKKAKNITM EGILTAEVYSFTKQVTS VSI  +FLS +D VLIIDDFLANGQAAKG
Sbjct: 61  MIFAKKAKNITMNEGILTAEVYSFTKQVTSTVSIAGKFLSPEDKVLIIDDFLANGQAAKG 120

Query: 136 LLEIIGQAGAKVAGIGIVIEKSFQDGRDLLEKTGVPVTSLAR 177
           L++II QAGA V  IGIVIEKSFQDGRDLLEK G PV SLAR
Sbjct: 121 LIQIIEQAGATVEAIGIVIEKSFQDGRDLLEKAGYPVLSLAR 162
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6613> which encodes the amino acid sequence <SEQ ID 6614>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2576(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 = 156/193 (80%), Positives = 172/193 (88%)

Query: 1   MKLLEERILKDGDVLGENILKVDSFLTHQVDFELMQEIGKVFADKYKEAGITKVVTIEAS 60
           M+LLEERIL DG++LGENILKVD+FLTHQVD+ LM+ IGKVFA KY EAGITKVVTIEAS
Sbjct: 1   MQLLEERILTDGNILGENILKVDNFLTHQVDYRLMKAIGKVFAQKYAEAGITKVVTIEAS 60
```

```
Query: 61   GIAPAVYAAQALGVPMIFAKKAKNITMTEGILTAEVYSFTKQVTSQVSIVSRFLSNDDTV 120
            GIAPAVYAA+A+ VPMIFAKK KNITMTEGILTAEVYSFTKQVTS VSI  +FLS +D V
Sbjct: 61   GIAPAVYAAEAMDVPMIFAKKHKNITMTEGILTAEVYSFTKQVTSTVSIAGKFLSKEDKV 120

Query: 121  LIIDDFLANGQAAKGLLEIIGQAGAKVAGIGIVIEKSFQDGRDLLEKTGVPVTSLARIKA 180
             LIIDDFLANGQAAKGL+EIIGQAGA+V G+GIVIEKSFQDGR L+E  G+ VTSLARIK
Sbjct: 121  LIIDDFLANGQAAKGLIEIIGQAGAQVVGVGIVIEKSFQDGRRLIEDMGIEVTSLARIKN 180

Query: 181  FENGRVVFAEADA 193
             FENG + F EADA
Sbjct: 181  FENGNLNFLEADA 193
```

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

Example 2141

A DNA sequence (GBSx2257) was identified in *S.agalactiae* <SEQ ID 6615> which encodes the amino acid sequence <SEQ ID 6616>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2546(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:CAB15203 GB:Z99120 similar to GMP reductase [Bacillus subtilis]
 Identities = 243/321 (75%), Positives = 286/321 (88%), Gaps = 2/321 (0%)

Query: 7    VFDYEDIQLIPNKCIISSRSQADTSVKLGNYTFKLPVIPANMQTIIDEEVAETLACEGYF 66
             VFDYEDIQLIP KCI++SRS+ DTSV+LG +TFKLPV+PANMQTIIDE++A +LA  GYF
Sbjct: 4    VFDYEDIQLIPAKCIVNSRSECDTSVRLGGHTFKLPVVPANMQTIIDEKLAISLAENGYF 63

Query: 67   YIMHRFNEEERKPFIKRMHDKGLIASISVGVKDYEYDFVTSLKED--APEFITIDIAHGH 124
             Y+MHRF  E R  FIK M+ +GL +SISVGVKD EY+FV  L E+   PE++TIDIAHGH
Sbjct: 64   YVMHRFEPETRIDFIKDMNARGLFSSISVGVKDEEYEFVRQLAEENLTPEYVTIDIAHGH 123

Query: 125  SNSVIEMIQHIKQELPETFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKVKTGF 184
             SN+VIEMIQH+K+ LP++FVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITK+KTGF
Sbjct: 124  SNAVIEMIQHLKKHLPDSFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKIKTGF 183

Query: 185  GTGGWQLAALRWCSKAARKPIIADGGIRTHGDIAKSIRFGASMVMIGSLFAGHLESPGKL 244
             GTGGWQLAALRWC+KAA KPIIADGGIRTHGDIAKSIRFGA+MVMIGSLFAGH ESPG+
Sbjct: 184  GTGGWQLAALRWCAKAASKPIIADGGIRTHGDIAKSIRFGATMVMIGSLFAGHEESPGQT 243

Query: 245  VEVEGQQFKEYYGSASEYQKGEHKNVEGKKILLPVKGRLEDTLTEMQQDLQSSISYAGGK 304
             +E +G+ +KEY+GSASE+ KGE KNVEGKK+ +  KG ++DTL EM+QDLQSSISYAGG
Sbjct: 244  IEKDGKLYKEYFGSASEFPKGEKKNVEGKKMHVAHKGSIKDTLIEMEQDLQSSISYAGGT 303

Query: 305  ELDSLRHVDYVIVKNSIWNGD 325
             +L+++R+VDYVIVKNSI+NGD
Sbjct: 304  KLNAIRNVDYVIVKNSIFNGD 324
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6617> which encodes the amino acid sequence <SEQ ID 6618>. Analysis of this protein sequence reveals the following:

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

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

```
bacterial cytoplasm --- Certainty=0.2405(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 = 297/327 (90%), Positives = 311/327 (94%)

Query: 1   MFNDIPVFDYEDIQLIPNKCIISSRSQADTSVKLGNYTFKLPVIPANMQTIIDEEVAETL 60
            MFNDIPVFDYEDIQLIPNKCII+SRSQADTSV LG Y FKLPVIPANMQTIIDE +AE L
Sbjct: 8   MFNDIPVFDYEDIQLIPNKCIITSRSQADTSVTLGKYQFKLPVIPANMQTIIDETIAEQL 67

Query: 61  ACEGYFYIMHRFNEEERKPFIKRMHDKGLIASISVGVKDYEYDFVTSLKEDAPEFITIDI 120
            A EGYFYIMHRF+E+ RKPFIKRMH++GLIASISVGVK  EY+FVTSLKEDAPEFITIDI
Sbjct: 68  AKEGYFYIMHRFDEDSRKPFIKRMHEQGLIASISVGVKACEYEFVTSLKEDAPEFITIDI 127

Query: 121 AHGHSNSVIEMIQHIKQELPETFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKV 180
            AHGH+NSVI+MI+HIK ELPETFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKV
Sbjct: 128 AHGHANSVIDMIKHIKTELPETFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKV 187

Query: 181 KTGFGTGGWQLAALRWCSKAARKPIIADGGIRTHGDIAKSIRFGASMVMIGSLFAGHLES 240
            KTGFGTGGWQLAALRWC+KAARKPIIADGGIRTHGDIAKSIRFGASMVMIGSLFAGH ES
Sbjct: 188 KTGFGTGGWQLAALRWCAKAARKPIIADGGIRTHGDIAKSIRFGASMVMIGSLFAGHFES 247

Query: 241 PGKLVEVEGQQFKEYYGSASEYQKGEHKNVEGKKILLPVKGRLEDTLTEMQQDLQSSISY 300
            PGK VEV+G+ FKEYYGSASEYQKGEHKNVEGKKILLP KG L DTLTEMQQDLQSSISY
Sbjct: 248 PGKTVEVDGETFKEYYGSASEYQKGEHKNVEGKKILLPTKGHLSDTLTEMQQDLQSSISY 307

Query: 301 AGGKELDSLRHVDYVIVKNSIWNGDSI 327
            AGGK+LDSLRHVDYVIVKNSIWNGDSI
Sbjct: 308 AGGKDLDSLRHVDYVIVKNSIWNGDSI 334
```

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

Example 2142

A DNA sequence (GBSx2258) was identified in *S.agalactiae* <SEQ ID 6619> which encodes the amino acid sequence <SEQ ID 6620>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-16.98   Transmembrane  421 - 437 ( 413 - 443)
   INTEGRAL    Likelihood = -8.81   Transmembrane  166 - 182 ( 159 - 186)
   INTEGRAL    Likelihood = -8.55   Transmembrane  220 - 236 ( 208 - 238)
   INTEGRAL    Likelihood = -6.69   Transmembrane  322 - 338 ( 319 - 353)
   INTEGRAL    Likelihood = -5.26   Transmembrane  199 - 215 ( 196 - 218)
   INTEGRAL    Likelihood = -4.35   Transmembrane  343 - 359 ( 342 - 361)
   INTEGRAL    Likelihood = -4.09   Transmembrane  291 - 307 ( 287 - 308)
   INTEGRAL    Likelihood = -3.66   Transmembrane    8 -  24 (   8 -  27)
   INTEGRAL    Likelihood = -3.66   Transmembrane  133 - 149 ( 133 - 151)
   INTEGRAL    Likelihood = -3.19   Transmembrane  254 - 270 ( 253 - 278)
   INTEGRAL    Likelihood = -2.50   Transmembrane   53 -  69 (  53 -  72)
   INTEGRAL    Likelihood = -1.81   Transmembrane   77 -  93 (  76 -  95)
   INTEGRAL    Likelihood = -1.33   Transmembrane  109 - 125 ( 109 - 125)

----- Final Results -----
              bacterial membrane --- Certainty=0.7793(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:CAB61253 GB:AJ250422 ORFC [Oenococcus oeni]
 Identities = 157/447 (35%), Positives = 252/447 (56%), Gaps = 13/447 (2%)
```

```
Query: 11  AIITTAILGFSGILIETSMNVTFPLLMKEFGVNPAVIQWVTTGNLLAVAVTVPLSAFMIK 70
           AI+  A L F G+LIETSMNVTFP LM++F ++   +QW+TT  LL VA T+ ++AF+ K
Sbjct: 15  AILGLAGLAFCGVLIETSMNVTFPTLMQQFSISLNKVQWLTTAYLLLVAATISIAAFIEK 74

Query: 71  NLSERQIFTLANVLFLSGVLIDSFAPNLAILLVGRVLQGVGTGLALPLLFHIILTQIPME 130
               ++IF  A +LF+ GV+  + APN  ILL+GR++Q + TGLA+PLL   I+ QIP +
Sbjct: 75  RFIFKKIFFWAGLLFIIGVICSALAPNFLILLIGRLIQALSTGLAIPLLITEIMQQIPQK 134

Query: 131 RRGLMMGVAAMVTLLAPAVGPTYGGVISGMLGWKMIFMLLAPILIISTFIGLASIPKRQV 190
           ++G  M +   + L  P++GPTYGGVI+  L W++IF  + PI +I+  IGL+ I ++
Sbjct: 135 KQGSYMELVEWLLLWQPSLGPTYGGVITQDLSWRLIFWFVLPIGLIAWLIGLSFIEQKSS 194

Query: 191 RINDKLNFPAFISLGIGLATLLLAIEKMSIF---------YLLVAIVSFVIFYYL--NKQ 239
                +  FISL + L ++ +A+    I+          +LL+A++  ++F  L  N +
Sbjct: 195 PSKIPFAWKQFISLILALLSITVAVNNAGIYGWTSIKFYGFLLIAVILLIVFIKLSTNSR 254

Query: 240 LEFLNLNVFKDKDFSILLYGVLAFQMIPLALSFLLPNLLQLVLHQTSTKAGLFMFPGAIA 299
             +++++FK  +F   L      Q I L+L+FLLPN  QL+L +    +G+ +  G++
Sbjct: 255 QALISISIFKKWEFVCPLLIYFLIQFIQLSLTFLLPNYAQLILKKGVMISGIMLLCGSLI 314

Query: 300 VVFLSPFAGYLLDKIGAFKPIMIGISLSLIGLIGTAIFIPAKSVVVLLAFDILTKIGMGI 359
             L P  G +LD      P++IG    +  I   IF    SV ++ A  ++  IG
Sbjct: 315 SAILQPLTGRMLDSFSVKIPLVIGAFFLITSTISFTIFQRYLSVFLIAALYVIYMIGFSF 374

Query: 360 GASNMVTTALTKLKPAQSADGNSILNTLQQFAGAFATAVASQIFTIGQVAIPKNGAIIGS 419
             +N +T AL KL     +DGN++ NTLQQ+AG+  T+VAS +   G     K     GS
Sbjct: 375 VFNNSLTYALQKLPLKLISDGNAVFNTLQQYAGSLGTSVASALLANGIGTDGKQSNYTGS 434

Query: 420 Q--FAVLFVIVVVILAIVGLTYLRKRK 444
           +  F + F+   +++ ++     +K K
Sbjct: 435 RHIFILNFISCAIVVILIFSIQRKKNK 461
```

There is also homology to SEQ ID 46.

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

Example 2143

A DNA sequence (GBSx2259) was identified in *S.agalactiae* <SEQ ID 6621> which encodes the amino acid sequence <SEQ ID 6622>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2151(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 6595> which encodes the amino acid sequence <SEQ ID 6596>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> 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 = 74/214 (34%), Positives = 112/214 (51%), Gaps = 5/214 (2%)

Query: 13  NESENNFFITLKTYFNYLFSIQIIT---DISTLNHADFDGSFAFHDIETSIPHLVIDSNY 69
           N+ E   F    L  +F++LF + I+T    +I +    + F G F+FH+ +   +P L    ++
Sbjct: 15  NQLEETFIRELSHHFSHLFEVTILTSKANIQSNQLSTFQGIFSFHEHDIDLPTLYFKTSQ 74

Query: 70  LAISQTNSKIEANDIKTFSELSKTMTEFHYMLNFDLFNHLPYRFRLHNKDGQTIYSNHKP 129
                ++   +        LS+ +T F+    +      +LP + RL + +G  I  NH
Sbjct: 75  HGQGFLVTESVFDQATAVLSLSQYLTGFYQKFDGHFLQYLPLQARLSDANGNIIVDNHAF 134

Query: 130 EDPFDIYPEEEYPIDKWVQNSLIEKKAKELHLLLPSASQDYILVQSYKRLENDSGQLVGY 189
              F   P  +  I+ W+   L          LLPS S D+I +Q Y+ L+N  GQLVG
Sbjct: 135 NGSF--LPTTDKEIEDWILAELRLSDNPCKTFLLPSGSLDHIYMQHYQALKNPQGQLVGV 192

Query: 190 IEHVHNIKPLLEGYLKESGQAIVGWSDVTSGASI 223
           ++ V +IKPLL  YL+E+GQAIVGWSDVTSG SI
Sbjct: 193 LDTVQDIKPLLNQYLEETGQAIVGWSDVTSGPSI 226
```

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

Example 2144

A DNA sequence (GBSx2260) was identified in *S.agalactiae* <SEQ ID 6623> which encodes the amino acid sequence <SEQ ID 6624>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.10   Transmembrane  431 - 447 ( 423 - 452)
   INTEGRAL    Likelihood = -8.92   Transmembrane  149 - 165 ( 147 - 174)
   INTEGRAL    Likelihood = -8.86   Transmembrane  404 - 420 ( 402 - 428)
   INTEGRAL    Likelihood = -7.91   Transmembrane  299 - 315 ( 293 - 318)
   INTEGRAL    Likelihood = -6.42   Transmembrane  380 - 396 ( 374 - 398)
   INTEGRAL    Likelihood = -5.31   Transmembrane  350 - 366 ( 347 - 367)
   INTEGRAL    Likelihood = -4.57   Transmembrane   56 -  72 (  54 -  74)
   INTEGRAL    Likelihood = -3.24   Transmembrane  172 - 188 ( 171 - 198)
   INTEGRAL    Likelihood = -1.33   Transmembrane  224 - 240 ( 224 - 240)
   INTEGRAL    Likelihood = -0.59   Transmembrane  101 - 117 ( 101 - 117)

----- Final Results -----
              bacterial membrane --- Certainty=0.5840(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:AAF84709 GB:AE004010 potassium uptake protein [Xylella
          fastidiosa]
 Identities = 201/570 (35%), Positives = 319/570 (55%), Gaps = 34/570 (5%)

Query: 1   MAEMQHVNHSSFDKASKAGFII--ALGIVYGDIGTSPLYTMQSLVENQGGISSVTESFIL 58
           M+   H    +   ++  G II  A+G+V+GDIGTSPLYT++       G++   ++ +L
Sbjct: 1   MSTSSHSGDCTAVPSNSNGTIILSAIGVVFGDIGTSPLYTLKEAFSPNYGLTPNHDT-VL 59

Query: 59  GSISLIIWTLTLITTIKYVLVALKADNHHEGGIFSLYTLVRKMTPW-------LIVPAVI 111
           G +SLI W + L+ TIKYV V ++ DN  EGGI +L  L ++  P+        + +  +
Sbjct: 60  GILSLIFWAMMLVVTIKYVAVIMRVDNDGEGGIMALTALTQRTMPFGSRSIYIVGILGIF 119

Query: 112 GGATLLSDGALTPAVTVTSAVEGLKVVPSLQHIFQNQSNVIFATLFILLLLFAIQRFGTG 171
           G +    DG +TPA++V SAVEGL+V        F     V+  TL +L+LLF  QRFGT
Sbjct: 120 GTSLFFGDGVITPAISVLSAVEGLEVAEPHMKAF-----VVPITLAVLILLFLCQRFGTE 174

Query: 172 VIGKLFGPIMFIWFAFLGISGLLNSFAHPEVFKAINPYYGLKLLFSPENHKGIFILGSIF 231
            +GK FGPI  +WF  +G+ G+ N     PEV  AINP +GL   F        +F+LG++
Sbjct: 175 RVGKTFGPITLLWFIAIGVVGVYNIAQAPEVLHAINPSWGLH-FFLEHGWHSMFVLGAVV 233
```

```
Query: 232 LATTGAEALYSDLGHVGRGNIHVSWPFVKVAII-LSYCGQGAWILANKNAGNELNPFFAS 290
           LA TG EALY+D+GH G    I  +W +V + ++ L+Y GQGA +L+N  A     NPF+ S
Sbjct: 234 LAVTGGEALYADMGHFGAKAIRHAWMYVVLPMLALNYLGQGALVLSNPTAIG--NPFYQS 291

Query: 291 IPSQFTMHVVILATLAAIIASQALISGSFTLVSEAMRLKIFPQFRSTYPGDN-IGQTYIP 349
           IP      ++ LAT AA+IASQALI+GS++L S+AM+L   P+    +   + IGQ Y+P
Sbjct: 292 IPDWGLYPMIALATAAAVIASQALITGSYSLSSQAMQLGYIPRMNVRHTSQSTIGQIYVP 351

Query: 350 VINWFLFAITTSIVLLFKTSAHMEAAYGLAITITMLMTTILLSFFL-IQKGVKRGLVLLM 408
            +NW L  +    V+ F  S  M +AYG+A+T TM++TT+L+  +      V R ++ +M
Sbjct: 352 TVNWTLLTLVILTVIGFGDSTSMASAYGVAVTGTMMITTVLMIIYARANPRVPRLMLWMM 411

Query: 409 MIFFGILEGIFFLASAVKFMHGGYVVVIIAVAIIFIMTIWYKGSKIVSRYVKL--LDLKD 466
            I F  ++G FF A+ +KFM G +  +++ V I   M  W +G K++    ++  ++L +
Sbjct: 412 AIVFIAVDGAFFYANIIKFMDGAWFPLLLGVVIFTFMRTWLRGRKLLHEEMRKDGINLDN 471

Query: 467 YIGQLDKLRHDHRYPIYHTNVVYLTNRMEEDMIDKSIMYSILDKRPKKAQVYWFVNIKVT 526
           ++  L  L    + P      V+LT   +  ++M+++   +    +   F+ +K
Sbjct: 472 FLPGL-MLAPPVKVP---GTAVFLT--ADSTVVPHALMHNLKHNKVLHERNV-FLTVKTL 524

Query: 527 DEPYTA---EYKVDMMGTDFIVKVELYLGF 553
             PY A     K++ +   F  +V +  GF
Sbjct: 525 KIPYAANSERLKIEPISNGF-YRVHIRFGF 553
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6625> which encodes the amino acid sequence <SEQ ID 6626>. Analysis of this protein sequence reveals the following:

```
     Possible site: 15
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.78   Transmembrane  428 - 444 ( 421 - 453)
   INTEGRAL    Likelihood = -8.70   Transmembrane  146 - 162 ( 144 - 171)
   INTEGRAL    Likelihood = -7.64   Transmembrane  404 - 420 ( 398 - 426)
   INTEGRAL    Likelihood = -4.88   Transmembrane  296 - 312 ( 294 - 315)
   INTEGRAL    Likelihood = -4.57   Transmembrane   53 -  69 (  51 -  71)
   INTEGRAL    Likelihood = -3.93   Transmembrane  347 - 363 ( 343 - 363)
   INTEGRAL    Likelihood = -2.50   Transmembrane  372 - 388 ( 371 - 388)
   INTEGRAL    Likelihood = -1.33   Transmembrane  169 - 185 ( 169 - 185)
   INTEGRAL    Likelihood = -1.33   Transmembrane  221 - 237 ( 221 - 237)

----- Final Results -----
              bacterial membrane --- Certainty=0.5713(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAF84709 GB:AE004010 potassium uptake protein [Xylella
           fastidiosa]
 Identities = 177/467 (37%), Positives = 270/467 (56%), Gaps = 20/467 (4%)

Query: 7   TAFDKASKAGFII-ALGIVYGDIGTSPLYTIQSLVENQGGVNQVSESFILGSISLIIWTL 65
           TA    S    I+ A+G+V+GDIGTSPLYT++       G+    ++ +LG +SLI W +
Sbjct: 11  TAVPSNSNGTIILSAIGVVFGDIGTSPLYTLKEAFSPNYGLTPNHDT-VLGILSLIFWAM 69

Query: 66  TLITTIKYVLIALKADNHHEGGIFSLFTLVRKMSPW-------LIIPAMIGGATLLSDGA 118
            L+ TIKYV + ++ DN  EGGI +L  L ++  P+       + I  + G +    DG
Sbjct: 70  MLVVTIKYVAVIMRVDNDGEGGIMALTALTQRTMPFGSRSIYIVGILGIFGTSLFFGDGV 129

Query: 119 LTPAVTVTSAIEGLKAVPGLSHIYQNQTNVIITTLVILIVLFGIQRFGTGFIGKIFGPVM 178
           +TPA++V SA+EGL+        +     V+  TL +LI+LF  QRFGT  +GK FGP+
Sbjct: 130 ITPAISVLSAVEGLEVAEPHMKAF-----VVPITLAVLILLFLCQRFGTERVGKTFGPIT 184

Query: 179 FIWFSFLGVSGFFNTLGHLEIFKAINPYYALHLLFSPENHRGIFILGSIFLATTGAEALY 238
            +WF  +GV G +N      E+  AINP + LH  F        +F+LG++ LA TG EALY
Sbjct: 185 LLWFIAIGVVGVYNIAQAPEVLHAINPSWGLH-FFLEHGWHSMFVLGAVVLAVTGGEALY 243

Query: 239 SDLGHVGRGNIYVSWPFVKM-CIVLSYCGQAAWILANKHSGIELNPFFASVPSQLRVYLV 297
```

```
           +D+GH G   I   +W +V +   + L+Y GQ A +L+N  +     NPF+ S+P     ++
Sbjct: 244 ADMGHFGAKAIRHAWMYVVLPMLALNYLGQGALVLSNPTA--IGNPFYQSIPDWGLYPMI 301

Query: 298 SLATLAAIIASQALISGSFTLVSEAMRLKIFPLFRVTYPG-ANLGQLYIPVINWILFAVT 356
           +LAT AA+IASQALI+GS++L S+AM+L   P   V +   + +GQ+Y+P +NW L  +
Sbjct: 302 ALATAAAVIASQALITGSYSLSSQAMQLGYIPRMNVRHTSQSTIGQIYVPTVNWTLLTLV 361

Query: 357 SCTVLAFRTSAHMEAAYGLAITITMLMTTILLKYYLIKKGTRPILAHLVMAF-FALVEFI 415
             TV+ F  S  M +AYG+A+T TM++TT+L+  Y       P L   +MA  F  V+
Sbjct: 362 ILTVIGFGDSTSMASAYGVAVTGTMMITTVLMIIYARANPRVPRLMLWMMAIVFIAVDGA 421

Query: 416 FFLASAIKFMHGGYAVVILALAIVFVMFIWHAGTRIVFKYVKSLNLN 462
           FF A+ IKFM G +   ++L + I   M  W  G +++ + ++   +N
Sbjct: 422 FFYANIIKFMDGAWFPLLLGVVIFTFMRTWLRGRKLLHEEMRKDGIN 468
```

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

```
Identities = 485/651 (74%), Positives = 575/651 (87%)

Query: 10  SSFDKASKAGFIIALGIVYGDIGTSPLYTMQSLVENQGGISSVTESFILGSISLIIWTLT 69
           ++FDKASKAGFIIALGIVYGDIGTSPLYT+QSLVENQGG++ V+ESFILGSISLIIWTLT
Sbjct: 7   TAFDKASKAGFIIALGIVYGDIGTSPLYTIQSLVENQGGVNQVSESFILGSISLIIWTLT 66

Query: 70  LITTIKYVLVALKADNHHEGGIFSLYTLVRKMTPWLIVPAVIGGATLLSDGALTPAVTVT 129
           LITTIKYVL+ALKADNHHEGGIFSL+TLVRKM+PWLI+PA+IGGATLLSDGALTPAVTVT
Sbjct: 67  LITTIKYVLIALKADNHHEGGIFSLFTLVRKMSPWLIIPAMIGGATLLSDGALTPAVTVT 126

Query: 130 SAVEGLKVVPSLQHIFQNQSNVIFATLFILLLLFAIQRFGTGVIGKLFGPIMFIWFAFLG 189
           SA+EGLK VP L HI+QNQ+NVI  TL IL++LF IQRFGTG IGK+FGP+MFIWF+FLG
Sbjct: 127 SAIEGLKAVPGLSHIYQNQTNVIITTLVILIVLFGIQRFGTGFIGKIFGPVMFIWFSFLG 186

Query: 190 ISGLLNSFAHPEVFKAINPYYGLKLLFSPENHKGIFILGSIFLATTGAEALYSDLGHVGR 249
           +SG  N+  H E+FKAINPYY L LLFSPENH+GIFILGSIFLATTGAEALYSDLGHVGR
Sbjct: 187 VSGFFNTLGHLEIFKAINPYYYALHLLFSPENHRGIFILGSIFLATTGAEALYSDLGHVGR 246

Query: 250 GNIHVSWPFVKVAIILSYCGQGAWILANKNAGNELNPFFASIPSQFTMHVVILATLAAII 309
           GNI+VSWPFVK+ I+LSYCGQ AWILANK++G ELNPFFAS+PSQ  +++V LATLAAII
Sbjct: 247 GNIYVSWPFVKMCIVLSYCGQAAWILANKHSGIELNPFFASVPSQLRVYLVSLATLAAII 306

Query: 310 ASQALISGSFTLVSEAMRLKIFPQFRSTYPGDNIGQTYIPVINWFLFAITTSIVLLFKTS 369
           ASQALISGSFTLVSEAMRLKIFP FR TYPG N+GQ YIPVINW LFA+T+  VL F+TS
Sbjct: 307 ASQALISGSFTLVSEAMRLKIFPLFRVTYPGANLGQLYIPVINWILFAVTSCTVLAFRTS 366

Query: 370 AHMEAAYGLAITITMLMTTILLSFFLIQKGVKRGLVLLMMIFFGILEGIFFLASAVKFMH 429
           AHMEAAYGLAITITMLMTTILL ++LI+KG +  L  L+M FF ++E IFFLASA+KFMH
Sbjct: 367 AHMEAAYGLAITITMLMTTILLKYYLIKKGTRPILAHLVMAFFALVEFIFFLASAIKFMH 426

Query: 430 GGYVVVIIAVAIIFIMTIWYKGSKIVSRYVKLLDLKDYIGQLDKLRHDHRYPIYHTNVVY 489
           GGY VVI+A+AI+F+M IW+ G++IV +YVK L+L DY  Q+ +LR D  + +Y TNVVY
Sbjct: 427 GGYAVVILALAIVFVMFIWHAGTRIVFKYVKSLNLNDYKEQIKQLRDDVCFDLYQTNVVY 486

Query: 490 LTNRMEEDMIDKSIMYSILDKRPKKAQVYWFVNIKVTDEPYTAEYKVDMMGTDFIVKVEL 549
           L+NRM++ MID+SI+YSILDKRPK+AQVYWFVN++VTDEPYTA+YKVDMMGTD++V+V L
Sbjct: 487 LSNRMQDHMIDRSILYSILDKRPKRAQVYWFVNVQVTDEPYTAKYKVDMMGTDYMVRVNL 546

Query: 550 YLGFKMRQTVSRYLRTIVEELLESGRLPKQGKTYSVRPDSNVGDFRFIVLDERFSSSQNL 609
           YLGF+M QTV RYLRTIV++L+ESGRLPKQ + Y++ P  +VGDFRF++++ER S+++ L
Sbjct: 547 YLGFRMPQTVPRYLRTIVQDLMESGRLPKQEQEYTITPGRDVGDFRFVLIEERVSNARQL 606

Query: 610 KPGERFVMLMKSSIKHWTATPIRWFGLQFSEVTTEVVPLIFTANRGLPIKE 660
             ERF+M  K+SIKH TA+P+RWFGLQ+SEVT EVVPLI +    LPIKE
Sbjct: 607 SNFERFIMQTKASIKHVTASPMRWFGLQYSEVTLEVVPLILSDVLKLPIKE 657
```

A related GBS gene <SEQ ID 8983> and protein <SEQ ID 8984> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
```

```
McG: Discrim Score:      5.84
GvH: Signal Score (-7.5): -4.59
     Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 10 value: -12.10 threshold:  0.0
   INTEGRAL    Likelihood =-12.10   Transmembrane  431 - 447 ( 423 - 452)
   INTEGRAL    Likelihood = -8.92   Transmembrane  149 - 165 ( 147 - 174)
   INTEGRAL    Likelihood = -8.86   Transmembrane  404 - 420 ( 402 - 428)
   INTEGRAL    Likelihood = -7.91   Transmembrane  299 - 315 ( 293 - 318)
   INTEGRAL    Likelihood = -6.42   Transmembrane  380 - 396 ( 374 - 398)
   INTEGRAL    Likelihood = -5.31   Transmembrane  350 - 366 ( 347 - 367)
   INTEGRAL    Likelihood = -4.57   Transmembrane   56 -  72 (  54 -  74)
   INTEGRAL    Likelihood = -3.24   Transmembrane  172 - 188 ( 171 - 198)
   INTEGRAL    Likelihood = -1.33   Transmembrane  224 - 240 ( 224 - 240)
   INTEGRAL    Likelihood = -0.59   Transmembrane  101 - 117 ( 101 - 117)
   PERIPHERAL  Likelihood =  0.85       20
 modified ALOM score:   2.92

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5840(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:

```
ORF02578(367 - 1680 of 2607)
GP|9106998|gb|AAF84709.1|AE004010_6|AE004010(25 - 463 of 634) potassium uptake protein
{Xylella fastidiosa}
%Match = 17.8
%Identity = 40.4  %Similarity = 63.7
Matches = 177  Mismatches = 150  Conservative Sub.s = 102

180       210       240       270       300       330       360       390
TSTCLS*LK**RPGNALIISGLFIDKCCFFNLICYNEFSHFFD*YYLIGGLAEMQHVNHSSFDKASKAGFIIALGIVYGD
                                                                |:|:|:||
                                                  MSTSSHSGDCTAVPSNSNGTIILSAIGVVFGD
                                                         10        20        30

420       450       480       510       540       570       600       612
IGTSPLYTMQSLVENQGGISSVTESFILGSISLIIWTLTLITTIKYVLVALKADNHHEGGIFSLYTLVRKMTP------W
||||||||::       |::   :: :|| :||| | : |: ||||| | :: ||  |||| :|  | ::  |      :
IGTSPLYTLKEAFSPNYGLTPNHDT-VLGILSLIFWAMMLVVTIKYVAVIMRVDNDGEGGIMALTALTQRTMPFGSRSIY
                50        60        70        80        90       100       110

639       669       699       729       759       789       819       849
LI-VPAVIGGATLLSDGALTPAVTVTSAVEGLKVVPSLQHIFQNQSNVIFATLFILLLLFAIQRFGTGVIGKLFGPIMFI
:: :  : | : :: || :|||::| ||||||:|       : :: |:  || :|:||| |||||  :|| |||| ::
IVGILGIFGTSLFFGDGVITPAISVLSAVEGLEV-----AEPHMKAFVVPITLAVLILLFLCQRFGTERVGKTFGPITLL
                130       140       150       160       170       180

879       909       939       969       999      1029      1059      1089
WFAFLGISGLLNSFAHPEVFKAINPYYGLKLLFSPENHKGIFILGSIFLATTGAEALYSDLGHVGRGNIHVSWPFVKVAI
||  :|: |: |    |||: |||| :|| ::::    | :|:||:: || || ||||:|:|| |   |  :| :| : :
WFIAIGVVGVYNIAQAPEVLHAINPSWGLHFFLEHGWHS-MFVLGAVVLAVTGGEALYADMGHFGAKAIRHAWMYVVLPM
           200       210       220       230       240       250       260

1116      1146      1176      1206      1236      1266      1296      1326
I-LSYCGQGAWILANKNAGNELNPFFASIPSQFTMHVVILATLAAIIASQALISGSFTLVSEAMRLKIFPQFRSTYPGDN
: |:| |||| :|:|  |    |||: |||      :: ||| ||:|||||||:||::| |:||:|   |:    :·  :
LALNYLGQGALVLSNPTA--IGNPFYQSIPDWGLYPMIALATAAAVIASQALITGSYSLSSQAMQLGYIPRMNVRHTSQS
            280       290       300       310       320       330       340

1353      1383      1413      1443      1473      1500      1530      1560
-IGQTYIPVINWFLFAITTSIVLLFKTSAHMEAAYGLAITITMLMTTILLSFFL-IQKGVKRGLVLLMMIFFGILEGIFF
 ||| |:| :|| |: :    |: |  |  | :|||:|:| ||::||:|: :      | | :: :| | |  ::| ||
TIGQIYVPTVNWTLLTLVILTVIGFGDSTSMASAYGVAVTGTMMITTVLMIIYARANPRVPRLMLWMMAIVFIAVDGAFF
```

```
               360       370       380       390       400       410       420

1590       1620       1650       1680       1710       1740       1770       1800
LASAVKFMHGGYVVVIIAVAIIFIMTIWYKGSKIVSRYVKLLDLKDYIGQLDKLRHDHRYPIYHTNVVYLTNRMEEDMID
 |: :||| | :  ::: | |   |  | :| |::   ::
YANIIKFMDGAWFPLLLGVVIFTFMRTWLRGRKLLHEEMRKDGINLDNFLPGLMLAPPVKVPGTAVFLTADSTVVPHALM
               440       450       460       470       480       490       500
```

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

Example 2145

A DNA sequence (GBSx2261) was identified in *S.agalactiae* <SEQ ID 6627> which encodes the amino acid sequence <SEQ ID 6628>. This protein is predicted to be serine dehydrogenase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3261(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:AAD07424 GB:AE000552 short chain alcohol dehydrogenase
           [Helicobacter pylori 26695]
 Identities = 18/31 (58%), Positives = 25/31 (80%)

Query: 3   WVASQPEHININRIEIMPVSQTYGPQPVYRD 33
           W+  QP H+NINRIEIMP+SQT+ P P +++
Sbjct: 219 WIYEQPLHVNINRIEIMPISQTFAPLPTHKN 249
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6629> which encodes the amino acid sequence <SEQ ID 6630>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1021(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 = 24/33 (72%), Positives = 29/33 (87%)

Query: 1  MSWVASQPEHININRIEIMPVSQTYGPQPVYRD 33
          +SWV  QP H+N+NRIE+MPVSQ+YGPQPV RD
Sbjct: 20 VSWVIHQPPHVNVNRIELMPVSQSYGPQPVTRD 52
```

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

Example 2146

A DNA sequence (GBSx2262) was identified in *S.agalactiae* <SEQ ID 6631> which encodes the amino acid sequence <SEQ ID 6632>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> May be a lipoprotein

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

A related GBS nucleic acid sequence <SEQ ID 9337> which encodes amino acid sequence <SEQ ID 9338> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10781> which encodes amino acid sequence <SEQ ID 10782> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10951> which encodes amino acid sequence <SEQ ID 10952> was also identified.

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

```
>GP:CAA32349 GB:X14130 ORF (AA 1 to 299) [Lactococcus lactis subsp.
          cremoris]
 Identities = 72/215 (33%), Positives = 110/215 (50%), Gaps = 8/215 (3%)

Query: 4   RSKLAAGFLTLMSVATLAACSGKTSNGTN--VVTMKGDTITVSDFYDQVKTSKAAQQSML 61
           + K+    L   +   L   SG  SN T+  V T  G  +T S FY ++K S   +  +
Sbjct: 2   KKKMRLKVLLASTATALLLLSGCQSNQTDQTVATYSGGKVTESSFYKELKQSPTTKTMLA 61

Query: 62  TLILSRVFDTQYGDKVSDKKVSEAYNKTAKGYGNSFSSALSQAGLTPEGYKQQIRTTMLV 121
            +++ R  +  YG  VS K V++AY+    + YG +F + LSQ G +    +K+ +RT  L
Sbjct: 62  NMLIYRALNHAYGKSVSTKTVNDAYDSYKQQYGENFDAFLSQNGFSRSSFKESLRTNFLS 121

Query: 122 EYAVKEAAKKELTEANYKEAYKNYTPETSVQVIKLDAEDKAKSVLKDVKADGADFAKIAK 181
           E A+K+   K+++E+  K A+K Y P+ +VQ I    ED AK V+ D+ A G DFA +AK
Sbjct: 122 EVALKKL--KKVSESQLKAAWKTYQPKVTVQHILTSDEDTAKQVISDLAA-GKDFAMLAK 178

Query: 182 E---KTTATDKKVEYKFDSAGTTLPKEVMSAAFKL 213
               T   D   +  F+    TL      AA+KL
Sbjct: 179 TDSIDTATKDNGGKISFELNNKTLDATFKDAAYKL 213
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6633> which encodes the amino acid sequence <SEQ ID 6634>. Analysis of this protein sequence reveals the following:

```
      Possible site: 21

>>> May be a lipoprotein

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

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

```
>GP:AAA25247 GB:M83946 maturation protein [Lactobacillus paracasei]
 Identities = 88/294 (29%), Positives = 146/294 (48%), Gaps = 14/294 (4%)

Query: 7   LIASVVTLASVMALAACQSTNDNTKVISMKGDTISVSDFYNETKNTEVSQKAMLNLVISR 66
           L+AS  T  +++ L+ CQS   + KV +  G  ++ S+FY E K +  ++  + N++I R
Sbjct: 10  LLASTAT--ALLLLSGCQSNQADQKVATYSGGKVTESNFYKELKQSPTTKTMLANMLIYR 67

Query: 67  VFEAQYGDKVSKKEVEKAYHKTAEQYGASFSAALAQSSLTPETFKRQIRSSKLVEYAVKE 126
               YG  VS K V  AY    +QYG +F A L+Q+  +  +FK  +R++ L E A+K+
Sbjct: 68  ALNHAYGKSVSTKTVNDAYDSYKQQYGENFDAFLSQNGFSRSSFKESLRTNFLSEVALKK 127

Query: 127 AAKKELTTQEYKKAYESYTPTMAVEMITLDNEETAKSVLEELKAEGADFTAIAKE---KT 183
              K+++  + K  +++Y P + V+ I   +E+TAK V+ +L A G DF  +AK     T
Sbjct: 128 L--KKVSESQLKAVWKTYQPKVTVQHILTSDEDTAKQVISDLAA-GKDFATLAKTDSIDT 184

Query: 184 TTPEKKVTYKFDSGATNVPTDVVKAASSLNEGGISDVISVLDPTSYQKKFYIVKVTKKAE 243
```

```
           T +      F+S    +      AA  L  G  +       P       + ++K+
Sbjct: 185 ATKDNGGKISFESNNKTLDATFKDAAYKLKNGDYTQT-----PVKVTNGYEVIKMINH-P 238

Query: 244 KKSDWQEYKKRLKAIIIAEKSKDMNFQNKVIANALDKANVKIKDKAFANILAQY 297
            K  +   KK L A + A+ S+D +   +VI+  L   +V IKDK  A+ L  Y
Sbjct: 239 AKGTFTSSKKALTASVYAKWSRDSSIMQRVISQVLKNQHVTIKDKDLADALDSY 292
```

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

```
 Identities = 125/213 (58%), Positives = 168/213 (78%), Gaps = 1/213 (0%)

Query: 1   MKTRSKLAAGFLTLMSVATLAACSGKTSNGTNVVTMKGDTITVSDFYDQVKTSKAAQQSM 60
            MK  +KL A  +TL SV  LAAC   T++ T V++MKGDTI+VSDFY++ K ++ +Q++M
Sbjct: 1   MKNSNKLIASVVTLASVMALAACQS-TNDNTKVISMKGDTISVSDFYNETKNTEVSQKAM 59

Query: 61  LTLILSRVFDTQYGDKVSDKKVSEAYNKTAKGYGNSFSSALSQAGLTPEGYKQQIRTTML 120
            L L++SRVF+ QYGDKVS K+V +AY+KTA+ YG SFS+AL+Q+ LTPE +K+QIR++ L
Sbjct: 60  LNLVISRVFEAQYGDKVSKKEVEKAYHKTAEQYGASFSAALAQSSLTPETFKRQIRSSKL 119

Query: 121 VEYAVKEAAKKELTEANYKEAYKNYTPETSVQVIKLDAEDKAKSVLKDVKADGADFAKIA 180
            VEYAVKEAAKKELT   YK+AY++YTP  +V++I LD E+ AKSVL+++KA+GADF  IA
Sbjct: 120 VEYAVKEAAKKELTTQEYKKAYESYTPTMAVEMITLDNEETAKSVLEELKAEGADFTAIA 179

Query: 181 KEKTTATDKKVEYKFDSAGTTLPKEVMSAAFKL 213
            KEKTT  +KKV YKFDS  T +P +V+ AA  L
Sbjct: 180 KEKTTTPEKKVTYKFDSGATNVPTDVVKAASSL 212
```

SEQ ID 10782 (GBS657) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 143 (lane 8-10; MW 62.8kDa) and in Figure 187 (lane 3; MW 63kDa). Purified GBS657-GST is shown in Figure 245, lanes 2 & 3.

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

Example 2147

A DNA sequence (GBSx2263) was identified in *S.agalactiae* <SEQ ID 6635> which encodes the amino acid sequence <SEQ ID 6636>. This protein is predicted to be methyltransferase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2576(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:CAA68045 GB:X99710 methyltransferase [Lactococcus lactis]
 Identities = 132/227 (58%), Positives = 169/227 (74%)

Query: 1   MVQSYSKNANHNMRRPVVKEEIVQYMRQHQKQNNGCLAELEAFAKQENIPIIPHETATYF 60
            MV++Y   +N  M RPVVK E+V++MR  Q Q  G LAE+  FAK+ NIP+IPHET  YF
Sbjct: 1   MVETYKSTSNPMMNRPVVKAELVEWMRSSQTQVTGELAEVLNFAKENNIPVIPHETVLYF 60

Query: 61  RFLMQTLQPKHILEIGTAIGFSALLMAENAPEAKITTIDRNEEMIALAKENFAKYDNHNQ 120
            + L+  L+PK ILEIGTAIGFSAL+MA+  PEA+I TIDRN EMI LAK+N AKYD+ NQ
Sbjct: 61  QMLLSLLKPKRILEIGTAIGFSALVMAQEVPEAEIVTIDRNPEMIELAKKNLAKYDHRNQ 120

Query: 121 ITLLEGDAVDVLQTLDKSYDFVFMDSAKSKYIVFLPQVLKHLDVGGVVVLDDIFQGGDIA 180
            I L EGDA DVLQ L    +D VFMDSAKSKY+ FLP+ L+ L   G++++DD+FQ G+I
Sbjct: 121 IQLKEGDAADVLQELKGPFDLVFMDSAKSKYVEFLPKSLELLSENGLIILMDDVFQAGEIL 180
```

```
Query: 181 KPIDEVRRGQRTIYRGLQRLFDSTLQHPDLTATLVPLGDGLLMIRKN 227
           PI EV+R QR + RGL++LFD    +P     +++PLGDGLLMI+K+
Sbjct: 181 LPIMEVKRNQRALERGLRKLFDEVFDNPKYMTSVLPLGDGLLMIKKH 227
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6637> which encodes the amino acid sequence <SEQ ID 6638>. Analysis of this protein sequence reveals the following:

```
     Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38   Transmembrane  153 - 169 ( 152 - 170)

----- 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 databases:

```
>GP:CAA68045 GB:X99710 methyltransferase [Lactococcus lactis]
 Identities = 134/227 (59%), Positives = 169/227 (74%)

Query: 1   MVKSYSKTANHNMRRPVVKEELVHYMRTRQKQTTGFLAELEQFARQENIPIIQPEVVAYF 60
           MV++Y  T+N  M RPVVK ELV +MR+ Q Q TG LAE+  FA++ NIP+I  E V YF
Sbjct: 1   MVETYKSTSNPMMNRPVVKAELVEWMRSSQTQVTGELAEVLNFAKENNIPVIPHETVLYF 60

Query: 61  RFLLQSLQPKHILEIGTAIGFSALLMAENAPDATIVTIDRNREMIDFAKANFAKYDSRQQ 120
           + LL  L+PK ILEIGTAIGFSAL+MA+  P+A IVTIDRN EMI+ AK N AKYD R Q
Sbjct: 61  QMLLSLLKPKRILEIGTAIGFSALVMAQEVPEAEIVTIDRNPEMIELAKKNLAKYDHRNQ 120

Query: 121 IRLLEGDAADILSTLEGNFDFVFMDSAKSKYIVFLPEILRLLKVGGVVILDDVFQGGDIT 180
           I+L EGDAAD+L  L+G FD VFMDSAKSKY+ FLP+ L LL   G++++DDVFQ G+I
Sbjct: 121 IQLKEGDAADVLQELKGPFDLVFMDSAKSKYVEFLPKSLELLSENGLILMDDVFQAGEIL 180

Query: 181 KPIEDIRRGQRTIYRGLQSLFDATLTHPNLTTSLVPLSDGLLMIRKN 227
           PI +++R QR + RGL+ LFD    +P   TS++PL DGLLMI+K+
Sbjct: 181 LPIMEVKRNQRALERGLRKLFDEVFDNPKYMTSVLPLGDGLLMIKKH 227
```

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

```
 Identities = 177/235 (75%), Positives = 199/235 (84%)

Query: 1   MVQSYSKNANHNMRRPVVKEEIVQYMRQHQKQNNGCLAELEAFAKQENIPIIPHETATYF 60
           MV+SYSK ANHNMRRPVVKEE+V YMR  QKQ  G LAELE FA+QENIPII  E   YF
Sbjct: 1   MVKSYSKTANHNMRRPVVKEELVHYMRTRQKQTTGFLAELEQFARQENIPIIQPEVVAYF 60

Query: 61  RFLMQTLQPKHILEIGTAIGFSALLMAENAPEAKITTIDRNEEMIALAKENFAKYDNHNQ 120
           RFL+Q+LQPKHILEIGTAIGFSALLMAENAP+A I TIDRN EMI  AK NFAKYD+  Q
Sbjct: 61  RFLLQSLQPKHILEIGTAIGFSALLMAENAPDATIVTIDRNREMIDFAKANFAKYDSRQQ 120

Query: 121 ITLLEGDAVDVLQTLDKSYDFVFMDSAKSKYIVFLPQVLKHLDVGGVVVLDDIFQGGDIA 180
           I LLEGDA D+L TL+ ++DFVFMDSAKSKYIVFLP++L+ L VGGVV+LDD+FQGGDI
Sbjct: 121 IRLLEGDAADILSTLEGNFDFVFMDSAKSKYIVFLPEILRLLKVGGVVILDDVFQGGDIT 180

Query: 181 KPIDEVRRGQRTIYRGLQRLFDSTLQHPDLTATLVPLGDGLLMIRKNADHIVLED 235
           KPI+++RRGQRTIYRGLQ LFD+TL HP+LT +LVPL DGLLMIRKN   IVL D
Sbjct: 181 KPIEDIRRGQRTIYRGLQSLFDATLTHPNLTTSLVPLSDGLLMIRKNQADIVLPD 235
```

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

Example 2148

A DNA sequence (GBSx2264) was identified in *S.agalactiae* <SEQ ID 6639> which encodes the amino acid sequence <SEQ ID 6640>. This protein is predicted to be phosphoglycolate phosphatase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2193(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 8985> which encodes amino acid sequence <SEQ ID 8986> was also identified. This protein appears to be a hydrolase *i.e.* an exposed protein.

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

```
>GP:CAA91552 GB:Z67740 unidentified [Streptococcus pneumoniae]
 Identities = 39/117 (33%), Positives = 67/117 (56%), Gaps = 9/117 (7%)

Query: 98  KEQESRDSKIHLM-PYAKEILEWTKEQDIPNFMYTHKGASTHSVLETLQISHYFDEILTG 156
           KE E+R+ +  ++      ++LE    Q   +F+ +H+       +LE   I+ YF E++T
Sbjct: 25  KENEARELEHPILFEGVSDLLEDILNQGGRHFLVSHRNDQVLEILEKTSIAAYFTEVVTS 84

Query: 157 VSGFERKPHPQGINYLVKRYSLDKSMTYYIGDRPLDLEVAQNAGIKS------INLR 207
            SGF+RKP+P+ + YL ++Y +    +   IGDRP+D+E  Q AG+ +       +NLR
Sbjct: 85  SSGFKRKPNPESMLYLREKYQISSGLV--IGDRPIDIEAGQAAGLDTHLFTSIVNLR 139
```

SEQ ID 8986 (GBS240) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 2; 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 61 (lane 3; MW 51.5kDa).

GBS240-GST was purified as shown in Figure 225, lane 12.

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

Example 2149

A DNA sequence (GBSx2265) was identified in *S.agalactiae* <SEQ ID 6641> which encodes the amino acid sequence <SEQ ID 6642>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2620(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6643> which encodes the amino acid sequence <SEQ ID 6644>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2967(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 = 463/599 (77%), Positives = 541/599 (90%)

Query: 1   MSDNRSHIEEKYQWDLTTVFATDELWETEVVELTQAIDNAKGFSGHLLDSSQSLLEITEV 60
           M+DNRSH+EEKY WDL+T+FATD+ WE EV +L   ++ +KGF+GHLLDSS +LL++T+
Sbjct: 1   MTDNRSHLEEKYTWDLSTIFATDKDWEAEVSDLATEVEASKGFAGHLLDSSANLLKVTKT 60

Query: 61  ELDLSRRLEKVYVYASMKNDQDTTVAKYQEFQAKATALYAKFSETFSFYEPELLQLSESD 120
            L+L+RR+EKVYVYA MKNDQDTTVAKYQE+QAKA+ LYAKFSE FSFY+PE++ L + D
Sbjct: 61  YLELARRVEKVYVYAHMKNDQDTTVAKYQEYQAKASGLYAKFSEVFSFYDPEVMMLHQED 120

Query: 121 YQSFLLEMPDLQKYDHFFEKIFANKPHVLSQNEEELLAGASEIFGAAGETFEILDNADMV 180
           YQ+FL E P+L+ Y+HFF+K+F  + HVLSQ EEELLAGA EIF  A ETF ILDNAD+V
Sbjct: 121 YQAFLTETPELKVYNHFFDKLFQAREHVLSQAEEELLAGAQEIFNGAEETFSILDNADIV 180

Query: 181 FPVVKNAKGEEVELTHGNFISLMESSDRTVRKEAYQAMYSTYEQFQHTYAKTLQTNVKSQ 240
           FPVVKN KGE+VELTHGNFISLMES DR+VR+ AY+AMYSTYEQFQHTYAKTLQTNVK Q
Sbjct: 181 FPVVKNDKGEDVELTHGNFISLMESKDRSVRQAAYEAMYSTYEQFQHTYAKTLQTNVKVQ 240

Query: 241 NFKARVHHYQSARQSALSANFIPEEVYETLIKTVNHHLPLLHRYMKLRQKVLGLDDLKMY 300
           N+KARVH Y SARQ+A++ANFIPE VY+TL++TVN HLPLLHRY+KLRQ+VLGLDDLKMY
Sbjct: 241 NYKARVHKYDSARQAAMAANFIPEAVYDTLLETVNKHLPLLHRYLKLRQEVLGLDDLKMY 300

Query: 301 DVYTPLSQMDMSFTYDEALKKSEEVLAIFGEAYSERVHRAFTERWIDVHVNKGKRSGAYS 360
           DVYTPLS+ D++ . YDEAL+K+E+VLA+FG+ Y++RVHRAFTERWIDVHVNKGKRSGAYS
Sbjct: 301 DVYTPLSETDLAIGYDEALEKAEKVLAVFGKDYADRVHRAFTERWIDVHVNKGKRSGAYS 360

Query: 361 GGSYDTNAFMLLNWQDTLDNLYTLVHETGHSLHSTFTRENQPYVYGDYSIFLAEIASTTN 420
           GGSYDTNAF+LLNWQDTLDNLYTLVHETGHSLHSTFTRE QPYVYGDYSIFLAEIASTTN
Sbjct: 361 GGSYDTNAFILLNWQDTLDNLYTLVHETGHSLHSTFTRETQPYVYGDYSIFLAEIASTTN 420

Query: 421 ENILTETLLKEVKDDKNRFAILNHYLDGFKGTIFRQTQFAEFEHAIHVADQEGQVLTSEY 480
           ENI+TE LL EV+D+K RFAILNHYLDGF+GT+FRQTQFAEFEHAIH ADQ+G+VLTSEY
Sbjct: 421 ENIMTEALLNEVQDEKERFAILNHYLDGFRGTVFRQTQFAEFEHAIHQADQKGEVLTSEY 480

Query: 481 LNNLYAELNEKYYGLTKEDNHFIQYEWARIPHFYYNYYVFQYATGFAAANYLAERIVNGN 540
           LN LYA+LNEKYYGL+K+DNHFIQYEWARIPHFYYNYYV+QYATGFAAA+YLA++IV+G
Sbjct: 481 LNQLYADLNEKYYGLSKKDNHFIQYEWARIPHFYYNYYVYQYATGFAAASYLADKIVHGT 540

Query: 541 PEDKEAYLNYLKAGNSDYPLNVIAKAGVDMTSADYLDAAFRVFEERLVELENLVAKGVH 599
            +D + YL YLK+GNSDYPL VIAKAGVDM   DYL+AAF+VF+ERL ELE LV+KG+H
Sbjct: 541 QDDIDHYLAYLKSGNSDYPLEVIAKAGVDMEKGDYLEAAFKVFDERLTELEVLVSKGIH 599
```

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

Example 2150

A DNA sequence (GBSx2266) was identified in *S.agalactiae* <SEQ ID 6645> which encodes the amino acid sequence <SEQ ID 6646>. This protein is predicted to be competence protein. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2955(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:AAC23746 GB:AF052209 competence protein [Streptococcus pneumoniae]
 Identities = 127/269 (47%), Positives = 176/269 (65%), Gaps = 8/269 (2%)

Query: 1   MLIAKDKQGNLINLLESHPGKGQYFCPTCCSAVRLKAGRIMRRHFAHISLKNCQFYHENE 60
            M +A+D +G L+N+LE    K  Y CP C   + L+ G  +R HFAH SLK+C F+ ENE
Sbjct: 1   MFVARDARGELVNVLEDKLEKQAYTCPACGGQLHLRQGPSVRTHFAHKSLKDCDFFFENE 60

Query: 61  SNEHLQLKAKLYMSLSRENETMLEHHLPEINQIADLFVNETLALE----VQCSRLSEQRL 116
            S EHL  K  LY  L +E +  LE+ L E+ QIAD+FVN  LALE    V C +   + L
Sbjct: 61  SPEHLANKESLYHWLKKETKVQLEYPLSELKQIADVFVNGNLALESSVVVPCLK---KVL 117

Query: 117 RERTKAYLQADFQVRWLLGEKLWLKHRLTNLHKQFLQFSQSIGFHIWELDLRLEVLRLKY 176
            +ER++ Y    +QV WLLG+KLWLK RLT L   FL FSQ++GF++WELD   +VLRLKY
Sbjct: 118 KERSEGYRSQGYQVLWLLGQKLWLKERLTRLQAGFLYFSQNMGFYVWELDKGKQVLRLKY 177

Query: 177 LIYEDLRGHVYYLSKTCPL-SGDVLAFLKWPYQSKNLNFYKVKQDRNIRDYVRQQLRYGN 235
            LIY+DLRG ++Y  K      G +L  L+ PY+ + ++ + V +D++I  Y+RQQL Y N
Sbjct: 178 LIYQDLRGKLHYQIKEFSYGQGSLLEILRLPYKKQKISHFTVSEDKDICRYIRQQLYYQN 237

Query: 236 QFWLRKQEKAYLSGQNLLTQELMMFFPQI 264
             FW+++Q +AY  G+N+LT  L  ++PQI
Sbjct: 238 LFWMKEQAEAYQKGENILTYGLKEWYPQI 266
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6647> which encodes the amino acid sequence <SEQ ID 6648>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1034(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 = 154/312 (49%), Positives = 204/312 (65%), Gaps = 1/312 (0%)

Query: 1   MLIAKDKQGNLINLL-ESHPGKGQYFCPTCCSAVRLKAGRIMRRHFAHISLKNCQFYHEN 59
            +L A D +  LI+L+ +    K  + CP C S VRL+ G I R HFAH+ L +CQF  EN
Sbjct: 4   ILTALDDKNQLISLVTQPISTKPPFRCPACKSPVRLRQGTIRRPHFAHVQLAHCQFQAEN 63

Query: 60  ESNEHLQLKAKLYMSLSRENETMLEHHLPEINQIADLFVNETLALEVQCSRLSEQRLRER 119
            ES EHL LKAKLY SL R      +E +LPE+ QIADL+VN+ LALE+QCS L  +RL++R
Sbjct: 64  ESEEHLTLKAKLYTSLVRTEAVCIEKYLPELQQIADLWVNDKLALEIQCSPLPVERLKKR 123

Query: 120 TKAYLQADFQVRWLLGEKLWLKHRLTNLHKQFLQFSQSIGFHIWELDLRLEVLRLKYLIY 179
            TKAY +  + VRWLLG KLWL   LT L KQFL FS S+GFH+WELD     +LRLKYLI+
Sbjct: 124 TKAYQEKGYPVRWLLGRKLWLNTHLTALQKQFLYFSSSLGFHLWELDAAANLLRLKYLIH 183

Query: 180 EDLRGHVYYLSKTCPLSGDVLAFLKWPYQSKNLNFYKVKQDRNIRDYVRQQLRYGNQFWL 239
            EDL G V YL+KT  L  +++    + PYQ + L  Y+ K   N+    +++ L   +  WL
Sbjct: 184 EDLFGKVSYLTKTISLDHNIMEMFRLPYQQEILYSYQKKMTVNLSKRIQRALLARHPKWL 243

Query: 240 RKQEKAYLSGQNLLTQELMMFFPQIQPPRVDTDFCQITNSLTSFYQNFTNYYQKNKNNLD 299
            R+QEKAYLSG NLL       F+PQ +P +  + FCQI  +L  +Y++F  YY+K K+
Sbjct: 244 RRQEKAYLSGYNLLMLTTDAFYPQWRPVQSSSGFCQIKGNLRPYYESFKVYYKKEKDKKV 303

Query: 300 QTLYPPVFYDKI 311
            QTL+ P +Y K+
Sbjct: 304 QTLFSPKYYVKM 315
```

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

Example 2151

A DNA sequence (GBSx2267) was identified in *S.agalactiae* <SEQ ID 6649> which encodes the amino acid sequence <SEQ ID 6650>. This protein is predicted to be bicyclomycin resistance protein. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.33   Transmembrane   78 -  94 (  75 -  96)
   INTEGRAL    Likelihood = -8.33   Transmembrane  269 - 285 ( 267 - 287)
   INTEGRAL    Likelihood = -7.38   Transmembrane  290 - 306 ( 287 - 314)
   INTEGRAL    Likelihood = -7.06   Transmembrane  203 - 219 ( 199 - 225)
   INTEGRAL    Likelihood = -6.69   Transmembrane  157 - 173 ( 143 - 184)
   INTEGRAL    Likelihood = -6.42   Transmembrane   53 -  69 (  44 -  73)
   INTEGRAL    Likelihood = -6.42   Transmembrane  362 - 378 ( 357 - 381)
   INTEGRAL    Likelihood = -3.72   Transmembrane  242 - 258 ( 240 - 261)
   INTEGRAL    Likelihood = -3.24   Transmembrane  329 - 345 ( 328 - 346)
   INTEGRAL    Likelihood = -1.28   Transmembrane  107 - 123 ( 106 - 123)

----- 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 homology with the following sequences in the GENPEPT database.

```
>GP:CAA15047 GB:AJ235272 BICYCLOMYCIN RESISTANCE PROTEIN (bcr1)
          [Rickettsia prowazekii]
 Identities = 86/336 (25%), Positives = 159/336 (46%), Gaps = 28/336 (8%)

Query: 73  GKKNTVLLGLCLILMSGFISFFTSNFSLAMASRLLLGIGIGLYNSLSISIITDLYEADER 132
           G++  VLLGL + ++S  IS F+ N  + M +R +   G+ + + +  S+  D Y+  E
Sbjct: 70  GRRPIVLLGLFIYIVSSIISIFSFNIEMLMIARFIQAFGVSVGSVIGQSMARDSYQGAEL 129

Query: 133 ASMIGLRTASLNIGKALTTFIVGLVLA-IGVNYIYLVYLLVIPVFF-FFWKNVPEVENQT 190
           + +  + +  L    AL ++I G ++  +  +Y+++ + L   +    +++ +PE
Sbjct: 130 SYVYAILSPWLLFIPALGSYIGGYIIEYLSWHYVFIFFSLAGTILLALYYQILPETNYYI 189

Query: 191 HTLKASTTFDT-----KAALLMLITFLVGI---AYIGATVKIPTLLVTKYHYATSFSSNM 242
              ++S  F+      K  +L L  F++G     Y G  ++ P +L+ +     SF   +
Sbjct: 190 AFSQSSKYFEVFNIIIKDKMLWLYAFIIGAFNGIYYGFFIEAPFILIDQMRVLPSFYGKL 249

Query: 243 LTLLAFSGILVGSVFGKLVK---VFQEKTLLIMILAMGIGNVLFALANNQIIFIVAS--I 297
             LL+F+ I  G + G L+K   V+ +K + I  +    G +LFA+ +  + FI+ S
Sbjct: 250 AFLLSFASIFGGFLGGYLIKKRQVYDKKVMSIGFIFSLCGCILFAVDSFILEFILVSNVF 309

Query: 298 LIGASFVGTM-----SSVFFYISKNYAKEHNNFITSLALTAGNI-GVILTPLI--LTKLP 349
            I   F+  M      S+   I+  YA E    +T   TAG+I G I   +I  +T
Sbjct: 310 AIAMIFMPMMIHMIGHSLLIAITLRYALEDYATVTG---TAGSIFGAIYYVVIASVTYCV 366

Query: 350 SQLHLEPFMTPFLITSGLMVINV--FVYLVLMSKNK 383
           S++H E      L+   L + +V  F Y+ L+ K K
Sbjct: 367 SKIHGETISNFSLLCLVLSISSVISFYYICLLYKKK 402
```

A related GBS gene <SEQ ID 8987> and protein <SEQ ID 8988> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:      6.28
GvH: Signal Score (-7.5): -2.45
     Possible site: 25
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 10 value:  -8.33 threshold:  0.0
   INTEGRAL    Likelihood = -8.33   Transmembrane   78 -  94 (  75 -  96)
   INTEGRAL    Likelihood = -8.33   Transmembrane  269 - 285 ( 267 - 287)
   INTEGRAL    Likelihood = -7.38   Transmembrane  290 - 306 ( 287 - 314)
```

```
    INTEGRAL     Likelihood = -7.06   Transmembrane  203 - 219 ( 199 - 225)
    INTEGRAL     Likelihood = -6.69   Transmembrane  157 - 173 ( 143 - 184)
    INTEGRAL     Likelihood = -6.42   Transmembrane   53 -  69 (  44 -  73)
    INTEGRAL     Likelihood = -6.42   Transmembrane  362 - 378 ( 357 - 381)
    INTEGRAL     Likelihood = -3.72   Transmembrane  242 - 258 ( 240 - 261)
    INTEGRAL     Likelihood = -3.24   Transmembrane  329 - 345 ( 328 - 346)
    INTEGRAL     Likelihood = -1.28   Transmembrane  107 - 123 ( 106 - 123)
    PERIPHERAL   Likelihood =  3.71     140
 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 homology with the following sequences in the databases:

```
ORF01955(517 - 1449 of 1749)
EGAD|163303|RP603(70 - 402 of 407) bicyclomycin resistance protein {Rickettsia prowazekii}
OMNI|NT01RP0626 conserved hypothetical protein GP|3861147|emb|CAA15047.1||AJ235272
BICYCLOMYCIN RESISTANCE PROTEIN (bcr1) {Rickettsia prowazekii} PIR|E71665|E71665
bicyclomycin resistance protein (bcr1) RP603 - Rickettsia prowazekii
%Match = 5.9
%Identity = 26.5  %Similarity = 52.0
Matches = 85  Mismatches = 141  Conservative Sub.s = 82

474       504       534       564       594       624       654       684
SLVTIPAMMITIFVILSNFVVTKLGKKNTVLLGLCLILMSGFISFFTSNFSLAMASRLLLGIGIGLYNSLSISIITDLYE
                         |:: |||||  : ::|  || |: |  : | :|::   |: : : :  |:  | |:
MTSTLYFLGFAVGILSLGRLSDIYGRRPIVLLGLFIYIVSSIISIFSFNIEMLMIARFIQAFGVSVGSVIGQSMARDSYQ
             60        70        80        90       100       110       120

714       744       774       801       831       858       888
ADERASMIGLRTASLNIGKALTTFIVGLVLA-IGVNYIYLVYLLVIPVFF-FFWKNVPEVENQTHTLKAST---TFDT--
  | : :  : :  |    || ::| | ::  :  :|::: : |   ::: :::: :||        ::|     |:
GAELSYVYAILSPWLLFIPALGSYIGGYIIEYLSWHYVFIFFSLAGTILLALYYQILPETNYYIAFSQSSKYFEVFNIII
            140       150       160       170       180       190       200

933       954       984      1014      1044      1074      1095      1125
KAALLMLITFLVG---IAYIGATVKIPTLLVTKYHYATSFSSNMLTLLAFSGILVGSVFGKLVK---VFQEKTLLIMILA
|  :| |  |::|     | |  :: | :|: :     ||   :  ||:|: |: | : | |:|   |: :| : |  :
KDKMLWLYAFIIGAFNGIYYGFFIEAPFILIDQMRVLPSFYGKLAFLLSFASIFGGFLGGYLIKKRQVYDKKVMSIGFIF
            220       230       240       250       260       270       280

1155      1182      1209      1224      1254      1284      1311      1335
MGIGNVLFALANNQIIFI-VASIL-IGASFVGTM-----SSVFFYISKNYAKEHNNFITSLALTAGNI-GVILTPLI--L
   | :|||: :  : || |:::: |   |:  |      |::  |:  || |    :|  |   |:| | |   :|  :
SLCGCILFAVDSFILEFILVSNVFAIAMIFMPMMIHMIGHSLLIAITLRYALEDYATVTGTA---GSIFGAIYYVVIASV
            300       310       320       330       340           350       360

1365      1395      1419      1449      1479      1509      1539      1569
TKLPSQLHLEPFMTPFLITSGLMVINV--FVYLVLMSKNK*KVIRKDNFFRIVKVGEKMLIAKDKQGNLINLLESHPGKG
|   |::| |       |:   | : :|  | |: |: | |
TYCVSKIHGETISNFSLLCLVLSISSVISFYYICLLYKKKSIIIN
               380       390       400
```

There is also homology to SEQ ID 400

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

Example 2152

A DNA sequence (GBSx2268) was identified in *S.agalactiae* <SEQ ID 6651> which encodes the amino acid sequence <SEQ ID 6652>. This protein is predicted to be 16S pseudouridylate synthase (rsuA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2645(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:BAB06992 GB:AP001518 16S pseudouridylate synthase [Bacillus halodurans]
 Identities = 106/234 (45%), Positives = 141/234 (59%), Gaps = 1/234 (0%)

Query: 1   MRLDKLLGQAGFGSRNQVKKLICSRQVSVDGQIVTKDNVIVDSGLQSIFVGKERVCLKES 60
           MR+DK L   GFGSR  VKKL+ +  V V GQ +   +  V+   +SI V  E V  K
Sbjct: 1   MRIDKFLANMGFGSRKDVKKLLKTGAVRVQGQPIKDPSTHVEPESESITVYGEEVEYKPY 60

Query: 61  SYYLLYKPSGVVSAVRDSEHKTVIDLISEKDKVEGLYPIGRLDRDTEGLLIVTNNGPLGY 120
            Y ++ KP GV+ A  D EH+TVIDL+ E+++      P+GRLD+DT GLL++TN+G   +
Sbjct: 61  VYLMMNKPKGVICATEDLEHETVIDLLGEEERHYEPSPVGRLDKDTVGLLLITNDGKFNH 120

Query: 121 RMLHPKHHVAKTYYVEVNGFLERDAITFFEEGVVFDDGTKCKPAELTIDTANNDKSTARI 180
            ++ PKHHV KTY   V G +  + +  F  GVV DDG   KPA L I  A   +S   +
Sbjct: 121 WLMSPKHHVPKTYRALVEGHVTEEDVGAFSHGVVLDDGYVTKPATLHILEA-GARSHIEL 179

Query: 181 TITEGKFHQVKKMFLAYGVKVIYLRRISFGDLRLDMNLKPGQYRRLRDSEAAIL 234
           +TEGKFHQVK+MF A G +V+ L RI  G+L LD  L  G+YR L   E A+L
Sbjct: 180 ILTEGKFHQVKRMFQAVGKRVLELERIKIGNLLLDPELARGEYRELTKEEIALL 233
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6653> which encodes the amino acid sequence <SEQ ID 6654>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3310(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 = 111/194 (57%), Positives = 138/194 (70%)

Query: 1   MRLDKLLGQAGFGSRNQVKKLICSRQVSVDGQIVTKDNVIVDSGLQSIFVGKERVCLKES 60
           MRLDKLL     GSR+QVKKLI ++ V VD          VD GLQ I V  +RV   +
Sbjct: 1   MRLDKLLEGTKVGSRSQVKKLIKAQGVWVDHMPARNGRQNVDPGLQLIEVTGQRVTHPKH 60

Query: 61  SYYLLYKPSGVVSAVRDSEHKTVIDLISEKDKVEGLYPIGRLDRDTEGLLIVTNNGPLGY 120
           SY +L KPSGVVSA +D+ + TVID ++E+DK   LYP+GRLDRDTEGL+++T+NGPLG+
Sbjct: 61  SYIILNKPSGVVSAKKDTNYLTVIDQLAEEDKSPDLYPVGRLDRDTEGLVLLTDNGPLGF 120

Query: 121 RMLHPKHHVAKTYYVEVNGFLERDAITFFEEGVVFDDGTKCKPAELTIDTANNDKSTARI 180
           RMLHP HHV+KTY V VNG L  DA  FF  G+ F  G +C+PA+LTI  A+ D+S A +
Sbjct: 121 RMLHPSHHVSKTYLVTVNGLLAEDASDFFAAGICFPTGEQCQPAQLTILKADTDQSQASL 180

Query: 181 TITEGKFHQVKKMF 194
           TI+EGKFHQVK F
Sbjct: 181 TISEGKFHQVKKCF 194
```

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

Example 2153

A DNA sequence (GBSx2269) was identified in *S.agalactiae* <SEQ ID 6655> which encodes the amino acid sequence <SEQ ID 6656>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> 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 9745> which encodes amino acid sequence <SEQ ID 9746> was also identified.

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

```
>GP:BAA18872 GB:D90917 hypothetical protein [Synechocystis sp.]
 Identities = 197/318 (61%), Positives = 243/318 (75%)

Query: 22  MGLLVDGKWVDQWYDTASTGGKFVRTVTQFRHWVTKDGSAGPSGDAGFKAESGRYHLYVS 81
           MGLLV+G W DQWYDT STGG+FVR  +QFRHW+T DGS GP+G  GFKAE+GRYHLYVS
Sbjct: 1   MGLLVNGIWQDQWYDTESTGGRFVRHDSQFRHWITPDGSPGPTGHGGFKAEAGRYHLYVS 60

Query: 82  LACPWASRVLIMRKLKNLESHISISIVNPLMLENGWTFQEYKGVIPDMINQSQYLYQIYQ 141
           LACPWA R LI RKLK LE  I +S+V+ LM ENGWTF    GV+PD +  ++YLYQIY
Sbjct: 61  LACPWAHRTLIFRKLKGLEGMIDVSVVHWLMRENGWTFAPGPGVMPDPLFNAEYLYQIYT 120

Query: 142 ASQSDYTGRVTVPVLWDKKFHTIVNNESSEIMRMLNTAFNHITGNTDDYYPDSLQGQIDE 201
            + + Y+GRVTVP+LWDK+  TIVNNESSEI+R+ N+AF+ +   + DYYP +L+ QID
Sbjct: 121 RADAQYSGRVTVPILWDKQKQTIVNNESSEIIRIFNSAFDGLGAKSGDYYPKALRTQIDA 180

Query: 202 MNNFIYPKINNGVYKAGFATSQNVYQKEVETLFTALDQLEKHLSDNHYLVGEQFTEADIR 261
           +N+ IY  INNGVYK GFAT+Q  Y++ +  LF +LD LE  L  + YL G++ TEAD R
Sbjct: 181 LNDRIYHTINNGVYKCGFATTQTAYEEAIAPLFESLDWLEGILQGHQYLTGDEITEADWR 240

Query: 262 LFTTLVRFDTVYYGHFKCNLKALHDYPHLWHYTKRIYNLPGIAETVNFDHIKKHYYGSHK 321
           LFTTL+RFD VY GHFKCNL+ + DYP+LW Y + +Y+ PGIAETVNF HIK HYY SH
Sbjct: 241 LFTTLIRFDVVYVGHFKCNLRRIQDYPNLWRYLRDLYHQPGIAETVNFQHIKGHYYESHL 300

Query: 322 TINPTGIIPAGPNLDWTI 339
            INPTGI+P GP LD ++
Sbjct: 301 NINPTGIVPMGPALDLSL 318
```

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

SEQ ID 6656 (GBS655) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 143 (lane 2-4; MW 27kDa).

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

Example 2154

A DNA sequence (GBSx2270) was identified in *S.agalactiae* <SEQ ID 6657> which encodes the amino acid sequence <SEQ ID 6658>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
```

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1116(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:CAB12030 GB:Z99105 similar to glucosamine-6-phosphate isomerase
          [Bacillus subtilis]
 Identities = 112/243 (46%), Positives = 163/243 (66%), Gaps = 10/243 (4%)

Query: 1   MRVITVKNDIEGGKIAFTLLEEKMKAGAQT-LGLATGSSPITFYEEIVKS----NLDFSN 55
           M+++  ++  E  K++  +++E+++A     LGLATGS+P+  Y++++        +DFS
Sbjct: 1   MKILIAEHYEELCKLSAAIIKEQIQAKKDAVLGLATGSTPVGLYKQLISDYQAGEIDFSK 60

Query: 56  MVSINLDEYVGIAASNDQSYSYFMHKHLFDAKPFKENNL--PNGLAKDLKEEIKRYDAVI 113
           + + NLDEY G++ S+ QSY++FMH+HLF     + +++  P G    L+   K Y+ +I
Sbjct: 61  VTTFNLDEYAGLSPSHPQSYNHFMHEHLFQHINMQPDHIHIPQGDNPQLEAACKVYEDLI 120

Query: 114 N-ANPIDFQILGIGRNGHIGFNEPGTPFDITTHVVDLAPSTIEANSRFFNSIDD-VPKQA 171
             A  ID QILGIG NGHIGFNEPG+ F+  T VV L+ STI+AN+RFF      VP+ A
Sbjct: 121 RQAGGIDVQILGIGANGHIGFNEPGSDFEDRTRVVKLSESTIQANARFFGGDPVLVPRLA 180

Query: 172 LSMGIGSIMK-SKTIVLVAYGIEKAEAIASMIKGPITEDMPASILQKHDDVVIIVDEAAA 230
           +SMGI +IM+ SK IVL+A G EKA+AI  M +GP+T D+PASILQKH+ V +I D  AA
Sbjct: 181 ISMGIKTIMEFSKHIVLLASGEEKADAIQKMAEGPVTTDVPASILQKHNHVTVIADYKAA 240

Query: 231 SKL 233
            KL
Sbjct: 241 QKL 243
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6659> which encodes the amino acid sequence <SEQ ID 6660>. Analysis of this protein sequence reveals the following:

```
     Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.59   Transmembrane  174 - 190 ( 174 - 190)

----- Final Results -----
             bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB12030 GB:Z99105 similar to glucosamine-6-phosphate isomerase
          [Bacillus subtilis]
 Identities = 120/244 (49%), Positives = 162/244 (66%), Gaps = 12/244 (4%)

Query: 1   MKIIRVQDQIEGGKIAFTLLKDSL-AKGAKTLGLATGSSPISFYQEMVKS----PLDFSD 55
           MKI+  +   E  K++  ++K+ + AK    LGLATGS+P+  Y++++        +DFS
Sbjct: 1   MKILIAEHYEELCKLSAAIIKEQIQAKKDAVLGLATGSTPVGLYKQLISDYQAGEIDFSK 60

Query: 56  LTSINLDEYVGLSVESDQSYDYFMRQNLF---NAKPFKKNYLPNGLATDVEAEAKRYNQI 112
           +T+ NLDEY GLS    QSY++FM ++LF   N +P    ++P G    +EA  K Y  +
Sbjct: 61  VTTFNLDEYAGLSPSHPQSYNHFMHEHLFQHINMQP-DHIHIPQGDNPQLEAACKVYEDL 119

Query: 113 IAEHP-IDFQVLGIGRNGHIGFNEPGTSFEEETHVVDLQESTIEANSRFFTSIED-VPKQ 170
           I +   ID Q+LGIG NGHIGFNEPG+ FE+ T VV L ESTI+AN+RFF      VP+
Sbjct: 120 IRQAGGIDVQILGIGANGHIGFNEPGSDFEDRTRVVKLSESTIQANARFFGGDPVLVPRL 179

Query: 171 AISMGIASIMK-SEMIVLLAFGQEKADAIKGMVFGPITEHLPASILQKHDHVIVIVDEAA 229
           AISMGI +IM+ S+ IVLLA G+EKADAI+ M  GP+T  +PASILQKH+HV VI D  A
Sbjct: 180 AISMGIKTIMEFSKHIVLLASGEEKADAIQKMAEGPVTTDVPASILQKHNHVTVIADYKA 239
```

```
Query: 230 ASQL 233
           A +L
Sbjct: 240 AQKL 243
```

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

```
 Identities = 163/233 (69%), Positives = 201/233 (85%)

Query: 1   MRVITVKNDIEGGKIAFTLLEEKMKAGAQTLGLATGSSPITFYEEIVKSNLDFSNMVSIN 60
           M++I V++ IEGGKIAFTLL++ +  GA+TLGLATGSSPI+FY+E+VKS LDFS++ SIN
Sbjct: 1   MKIIRVQDQIEGGKIAFTLLKDSLAKGAKTLGLATGSSPISFYQEMVKSPLDFSDLTSIN 60

Query: 61  LDEYVGIAASNDQSYSYFMHKHLFDAKPFKENNLPNGLAKDLKEEIKRYDAVINANPIDF 120
           LDEYVG++  +DQSY YFM ++LF+AKPFK+N LPNGLA D++ E KRY+ +I  +PIDF
Sbjct: 61  LDEYVGLSVESDQSYDYFMRQNLFNAKPFKKNYLPNGLATDVEAEAKRYNQIIAEHPIDF 120

Query: 121 QILGIGRNGHIGFNEPGTPFDITTHVVDLAPSTIEANSRFFNSIDDVPKQALSMGIGSIM 180
           Q+LGIGRNGHIGFNEPGT F+  THVVDL  STIEANSRFF SI+DVPKQA+SMGI SIM
Sbjct: 121 QVLGIGRNGHIGFNEPGTSFEEETHVVDLQESTIEANSRFFTSIEDVPKQAISMGIASIM 180

Query: 181 KSKTIVLVAYGIEKAEAIASMIKGPITEDMPASILQKHDDVVIIVDEAAASKL 233
           KS+ IVL+A+G EKA+AI  M+ GPITE +PASILQKHD V++IVDEAAAS+L
Sbjct: 181 KSEMIVLLAFGQEKADAIKGMVFGPITEHLPASILQKHDHVIVIVDEAAASQL 233
```

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

Example 2155

A DNA sequence (GBSx2271) was identified in *S.agalactiae* <SEQ ID 6661> which encodes the amino acid sequence <SEQ ID 6662>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.12   Transmembrane  169 - 185 ( 161 - 194)
   INTEGRAL    Likelihood = -6.37   Transmembrane  151 - 167 ( 145 - 168)
   INTEGRAL    Likelihood = -5.15   Transmembrane   42 -  58 (  41 -  62)
   INTEGRAL    Likelihood = -1.59   Transmembrane  207 - 223 ( 207 - 224)
   INTEGRAL    Likelihood = -1.12   Transmembrane   24 -  40 (  23 -  40)

----- Final Results -----
               bacterial membrane --- Certainty=0.4248(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:AAF13747 GB:AF117351 unknown [Zymomonas mobilis]
 Identities = 88/216 (40%), Positives = 123/216 (56%)

Query: 9   QQLNILRAGVLGANDGIISVAGVVIGVASATHNLWIIFLSAASAILAGAFSMAGGEYVSV 68
           +Q+  LRA VLGANDGI+S + ++IGVASA  +   I L+  S ++AGA SMA GEYVSV
Sbjct: 17  RQMGWLRASVLGANDGILSTSSLMIGVASAHGSSGNILLAGMSGLIAGALSMAAGEYVSV 76

Query: 69  STQKDTEQAAVAREEKLLENNPELAKKSLVDIYLAKGESHEHAQWLVDKAFSKNAIEHLV 128
           S+Q D EQA VARE   L+ NP   K  L +IY+ +G   E A  + ++  + NA+E  +
Sbjct: 77  SSQHDMEQADVAREHAELKANPHAEKHELAEIYVERGLDRELALQVAEQLMAHNALEAHL 136

Query: 129 EEKYGIEFGEYTSPWHAAISSFIAFAIGSIFPTITILLLPFSVRIVGTVIIVIVSLLSTG 188
            ++ G+       P  AA++S I+F+ G+I P +T L  P  +  +I I+ L   G
Sbjct: 137 RDELGLTDSLIARPVQAALASAISFSGGAIVPFLTALFSPPEIINITISLISILCLAVLG 196

Query: 189 YVSAKLGQAPTVPAMRRNVMIGCLTMLATYVIGQLF 224
           V A LG A    A  R    G L M+ T  IG  F
Sbjct: 197 MVGAHLGGANVPKAALRVTFCGALAMIGTAAIGSFF 232
```

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 2156

A DNA sequence (GBSx2272) was identified in *S.agalactiae* <SEQ ID 6663> which encodes the amino acid sequence <SEQ ID 6664>. This protein is predicted to be S-adenosylmethionine tRNA ribosyltransferase (queA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3438(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:CAB14732 GB:Z99118 S-adenosylmethionine tRNA ribosyltransferase
          [Bacillus subtilis]
 Identities = 228/341 (66%), Positives = 279/341 (80%)

Query: 1   MNTNDFDFYLPEELIAQTPLEKRDASKLLVIDHKNKTMTDSHFDHILDELKPGDALVMNN 60
            M  + FDF LPE LIAQ PLE+RDAS+L+V+D      +TDS F HI+      GD LV+NN
Sbjct: 1   MKVDLFDFELPERLIAQVPLEQRDASRLMVLDKHTGELTDSSFKHIISFFNEGDCLVLNN 60

Query: 61  TRVLPARLYGEKQDTHGHVELLLLKNTEGDQWEVLAKPAKRLRVGTKVSFGDGRLIATVT 120
            TRVLPARL+G K+DT   VELLLLK   GD+WE LAKPAKR++ GT V+FGDGRL A  T
Sbjct: 61  TRVLPARLFGTKEDTGAKVELLLLKQETGDKWETLAKPAKRVKKGTVVTFGDGRLKAICT 120

Query: 121 KELEHGGRIVEFSYDGIFLEVLESLGEMPLPPYIHEKLEDRDRYQTVYAKENGSAAAPTA 180
            +ELEHGGR +EF YDGIF EVLESLGEMPLPPYI E+L+D++RYQTVY+KE GSAAAPTA
Sbjct: 121 EELEHGGRKMEFQYDGIFYEVLESLGEMPLPPYIKEQLDDKERYQTVYSKEIGSAAAPTA 180

Query: 181 GLHFTKELLEKIETKGVKLVYLTLHVGLGTFRPVSVDNLDEHEMHSEFYQLSKEAADTLN 240
            GLHFT+E+L++++ KGV++ ++TLHVGLGTFRPVS D ++EH MH+EFYQ+S+E A  LN
Sbjct: 181 GLHFTEEILQQLKDKGVQIEFITLHVGLGTFRPVSADEVEEHNMHAEFYQMSEETAAALN 240

Query: 241 AVKESGGRIVAVGTTSIRTLETIGSKFNGELKADSGWTNIFIKPGYQFKVVDAFSTNFHL 300
             V+E+GGRI++VGTTS RTLETI  + +G+ KA SGWT+IFI PGY+FK +D   TNFHL
Sbjct: 241 KVRENGGRIISVGTTSTRTLETIAGEHDGQFKASSGWTSIFIYPGYEFKAIDGMITNFHL 300

Query: 301 PKSTLVMLVSAFAGRDFVLEAYNHAVEERYRFFSFGDAMFV 341
            PKS+L+MLVSA AGR+ +L AYNHAVEE YRFFSFGDAM +
Sbjct: 301 PKSSLIMLVSALAGRENILRAYNHAVEEEYRFFSFGDAMLI 341
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6665> which encodes the amino acid sequence <SEQ ID 6666>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3864(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 = 297/341 (87%), Positives = 322/341 (94%)
```

```
Query: 1   MNTNDFDFYLPEELIAQTPLEKRDASKLLVIDHKNKTMTDSHFDHILDELKPGDALVMNN 60
           MNTN+FDF LPEELIAQTPLEKRD+SKLL+IDH+ KTM DSHFDHI+D+L PGDALVMNN
Sbjct: 1   MNTNNFDFELPEELIAQTPLEKRDSSKLLIIDHRQKTMVDSHFDHIIDQLNPGDALVMNN 60

Query: 61  TRVLPARLYGEKQDTHGHVELLLLKNTEGDQWEVLAKPAKRLRVGTKVSFGDGRLIATVT 120
           TRVLPARLYGEK DTHGHVELLLLKNT+GDQWEVLAKPAKRL+VG++V+FGDGRL AT+
Sbjct: 61  TRVLPARLYGEKPDTHGHVELLLLKNTQGDQWEVLAKPAKRLKVGSQVNFGDGRLKATII 120

Query: 121 KELEHGGRIVEFSYDGIFLEVLESLGEMPLPPYIHEKLEDRDRYQTVYAKENGSAAAPTA 180
            ELEHGGRIVEFSYDGIFLEVLESLGEMPLPPYIHEKLED +RYQTVYAKENGSAAAPTA
Sbjct: 121 DELEHGGRIVEFSYDGIFLEVLESLGEMPLPPYIHEKLEDAERYQTVYAKENGSAAAPTA 180

Query: 181 GLHFTKELLEKIETKGVKLVYLTLHVGLGTFRPVSVDNLDEHEMHSEFYQLSKEAADTLN 240
           GLHFT +LL+KIE KGV LVYLTLHVGLGTFRPVSVDNLDEH+MHSEFY LS+EAA TL
Sbjct: 181 GLHFTTDLLKKIEAKGVHLVYLTLHVGLGTFRPVSVDNLDEHDMHSEFYSLSEEAAQTLR 240

Query: 241 AVKESGGRIVAVGTTSIRTLETIGSKFNGELKADSGWTNIFIKPGYQFKVVDAFSTNFHL 300
            VK++GGR+VAVGTTSIRTLETIG KF G+++ADSGWTNIFIKPGYQFKVVDAFSTNFHL
Sbjct: 241 DVKQAGGRVVAVGTTSIRTLETIGGKFQGDIQADSGWTNIFIKPGYQFKVVDAFSTNFHL 300

Query: 301 PKSTLVMLVSAFAGRDFVLEAYNHAVEERYRFFSFGDAMFV 341
           PKSTLVMLVSAFAGRDFVLEAY HAV+E+YRFFSFGDAMFV
Sbjct: 301 PKSTLVMLVSAFAGRDFVLEAYRHAVDEKYRFFSFGDAMFV 341
```

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

Example 2157

A DNA sequence (GBSx2273) was identified in *S.agalactiae* <SEQ ID 6667> which encodes the amino acid sequence <SEQ ID 6668>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-14.22   Transmembrane   14 -  30 (   6 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.6689(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 6669> which encodes the amino acid sequence <SEQ ID 6670>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2655(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 = 126/195 (64%), Positives = 155/195 (78%), Gaps = 1/195 (0%)

Query: 160 MEERFDITETDYEYIGEHNNYVAAFSGAMSIDDMQKYSLVYSENTPAYALAERIGGMDSA 219
           M ERFDITETDYEY  EH+ YVA F+GAMSI DMQ+YSLVYSENTPAYALAER+GGM+ A
Sbjct: 1   MTERFDITETDYEYDQEHHAYVAQFNGAMSIPDMQEYSLVYSENTPAYALAERLGGMNKA 60

Query: 220 YSKFGRYGQSKGDIKNIQKNGNKVTTDYYIQVLDYLWKHRKKYDSLITYLEEAFPTDYYR 279
           Y  F RYG+  G I  I +NGNK+TT YY+QVLDYLW+H+ KY  ++ Y+ E+FP  YY+
Sbjct: 61  YQLFDRYGKVSGAITTIDRNGNKITTAYYLQVLDYLWQHQDKYKDILYYIGESFPDLYYK 120
```

```
Query: 280 ALIPSDVVVAQKPGYVREALNVGAIVKEEVPYIVAIYTAGLGGSTQEDSEINGVGLYQLE 339
             +P  V V QKPGYVREALNVGAIV EE PY++A+Y++GLGG+TQ   E+NG+G  QL
Sbjct: 121 TYLP-HVKVYQKPGYVREALNVGAIVCEESPYLIALYSSGLGGATQASEEVNGLGYVQLV 179

Query: 340 QLCFVINQWHRVNMN 354
           QL +VIN+W+R N+N
Sbjct: 180 QLPYVINEWYRGNLN 194
```

SEQ ID 6668 (GBS680) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 164 (lane 10-12; MW 64kDa) and in Figure 239 (lane 9; MW 64 kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 164 (lane 15; MW 40kDa) and in Figure 188 (lane 9; MW 40kDa). Purified GBS680-His is shown in Figure 242, lane 8. Purified GBS680-GST is shown in Figure 246, lanes 6 & 7.

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

Example 2158

A DNA sequence (GBSx2274) was identified in *S.agalactiae* <SEQ ID 6671> which encodes the amino acid sequence <SEQ ID 6672>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.57   Transmembrane    8 -  24 (   4 -  25)
   INTEGRAL    Likelihood = -2.13   Transmembrane   66 -  82 (  65 -  84)
   INTEGRAL    Likelihood = -1.65   Transmembrane  107 - 123 ( 107 - 125)
   INTEGRAL    Likelihood = -0.69   Transmembrane   36 -  52 (  36 -  52)
   INTEGRAL    Likelihood = -0.48   Transmembrane   89 - 105 (  89 - 105)

----- 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 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 2159

A DNA sequence (GBSx2275) was identified in *S.agalactiae* <SEQ ID 6673> which encodes the amino acid sequence <SEQ ID 6674>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.87   Transmembrane  108 - 124 (  97 - 133)
   INTEGRAL    Likelihood = -9.08   Transmembrane  181 - 197 ( 173 - 201)
   INTEGRAL    Likelihood = -7.43   Transmembrane  220 - 236 ( 216 - 248)
   INTEGRAL    Likelihood = -6.69   Transmembrane    6 -  22 (   3 -  28)
   INTEGRAL    Likelihood = -3.72   Transmembrane  401 - 417 ( 400 - 417)
   INTEGRAL    Likelihood = -3.35   Transmembrane  279 - 295 ( 278 - 295)
   INTEGRAL    Likelihood = -2.87   Transmembrane   31 -  47 (  30 -  50)
   INTEGRAL    Likelihood = -2.87   Transmembrane  244 - 260 ( 242 - 264)
   INTEGRAL    Likelihood = -0.80   Transmembrane   62 -  78 (  62 -  78)

----- 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:AAC21770 GB:U32694 H. influenzae predicted coding region HI0092
        [Haemophilus influenzae Rd]
 Identities = 232/416 (55%), Positives = 314/416 (74%), Gaps = 3/416 (0%)

Query: 4   TFTTTGALIGLALAILLIIKKVHPAYSLILGALVGGLIGGGDLVTIVNTMVLGAQGMMSS 63
           T +  GAL+ L +AI LI+KKV PAY +++GALVGGLIGG DL   V+ M+ GAQG+ ++
Sbjct: 3   TVSAIGALVALIVAIFLILKKVSPAYGMLVGALVGGLIGGADLSQTVSLMIGGAQGITTA 62

Query: 64  ILRILTSGILAGALIKTGSAEKIAESIIKKLGQQRAITALAIATMIICAVGVFIDIAVIT 123
           ++RIL +G+LAG LI++G+A  I E+I  KLG+ RA+ ALA+ATMI+ AVGVF+D+AVIT
Sbjct: 63  VMRILAAGVLAGVLIESGAANSITETITNKLGETRALLALALATMILTAVGVFVDVAVIT 122

Query: 124 VAPIALAIGKKANLSKSSILLAMIGGGKAGNIISPNPNTIAASEAFKVDLTSLMVQNIIP 183
           V+PIALA+ ++++LSK++ILLAMIGGGKAGNI+SPNPN IAA++ F + LTS+M+  IIP
Sbjct: 123 VSPIALALSRRSDLSKAAILLAMIGGGKAGNIMSPNPNAIAAADTFHLPLTSVMMAGIIP 182

Query: 184 AIAALVVTIILAKIVSKKNNDISYDSEEQV--GSDLPAFLPAISGPLVVICLLALRPLFG 241
           A+  L++T  LAK +  K + ++ D E  V    +LP+FL A+  PLV I LLALRPLF
Sbjct: 183 ALFGLILTYFLAKRLINKGSKVT-DKEVIVLETQNLPSFLTALVAPLVAILLLALRPLFD 241

Query: 242 ITIDPLIALPLGGLISILATGYLKETVPFVEYGLSKVVGVSILLIGTGTLSGIIKASNLQ 301
           I +DPLIALPLGGLI     G L+    +   GLSK+  V+I+L+GTG L+GII  S L+
Sbjct: 242 IKVDPLIALPLGGLIGAFCMGKLRNINSYAINGLSKMTPVAIMLLGTGALAGIIANSGLK 301

Query: 302 FDMIHLLEFLNMPTFILAPLSGIFMGAATASTTSGTTIASQTFAETLIKSGVPAVSGAAM 361
             +I  LE   +P++ILAP+SG+ M  ATASTT+GT +AS  F+ TL++ GV +++GAAM
Sbjct: 302 EVLIQGLEHSGLPSYILAPISGVLMSLATASTTAGTAVASNVFSSTLLELGVSSLAGAAM 361

Query: 362 IHAGATVLDSLPHGSFFHATGGAVNMAIKDRMKLISYEALIGLTSTIVAVVYYCFF 417
           IHAGATV D +PHGSFFHATGG+VNM IK+R+KLI YE+ +GL  TIV+ + +  F
Sbjct: 362 IHAGATVFDHMPHGSFFHATGGSVNMDIKERLKLIPYESAVGLMMTIVSTLIFGVF 417
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6675> which encodes the amino acid sequence <SEQ ID 6676>. Analysis of this protein sequence reveals the following:

```
     Possible site: 51
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.15   Transmembrane  240 - 256 ( 236 - 265)
   INTEGRAL    Likelihood =-10.88   Transmembrane    3 -  19 (   1 -  32)
   INTEGRAL    Likelihood =-10.14   Transmembrane  269 - 285 ( 263 - 289)
   INTEGRAL    Likelihood = -7.27   Transmembrane  107 - 123 ( 102 - 141)
   INTEGRAL    Likelihood = -7.17   Transmembrane  307 - 323 ( 303 - 330)
   INTEGRAL    Likelihood = -6.64   Transmembrane   24 -  40 (  23 -  43)
   INTEGRAL    Likelihood = -5.63   Transmembrane  422 - 438 ( 420 - 442)
   INTEGRAL    Likelihood = -3.77   Transmembrane  124 - 140 ( 124 - 141)
   INTEGRAL    Likelihood = -3.24   Transmembrane  189 - 205 ( 184 - 207)
   INTEGRAL    Likelihood = -2.60   Transmembrane   65 -  81 (  65 -  82)
   INTEGRAL    Likelihood = -2.34   Transmembrane  393 - 409 ( 393 - 409)
   INTEGRAL    Likelihood = -0.11   Transmembrane  149 - 165 ( 149 - 166)

----- Final Results -----
               bacterial membrane --- Certainty=0.5458(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:BAB07616 GB:AP001520 unknown conserved protein [Bacillus halodurans]
 Identities = 155/435 (35%), Positives = 248/435 (56%), Gaps = 21/435 (4%)

Query: 7   LGVLVGVIVIIYLYVKEVNIIIAAPLATSLVILFNQMDPTTTLLGKEPNQFMGALSTYIL 66
```

```
           LG+++G+++++ L  +  +II  AP+A  +V LF  +D    LL    + +M     +
Sbjct: 2   LGIVLGLVILMVLAYRGWSIIWVAPIAAGVVALFGGLD----LLPAYTDTYMEGFVNFAK 57

Query: 67  NYFAIFLLGSILAKLMETSGATTSIADYILKKVGHDSPYKVLVAIFLISAILTYGGISLF 126
            +F +F+LG+I  KLME +GA  S+A  I K +G     + ++ + L  A+LTYGGISLF
Sbjct: 58  QWFPVFMLGAIFGKLMEDTGAARSVASAITKLIGTK---RAILGVMLGCAVLTYGGISLF 114

Query: 127 VVMFAVLPLARSLFKKMDLAWNLIQVPLWLGIATFTMTILPGTPAIQNVIPIQYLDTSLT 186
           VV+FA+ PLA +LF++ +++  LI   + LG  TFTMT +PGTP IQN+IP  Y  T+
Sbjct: 115 VVVFAMYPLALALFREANISRRLIPGTIALGAFTFTMTAVPGTPQIQNLIPTSYYGTNAM 174

Query: 187 AAAIPSIVGSIGCVAFGLFYMKYCLAKSMARGETYATYAFDNEIQVKTKNLPHFLASILP 246
           AA +  ++ ++     G  Y+ +   K    GE + T   + E + + + +P+   S LP
Sbjct: 175 AAPMMGVIAALIMGIGGYTYLVWREKKLKEAGE-FFTEPKNGEKEEEGEKVPNPWLSFLP 233

Query: 247 LLLLIIIALTGSLFGNDFFKKNIIFIALLAVILTASWLFRQFIPNKIAVFNLGASSSIAP 306
           L+ +I+   T +L   D      I +AL++ I+    L    +  I   N GA  S+
Sbjct: 234 LVSVIV---TLNLLQWD------IVLALISGIVLIMLLNVGKVKGFIQSMNQGAGGSVLA 284

Query: 307 IFATASAVAFGAVVMIVPGFTFFSDLILNIPGNPLISLAVLTSSMSAITGSSSGALGIVM 366
           I  T++AV FG+VV  VPGF   ++L+L I G+PLIS AV  + ++  TGS+SG +GI +
Sbjct: 285 IINTSAAVGFGSVVRAVPGFERLTELLLGIQGSPLISQAVAINVLAGATGSASGGMGIAL 344

Query: 367 ----PNFAQYYLDQGLNPEMIHRVATIASNIFTIVPQSGVFLTFLALTGLNHKNAFKETF 422
                + Q  ++ G++PE  HRVA+IAS     +P +G  LT LA+TGL+HK ++K+ F
Sbjct: 345 EALGDRYMQLAMETGMSPEAFHRVASIASGGLDTLPHNGAVLTLLAITGLSHKESYKDIF 404

Query: 423 ITVSVSTFIAQVIVI 437
           +  V   ++   I
Sbjct: 405 VVGCVIPIVSVAFAI 419
```

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

```
 Identities = 88/395 (22%), Positives = 167/395 (42%), Gaps = 40/395 (10%)

Query: 9   GALIGLALAILLIIKKVHPAYSLILGALVGGLIGGGDLVTIV----NTMVLGAQG--MMS 62
           G L+G+ + I L +K+V+   + L   + L    D  T +         +GA    +++
Sbjct: 8   GVLVGVIVIIYLYVKEVNIIIAAPLATSLVILFNQMDPTTTLLGKEPNQFMGALSTYILN 67

Query: 63  SILRILTSGILAGALIKTGSAEKIAESIIKKLGQQ---RAITALAIATMIICAVGVFIDI 119
               L   ILA  +G+   IA+ I+KK+G     + + A+ + + I+   G+ + +
Sbjct: 68  YFAIFLLGSILAKLMETSGATTSIADYILKKVGHDSPYKVLVAIFLISAILTYGGISLFV 127

Query: 120 AVITVAPIALAIGKKANLSKSSILLAMIGGGKAGNII----SPNPNTIAASEAFKVDLTS 175
            + V P+A ++ KK +L+ + I + +  G     +     +P    +   +     LT+
Sbjct: 128 VMFAVLPLARSLFKKMDLAWNLIQVPLWLGIATFTMTILPGTPAIQNVIPIQYLDTSLTA 187

Query: 176 LMVQNIIPAIAALVVTII-----LAKIVSKKNNDISY--DSEEQVGS-DLPAFLPAISGP 227
            + +I+ +I  +    +      LAK +++     +Y  D+E QV + +LP FL +I
Sbjct: 188 AIPSIVGSIGCVAFGLFYMKYCLAKSMARGETYATYAFDNEIQVKTKNLPHFLASILPL 247

Query: 228 LVVICLLALRPLFG-------ITIDPLIALPLGGLISILATGYLKETVPFVEYGLSKVVG 280
           L++I +    LFG       I    L+A+ L    S L   ++   +     G S +
Sbjct: 248 LLLIIIALTGSLFGNDFFKKNIIFIALLAVIL--TASWLFRQFIPNKIAVFNLGASSSIA 305

Query: 281 ---VSILLIGTGTLSGIIKASNLQFDMIHLLEFLNMPTFILAPLSGIFMGAATASTTSGT 337
              +   + G +  I+       D+I  L     P   LA L+  SM A T S+SGA
Sbjct: 306 PIFATASAVAFGAVVMIVPGFTFFSDLI--LNIPGNPLISLAVLTS-SMSAITGSSSGAL 362

Query: 338 TIASQTFAETLIKSGVPAVSGAAMIHAGATVLDSL 372
            I   FA+  + G+      MIH  AT+  ++
Sbjct: 363 GIVMPNFAQYYLDQGL----NPEMIHRVATIASNI 393
```

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

Example 2160

A DNA sequence (GBSx2277) was identified in *S.agalactiae* <SEQ ID 6677> which encodes the amino acid sequence <SEQ ID 6678>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.24   Transmembrane   85 - 101 (  84 - 101)

----- Final Results -----
               bacterial membrane --- Certainty=0.2296(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:CAB16041 GB:Z99124 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 176/377 (46%), Positives = 234/377 (61%), Gaps = 2/377 (0%)

Query: 1   MKVVVAIDSLKGSLSSLEAGNAIKESINEVISGADVEVHPLADGGEGTVEALTLGMGGTI 60
           MK+++A DS K SLS+LEA  AI+     V  GAD    P+ADGGEGTV++L     G I
Sbjct: 1   MKIIIAPDSFKESLSALEAAEAIERGFKSVFPGADYRKLPVADGGEGTVQSLVDATNGRI 60

Query: 61  ETIPVKGPLGEKVHASYGIIPQRQLAIIEMAAAAGITLIATEERNPLHTTTYGVGEMIKD 120
              V GPLGE V A +G++    + A+IEMAAA+G+ L+  ++RNPL TTT G GE+I
Sbjct: 61  IEQVVTGPLGEPVRAFFGMMGDGRTAVIEMAAASGLHLVPVDKRNPLITTTRGTGELIGA 120

Query: 121 AISKGCRHFIIGIGGSATNDGGAGMLQALGYALLDKDNQEISLGAQGLADLKSISTDKVI 180
           A+  G    IIGIGGSATNDGGAGM+QALG  LLD    EI  G   L+ L SI    +
Sbjct: 121 ALDAGAERLIIGIGGSATNDGGAGMIQALGGRLLDNSGSEIGPGGGALSQLASIDVSGLD 180

Query: 181 EELKECDFKIACDVTNPLCGAQGCSSIFGPQKGADEDMITKMDTWLSNYATLATSVSEKA 240
             L+    ++AC+V NPL G +G +++FGPQKGA  DM+  +D  +S++A +A
Sbjct: 181 SRLRNVKLEVACNVDNPLTGPKGATAVFGPQKGATADMLDVLDQNVSHFADMAEKALGST 240

Query: 241 DATIEGTGAAGGLGFAFLAFTNATLEPGIDIILSEINIEKAISEADLVVTGEGRLDGQTV 300
              EG GAAGGLG++ L +  A L+ GIDI+L   ++ E  + +ADLV+TGEGR+D QTV
Sbjct: 241 FRDTEGAGAAGGLGWSLLTYLQADLKRGIDIVLEAVDFESIVQDADLVITGEGRIDSQTV 300

Query: 301 MGKAPIGVAKLAKKYGKKVVAFSGSVTEDAILCNQHGIDAFFPIVRRLISLDEAMSKEVA 360
            GK PIGVAK AK Y   V+  +GS++ D+    QHGIDA F IV   + L++A
Sbjct: 301 HGKTPIGVAKAAKSYDVPVIGIAGSISRDSNAVYQHGIDALFSIVPGAVPLEDAFEHAAE 360

Query: 361 YKNMKETATQVFRLINL 377
           Y  M+ TA  +   I L
Sbjct: 361 Y--MERTARDIAASIKL 375
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6679> which encodes the amino acid sequence <SEQ ID 6680>. Analysis of this protein sequence reveals the following:

```
      Possible site: 49

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.27   Transmembrane  360 - 376 ( 360 - 376)

----- 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 databases:

```
>GP:AAA57927 GB:U18997 ORF_f408 [Escherichia coli]
 Identities = 115/345 (33%), Positives = 182/345 (52%), Gaps = 25/345 (7%)

Query: 24  MKILVAIDSFKGSVTSPELNTSVAQALLSVDKQLVIETRAIADGGEGSLVALSQTVAGRW 83
```

```
            MKI++A DS+K S+++ E+  ++ +    +       +  +ADGGEG++ A+     G
Sbjct: 28   MKIVIAPDSYKESLSASEVAQAIEKGFREIFPDAQYVSVPVADGGEGTVEAMIAATQGAE 87

Query: 84   HQVKTIDLLRRPIKVAY--YRHAKQAFIESASIIGIDKITSNSVTYAQATSYGLGLAVKD 141
                     L   +  ++      K AFIE A+  G++ + +         TS G G  +
Sbjct: 88   RHAWVTGPLGEKVNASWGISGDGKTAFIEMAAASGLELVPAEKRDPLVTTSRGTGELILQ 147

Query: 142  AIQKGATQIEIMLGGTGTSDGGKGFLESLNYDFMT--------GRSYLDTLASPVTLLGL 193
             A++ GAT I I +GG+ T+DGG G +++L               G   L+TL + + + GL
Sbjct: 148  ALESGATNIIIGIGGSATNDGGAGMVQALGAKLCDANGNEIGFGGGSLNTL-NDIDISGL 206

Query: 194  T------------DVTNPYHGPQGFAAVFGPQKGGSLSQIEETDQIASNFAKKVFCQTTI 241
                         DVTNP  G  G + +FGPQKG S + I E D   S++A+ +     +
Sbjct: 207  DPRLKDCVIRVACDVTNPLVGDNGASRIFGPQKGASEAMIVELDNNLSHYAEVIKKALHV 266

Query: 242  DLQTIPGSGAAGGLGGAIV-LLGGTLTSGFSRIAELLNLDNSLQSCDLVITGEGCLDTQS 300
             D++ +PG+GAAGG+G A++  LG  L SG   +   LNL+  +  C LVITGEG +D+QS
Sbjct: 267  DVKDVPGAGAAGGMGAALMAFLGAELKSGIEIVTTALNLEEHIHDCTLVITGEGRIDSQS 326

Query: 301  QSGKVPVAIARMAKKYQVPTIALCGSVKIETGLAAEDFL-AVFSI 344
              GKVP+ +A +AKKY  P I + GS+  + G+  +  + AVFS+
Sbjct: 327  IHGKVPIGVANVAKKYHKPVIGIAGSLTDDVGVVHQHGIDAVFSV 371
```

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

```
 Identities = 128/379 (33%), Positives = 194/379 (50%), Gaps = 23/379 (6%)

Query: 1    MKVVVAIDSLKGSLSSLEAGNAIKESINEVISGADVEVHPLADGGEGTVEALTLGMGGTI 60
             MK++VAIDS KGS++S E   ++ +++  V     +E   +ADGGEG++ AL+  + G
Sbjct: 24   MKILVAIDSFKGSVTSPELNTSVAQALLSVDKQLVIETRAIADGGEGSLVALSQTVAGRW 83

Query: 61   ETIPVKGPLGEKVHASYGIIPQRQLAIIEMAAAAGITLIATEERNPLHTTTYGVGEMIKD 120
               +     L   +  +Y      + A IE A+  GI  I +          T+YG+G  +KD
Sbjct: 84   HQVKTIDLLRRPIKVAY--YRHAKQAFIESASIIGIDKITSNSVTYAQATSYGLGLAVKD 141

Query: 121  AISKGCRHFIIGIGGSATNDGGAGMLQALGYALLDKDNQEISLGAQGLADLKSISTDKVI 180
             AI KG      I +GG+ T+DGG G L++L Y  +             G + L ++++   +
Sbjct: 142  AIQKGATQIEIMLGGTGTSDGGKGFLESLNYDFMT----------GRSYLDTLASPVTL 190

Query: 181  EELKECDFKIACDVTNPLCGAQGCSSIFGPQKGADEDMITKMDTWLSNYATLATSVSEKA 240
               L          DVTNP  G QG +++FGPQKG      I + D   SN+A       +
Sbjct: 191  LGLT--------DVTNPYHGPQGFAAVFGPQKGGSLSQIEETDQIASNFAKKVFCQTTID 242

Query: 241  DATIEGTGAAGGLGFAFLAFTNATLEPGIDIILSEINIEKAISEADLVVTGEGRLDGQTV 300
               TI G+GAAGGLG A +      TL  G    I   +N++ ++   DLV+TGEG LD Q+
Sbjct: 243  LQTIPGSGAAGGLGGA-IVLLGGTLTSGFSRIAELLNLDNSLQSCDLVITGEGCLDTQSQ 301

Query: 301  MGKAPIGVAKLAKKYGKKVVAFSGSVTEDAILCNQHGIDAFFPIVRRLISLDEAMSKEVA 360
              GK P+ +A++AKKY    +A  GSV  +  L  +  + A F I ++ ISL+ A+ K
Sbjct: 302  SGKVPVAIARMAKKYQVPTIALCGSVKIETGLAAEDFL-AVFSIQQQPISLEAAIDKTTT 360

Query: 361  YKNMKETATQVFRLINLYN 379
              N+K  A  +  LI  +N
Sbjct: 361  LSNIKILAANLMLLIAQFN 379
```

SEQ ID 6678 (GBS409) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 7; MW 45.4kDa).

GBS409-His was purified as shown in Figure 214, lane 6.

GBS409d was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 166 (lane 3 & 4; MW 35kDa) and in Figure 188 (lane 12; MW 35kDa). Purified protein is shown in Figure 240, lanes 9-10.

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

Example 2161

A DNA sequence (GBSx2278) was identified in *S.agalactiae* <SEQ ID 6681> which encodes the amino acid sequence <SEQ ID 6682>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1886(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:AAC21771 GB:U32695 conserved hypothetical protein [Haemophilus influenzae Rd]
 Identities = 97/383 (25%), Positives = 175/383 (45%), Gaps = 52/383 (13%)

Query: 1   MKLRKQLAQQIVTSIKDVCQQDINFINTKGIIFASTNPKRVGEFHEIGLKVAQTGQMIEV 60
           M+L K  A++IV     +    +N ++  G+I AS N  R+ + H   +   +  +++E+
Sbjct: 1   MQLDKYTAKKIVKRAMKIIHHSVNVMDHDGVIIASGNSTRLNQRHTGAVLALRENRVVEI 60

Query: 61  TD---QESYFGTQAGINIPFYYNCELLATIGISGNPNQVGKYALLAQKMTRLILKEHE-L 116
               Q+  F  Q GIN+P +Y  + +  +GISG P QV +YA L +     LI+++   L
Sbjct: 61  DQALAQKWNFEAQPGINLPIHYLGKNIGVVGISGEPTQVKQYAELVKMTAELIVEQQALL 120

Query: 117 DYLDFGRKNEASIVLHHLVEGRELDYYYLNQFLNQYHLSEKTDYRLLTFEINSQKQKLLL 176
           +   + R+ +   +L              L+  LN   + ++   +  +F++N  +  +L+
Sbjct: 121 EQESWHRRYKEEFILQ-----------LLHCNLNWKEMEQQA--KFFSFDLNKSRVVVLI 167

Query: 177 S------QSEMSLLNFFDK-----------LDTAIYTFNYPNQYWLLLSDHMFDYYYPNI 219
                 +  +L+N+ ++             LD  +    + N    +LS  M
Sbjct: 168 KLLNPALDNLQNLINYLEQSEFAQDVAILSLDQVVVLKTWQNS--TVLSAQM------KT 219

Query: 220 LSKFECEKGLYKVGIGQKSSLSLLKR---SYETSILALK-ALKGQQK--VNLVDDLDLEL 273
           L   +  K  YK+ +G   +L L ++   S++++   L   LK   +  + + D+  L +
Sbjct: 220 LLPADYSKQDYKIAVGACLNLPLFEQLPLSFQSAQSTLSYGLKHHPRKGIYVFDEHRLPV 279

Query: 274 LLTSIDSNIKQYVLNKALVNL-SENDKIL---LNSYFKHNLSLKECSQELFIHKNTVQYR 329
           LL  +  + +   L K L  L SE + IL   L  YF  N  L    +++LF+H NT++YR
Sbjct: 280 LLAGLSHSWQGNELIKPLSPLFSEENAILYKTLQQYFLSNCDLYLTAEKLFVHPNTLRYR 339

Query: 330 LNKIYESTQLNPRNFKDATLLYL 352
           LNKI + T L       D   LYL
Sbjct: 340 LNKIEQITGLFFNKIDDKLTLYL 362
```

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 2162

A DNA sequence (GBSx2279) was identified in *S.agalactiae* <SEQ ID 6683> which encodes the amino acid sequence <SEQ ID 6684>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0290(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:AAF89979 GB:AF206272 beta-glucosidase [Streptococcus mutans]
 Identities = 334/475 (70%), Positives = 392/475 (82%), Gaps = 8/475 (1%)

Query: 4   FPKHFLWGGAVAANQVEGAFRTDGKGLSVQDVLPNGGLGD-------FTAKPTPDNLKLE 56
           FP++FLWGGA AANQ EGA+  DGKGLSVQDV P GG+          T KPT DNLKL
Sbjct: 6   FPENFLWGGATAANQFEGAYNQDGKGLSVQDVTPKGGVAQSGSSSPLITEKPTEDNLKLV 65

Query: 57  AIDFYHNYKNDIKLFAEMGFKVFRTSIAWSRIFPNGDDSAPNEAGLQFYDNLFDELLKYN 116
            IDFY+ YK DI LFAEMGFKVFR SIAW+RIFPNGDD  PNEAGL FYD +FDEL KY+
Sbjct: 66  GIDFYNRYKEDIALFAEMGFKVFRLSIAWTRIFPNGDDLEPNEAGLAFYDKVFDELAKYD 125

Query: 117 IEPLVTLSHYETPLHLAKTYNGWADRRLIAFFEKFAQTVMERYKDKVKYWLTFNEVNSIL 176
           IEPLVTLSHYETPLHLA+ YNGWA+R LIAF+E++A+TV  RYKDKVKYWLTFNEVNS+L
Sbjct: 126 IEPLVTLSHYETPLHLARKYNGWANRELIAFYERYARTVFTRYKDKVKYWLTFNEVNSVL 185

Query: 177 HMPFTSGAIMTDKSQLSPQELYQAIHHELVASARVTKLGRSINPNFKIGCMILAMPAYPM 236
           H PF SG I+TD  QLS Q+LYQA+HHELV SA  TK+G  INP+FKIGCM+LAMPAYPM
Sbjct: 186 HAPFMSGGIITDPEQLSKQDLYQAVHHELVVSALATKVGHEINPDFKIGCMVLAMPAYPM 245

Query: 237 TSDPRDVLAARQFEQHNLLFSDIHVRGKYPTYIQSYFKNNGIKIKFEEGDEEVLAQNTVD 296
           T+DP D LA R+FE  N LFSD+H RGKYP YI+ YFK+N I IK  EGD+E++ +NTVD
Sbjct: 246 TADPLDQLAVREFENQNYLFSDLHARGKYPNYIKRYFKDNNIDIKMGEGDKELMLENTVD 305

Query: 297 FLSFSYYMSVTQAYDFENYQSGQGNILGGLTNPHLTTSEWGWQIDPIGLRLVLNQYYERY 356
           F+SFSYYMSV  A++ E+Y SG+GN+LGGL+NP+L  SEWGWQIDP+GLRLVLN  Y+RY
Sbjct: 306 FISFSYYMSVAAAHNPEDYNSGRGNVLGGLSNPYLQASEWGWQIDPVGLRLVLNDSYDRY 365

Query: 357 QIPLPFIVENGLGAKDQLIETLDGDYTVEDDYRIDYMNQHLVQVAKAIEDGVEIMGYTSWG 416
           Q+PLPFIVENGLGAKD L++  DG  TVEDDYRIDY+ +HL+QV +A++DGV+++GYT+WG
Sbjct: 366 QLPLPFIVENGLGAKDVLVQGPDGP-TVEDDYRIDYLQKHLMQVGEALQDGVDLLGYTTWG 424

Query: 417 CIDCVSMSTAQLSKRYGLIYVDRNDDGTGSLQRYKKKSFGWYQKVIKTNGQSLFE 471
            ID VS ST +LSKRYG IYV  NDDG+GSL RYKKKSF WY+KVI+TNG SL+E
Sbjct: 425 PIDLVSESTVELSKRYGFIYVACNDDGSGSLARYKKKSFAWYKKVIETNGASLYE 479
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5287> which encodes the amino acid sequence <SEQ ID 5288>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0763(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 = 390/469 (83%), Positives = 423/469 (90%)

Query: 1   MTVFPKHFLWGGAVAANQVEGAFRTDGKGLSVQDVLPNGGLGDFTAKPTPDNLKLEAIDF 60
           M +FPK FLWGGAVAANQVEGAF  D KGLSVQDVLPNGGLG++T  PT DNL LEAIDF
Sbjct: 1   MGIFPKDFLWGGAVAANQVEGAFEADAKGLSVQDVLPNGGLGEWTDSPTSDNLTLEAIDF 60

Query: 61  YHNYKNDIKLFAEMGFKVFRTSIAWSRIFPNGDDSAPNEAGLQFYDNLFDELLKYNIEPL 120
           YH YK DI LFAEMGFKVFRTSIAWSRIFPNGDD  PNEAGLQFYD+LFDELL Y IEPL
Sbjct: 61  YHRYKEDIALFAEMGFKVFRTSIAWSRIFPNGDDDQPNEAGLQFYDDLFDELLNYGIEPL 120

Query: 121 VTLSHYETPLHLAKTYNGWADRRLIAFFEKFAQTVMERYKDKVKYWLTFNEVNSILHMPF 180
           VTLSHYETPLHLAK YNGW DRRLI FFE+FAQTVMERYKDKVKYWLTFNEVNSILHMPF
Sbjct: 121 VTLSHYETPLHLAKAYNGWTDRRLIGFFERFAQTVMERYKDKVKYWLTFNEVNSILHMPF 180
```

```
Query: 181 TSGAIMTDKSQLSPQELYQAIHHELVASARVTKLGRSINPNFKIGCMILAMPAYPMTSDP 240
           TSG IMT+K +LS Q+LYQAIHHELVASA VTKL   INP+ K+GCMILAMPAYPMTSDP
Sbjct: 181 TSGGIMTEKEKLSLQDLYQAIHHELVASASVTKLAHEINPDVKVGCMILAMPAYPMTSDP 240

Query: 241 RDVLAARQFEQHNLLFSDIHVRGKYPTYIQSYFKNNGIKIKFEEGDEEVLAQNTVDFLSF 300
           RD+LAA  FE  NLLFSDIHVRGKYP+YI+SYFK NGI+I  FE+GD+E+LA++TVDFLSF
Sbjct: 241 RDILAAHAFENLNLLFSDIHVRGKYPSYIKSYFKENGIEIVFEDGDKELLAEHTVDFLSF 300

Query: 301 SYYMSVTQAYDFENYQSGQGNILGGLTNPHLTTSEWGWQIDPIGLRLVLNQYYERYQIPL 360
           SYYMSVTQA++ E Y SGQGNILGGL+NP+L +SEWGWQIDPIGLRLVLNQYY+RYQIPL
Sbjct: 301 SYYMSVTQAHNPEAYTSGQGNILGGLSNPYLESSEWGWQIDPIGLRLVLNQYYDRYQIPL 360

Query: 361 FIVENGLGAKDQLIETLDGDYTVEDDYRIDYMNQHLVQVAKAIEDGVEIMGYTSWGCIDC 420
           FIVENGLGAKDQL++T DG  TV DDYRIDYM+QHLVQVAKAIEDGVE+MGYTSWGCIDC
Sbjct: 361 FIVENGLGAKDQLVQTADGSMTVHDDYRIDYMSQHLVQVAKAIEDGVEVMGYTSWGCIDC 420

Query: 421 VSMSTAQLSKRYGLIYVDRNDDGTGSLQRYKKKSFGWYQKVIKTNGQSL 469
           VSMSTAQLSKRYG IYVDRNDDGTG L RYKKKSF WY++VI+TNG+ L
Sbjct: 421 VSMSTAQLSKRYGFIYVDRNDDGTGQLTRYKKKSFDWYRQVIQTNGRYL 469
```

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

Example 2163

A DNA sequence (GBSx2280) was identified in *S.agalactiae* <SEQ ID 6685> which encodes the amino acid sequence <SEQ ID 6686>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.40   Transmembrane  247 - 263 ( 241 - 273)
   INTEGRAL    Likelihood = -8.55   Transmembrane  429 - 445 ( 424 - 450)
   INTEGRAL    Likelihood = -4.88   Transmembrane  285 - 301 ( 280 - 303)
   INTEGRAL    Likelihood = -3.82   Transmembrane  207 - 223 ( 205 - 225)
   INTEGRAL    Likelihood = -3.40   Transmembrane  113 - 129 ( 112 - 139)
   INTEGRAL    Likelihood = -1.97   Transmembrane  309 - 325 ( 305 - 328)
   INTEGRAL    Likelihood = -1.59   Transmembrane  395 - 411 ( 395 - 411)
   INTEGRAL    Likelihood = -1.49   Transmembrane  174 - 190 ( 173 - 193)

----- 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:CAA84286 GB:Z34526 beta-glucoside permease [Bacillus subtilis]
 Identities = 225/594 (37%), Positives = 351/594 (58%), Gaps = 11/594 (1%)

Query: 4   YQETAKAILAAVGGEKNIQHVTHCVTRLRLVLDNDEIVNDQVIKTIPNVIGVMRKNDQYQ 63
           Y + +K IL  VGGE+N+Q V HC+TRLR  L ++   +   ++ +P V+G     +Q+Q
Sbjct: 3   YDKLSKDILQLVGGEENVQRVIHCMTRLRFNLHDNAKADRSQLEQLPGVMGTNISGEQFQ 62

Query: 64  IILGNDVNNYYNAFLALGHFENTTREFSSQKKSSILEKLIETIAGVITPLIPALLGGGML 123
           II+GNDV   Y A +   +  +     SS +K ++L  + + I+GV TP++PA+ G GM+
Sbjct: 63  IIIGNDVPKVYQAIVRHSNLSDEKSAGSSSQKKNVLSAVFDVISGVFTPILPAIAGAGMI 122

Query: 124 KVIGILLPMLGIASSSSQTVAFINFFGDAAYYFMPIMIAYSAASRFKVTPVLAATVGGIL 183
           K +  L    G  +  SQ    +   GD A+YF+P+++A SAA +F   P +AA +   +
Sbjct: 123 KGLVALAVTFGWMAEKSQVHVILTAVGDGAFYFLPLLLAMSAARKFGSNPYVAAAIAAAI 182

Query: 184 LHPAFVTMVAEGKPLSLFGAPVTLASYGSSVIPILIMVFLMQYIERWINKIVPSVMKSFL 243
           LHP    ++  GKP+S  G PVT A+Y S+VIPIL+ +++  Y+E+WI++   + +K  +
Sbjct: 183 LHPDLTALLGAGKPISFIGLPVTAATYSSTVIPILLSIWIASYVEKWIDRFTHASLKLIV 242

Query: 244 QPTLIILISGFLALVVVGPLGVIIGKGLSSAMLSIYHVAPWLALSILGAIMPLVVMTGMH 303
```

```
            PT  +LI    L L+ VGPLG I+G+ LSS +  ++  A  +A+ +L     L++MTGMH
Sbjct: 243 VPTFTLLIVVPLTLITVGPLGAILGEYLSSGVNYLFDHAGLVAMILLAGTFSLIIMTGMH 302

Query: 304 WAFAPIFLAASVATPDVLILPAMLASNLAQGAASLAVAVKAKQKQTRQVAFAAGLSALLA 363
           +AF PI +           +LPAM  +N+ Q  AS AV ++++ K+ + +A    ++AL+
Sbjct: 303 YAFVPIMINNIAQNGHDYLLPAMFLANMGQAGASFAVFLRSRNKKFKSLALTTSITALM- 361

Query: 364 GITEPALYGVTLKFKKPLYAAMISGGLVGAYIGLVNIASYTFVVPSIIGLPQYINPQGGN 423
           GITEPA+YGV ++ KKP  AA+I G   GA+ G+  +ASY  +V    GLP  I    G
Sbjct: 362 GITEPAMYGVNMRLKKPFAAALIGGAAGGAFYGMTGVASY--IVGGNAGLPS-IPVFIGP 418

Query: 424 NFSNAVIAAIATIILTFIITWFLGIDEGENEKSSINAQEHTHIRSGLSKKETLYSPMVGN 483
            F  A+I  +          + LG ++  ++ S    Q   H  S    +E ++SP+ G
Sbjct: 419 TFIYAMIGLVIAFAAETAAAYLLGFEDVPSDGSQ---QPAVHEGS----REIIHSPIKGE 471

Query: 484 VLPLSKVPDETFSSKLLGEGLAITPSVGEVYAPFDGEIISLFPTKHAIALKDDKGVEVLI 543
           V  LS+V D   FS+ ++G+G AI P  GEV +P  G + ++F TKHAI +  D+G E+LI
Sbjct: 472 VKALSEVKDGVFSAGVMGKGFAIEPEEGEVVSPVRGSVTTIFKTKHAIGITSDQGAEILI 531

Query: 544 HIGIDTVELNGEGFEQLVKVGDFVKRGQLLLRMDIDFISSKGYSLISPVVVTNS 597
           HIG+DTV+L G+ F    +K GD V  G  L+  D++ I + GY +I+PV+VTN+
Sbjct: 532 HIGLDTVKLEGQWFTAHIKEGDKVAPGDPLVSFDLEQIKAAGYDVITPVIVTNT 585
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2883> which encodes the amino acid sequence <SEQ ID 2884>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.40   Transmembrane  246 - 262 ( 240 - 271)
   INTEGRAL    Likelihood = -6.26   Transmembrane  284 - 300 ( 279 - 304)
   INTEGRAL    Likelihood = -4.14   Transmembrane  173 - 189 ( 172 - 194)
   INTEGRAL    Likelihood = -3.24   Transmembrane  112 - 128 ( 111 - 137)
   INTEGRAL    Likelihood = -2.39   Transmembrane  428 - 444 ( 425 - 445)
   INTEGRAL    Likelihood = -2.13   Transmembrane  383 - 399 ( 380 - 401)
   INTEGRAL    Likelihood = -1.97   Transmembrane  308 - 324 ( 304 - 327)

----- 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>
```

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

```
 Identities = 508/619 (82%), Positives = 561/619 (90%), Gaps = 1/619 (0%)

Query: 4   YQETAKAILAAVGGEKNIQHVTHCVTRLRLVLDNDEIVNDQVIKTIPNVIGVMRKNDQYQ 63
           YQETAKAILAAVGG+ NIQ VTHCVTRLRLVL  DE V DQ +K I NVIGVMRKN QYQ
Sbjct: 3   YQETAKAILAAVGGKTNIQRVTHCVTRLRLVLKNDEKVKDQQVKAISNVIGVMRKNGQYQ 62

Query: 64  IILGNDVNNYYNAFLALGHFENTTREFSSQKKSSILEKLIETIAGVITPLIPALLGGGML 123
           IILGNDVNNYY AFL+LGHF+N   + SS+ K SILE+LIETIAGVITPLIPALLGGGML
Sbjct: 63  IILGNDVNNYYQAFLSLGHFDNQDEDHSSKAKGSILERLIETIAGVITPLIPALLGGGML 122

Query: 124 KVIGILLLPMLGIASSSSQTVAFINFFGDAAYYFMPIMIAYSAASRFKVTPVLAATVGGIL 183
           KV+GILLLPMLG+AS+ SQTVAFINFFGDAAYYFMP+MIAYSAA+RFKVTPVLAAT+ GIL
Sbjct: 123 KVVGILLLPMLGLASADSQTVAFINFFGDAAYYFMPVMIAYSAAARFKVTPVLAATIAGIL 182

Query: 184 LHPAFVTMVAEGKPLSLFGAPVTLASYGSSVIPILIMVFLMQYIERWINKIVPSVMKSFL 243
           LHPAFV MVAEGKPL+LFGAPVT ASYGSSVIPIL+MV+LMQYIE+W+N++VPSVMKSFL
Sbjct: 183 LHPAFVAMVAEGKPLTLFGAPVTPASYGSSVIPILMMVYLMQYIEKWVNRLVPSVMKSFL 242

Query: 244 QPTLIILISGFLALVVVGPLGVIIGKGLSSAMLSIYHVAPWLALSILGAIMPLVVMTGMH 303
           QPTLIILISGFLALVVVGPLGVIIG+GLS+ ML+IYHVAPWLAL+ILGAIMPLVVMTGMH
Sbjct: 243 QPTLIILISGFLALVVVGPLGVIIGQGLSNTMLAIYHVAPWLALAILGAIMPLVVMTGMH 302

Query: 304 WAFAPIFLAASVATPDVLILPAMLASNLAQGAASLAVAVKAKQKQTRQVAFAAGLSALLA 363
           WAFAPIFLAASVATPDVLILPAMLASNLAQGAASLAVA K KQKQTRQVA AAG+SALLA
```

```
Sbjct: 303 WAFAPIFLAASVATPDVLILPAMLASNLAQGAASLAVAFKTKQKQTRQVALAAGISALLA 362

Query: 364 GITEPALYGVTLKFKKPLYAAMISGGLVGAYIGLVNIASYTFVVPSIIGLPQYINPQGGN 423
           GITEPALYGVTLKFKKPLYAAMISGGLVGA+IG VNIASYTFVVPSIIGLPQYINP GG
Sbjct: 363 GITEPALYGVTLKFKKPLYAAMISGGLVGAFIGFVNIASYTFVVPSIIGLPQYINPSGGA 422

Query: 424 NFSNAVIAAIATIILTFIITWFLGIDEGENEKSSINAQEHTHIRSGLSKKETLYSPMVGN 483
           NF+NA+IA  ATI+L F +TWF+GIDE E+ K    A + + ++SGLS K+TLY+PM G
Sbjct: 423 NFTNALIAGTATIVLAFSLTWFMGIDE-ESPKQVSVAADMSQVKSGLSTKQTLYAPMTGE 481

Query: 484 VLPLSKVPDETFSSKLLGEGLAITPSVGEVYAPFDGEIISLFPTKHAIALKDDKGVEVLI 543
           +L LS+VPDETFSSKLLGEG AI PS GEVYAPFDGE+I+ FPTKHA+ALK+ +GVEVLI
Sbjct: 482 MLFLSEVPDETFSSKLLGEGFAILPSEGEVYAPFDGEVITFFPTKHAVALKNTRGVEVLI 541

Query: 544 HIGIDTVELNGEGFEQLVKVGDFVKRGQLLLRMDIDFISSKGYSLISPVVVTNSIDQLEI 603
           H+GIDTVEL G+GFEQLV VGD VKRGQ LL+MDIDFI+SKGYSLISPVVVTNS +QLEI
Sbjct: 542 HVGIDTVELKGQGFEQLVSVGDVVKRGQALLKMDIDFITSKGYSLISPVVVTNSAEQLEI 601

Query: 604 IVKDAETMVTNEDDLLVIL 622
           I++D + MVT ED LLVIL
Sbjct: 602 IIQDDKKMVTKEDALLVIL 620
```

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

Example 2164

A DNA sequence (GBSx2281) was identified in *S.agalactiae* <SEQ ID 6687> which encodes the amino acid sequence <SEQ ID 6688>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1148(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:CAB15944 GB:Z99124 transcriptional antiterminator (BglG family)
           [Bacillus subtilis]
 Identities = 118/275 (42%), Positives = 183/275 (65%)

Query: 1   MIIKRVLNHNAVISVTHQGLDVLLMGKGIAFKKRIGDRINSDAIEKSFVLKNSDNMNRFT 60
           M I +V+N+N +  V  QG ++++MG+G+AF+K+ GD ++   IEK F L N D   +F
Sbjct: 1   MKIAKVINNNVISVVNEQGKELVVMGRGLAFQKKSGDDVDEARIEKVFTLDNKDVSEKFK 60

Query: 61  ELFITVPEEVVACSERIINLGKIKLGKNLDEILYINLTDHIHSAIERHEQGMVIQNPLRL 120
            L   +P E +  SE II+  K++LGK L++ +Y++LTDHI+ AI+R+++G+ I+N L
Sbjct: 61  TLLYDIPIECMEVSEEIIHYAKLQLGKKLNDSIYVSLTDHINFAIQRNQKGLDIKNALLW 120

Query: 121 EIQRYYPDEYSIGMKALELIKDELGICLTIDESAFIAMHFVNAGLDNPFNEAHKITEIVS 180
           E +R Y DE++IG +AL ++K++ G+ L  DE+ FIA+H VNA L+         IT+++
Sbjct: 121 ETKRLYKDEFAIGKEALVMVKNKTGVSLPEDEAGFIALHIVNAELNEEMPNIINITKVMQ 180

Query: 181 YIEQKVKIDFRTELDESSIDYYRFMTHTKLFAQRVLSGMKYEDDDADLLLVVKKKYPREY 240
            I   VK  F+ E +E S+ YYRF+TH K FAQR+ +G   E  D  LL  VK+KY R Y
Sbjct: 181 EILSIVKYHFKIEFNEESLHYYRFVTHLKFFAQRLFNGTHMESQDDFLLDTVKEKYHRAY 240

Query: 241 KCVKEIGNNMAIQYQYQLNSSELLYLTVHVKRLVK 275
           +C K+I   +  +Y+++L S ELLYLT+H++R+VK
Sbjct: 241 ECTKKIQTYIEREYEHKLTSDELLYLTIHIERVVK 275
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6689> which encodes the amino acid sequence <SEQ ID 6690>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0680(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 = 220/279 (78%), Positives = 246/279 (87%)

Query: 1   MIIKRVLNHNAVISVTHQGLDVLLMGKGIAFKKRIGDRINSDAIEKSFVLKNSDNMNRFT 60
           M+IKRVLNHNA IS  HQGLD+LLMGKGI F K++GD I  +AIE SFVLKNSDNMNRFT
Sbjct: 1   MLIKRVLNHNAAISTNHQGLDILLMGKGITFGKKVGDSIELNAIETSFVLKNSDNMNRFT 60

Query: 61  ELFITVPEEVVACSERIINLGKIKLGKNLDEILYINLTDHIHSAIERHEQGMVIQNPLRL 120
           ELFITVP+EVVACSERIINLGKIKLGK LDEILYINLTDHIHSAIERHEQGM+I NPLR
Sbjct: 61  ELFITVPQEVVACSERIINLGKIKLGKTLDEILYINLTDHIHSAIERHEQGMLIHNPLRW 120

Query: 121 EIQRYYPDEYSIGMKALELIKDELGICLTIDESAFIAMHFVNAGLDNPFNEAHKITEIVS 180
           EIQRYYPDEYS+G+KALELI+  LG+ L IDE+AFIAMHFVNA LD PF E H++TEIVS
Sbjct: 121 EIQRYYPDEYSLGVKALELIERNLGVTLAIDEAAFIAMHFVNASLDTPFKEPHRLTEIVS 180

Query: 181 YIEQKVKIDFRTELDESSIDYYRFMTHTKLFAQRVLSGMKYEDDDADLLLVVKKKYPREY 240
           YIEQK+K DF+TELD++SIDYYRFMTH KLFAQRVLS M Y+DDDA+LLLVVK KYP+EY
Sbjct: 181 YIEQKIKTDFKTELDDTSIDYYRFMTHIKLFAQRVLSQMSYDDDDAELLLVVKTKYPKEY 240

Query: 241 KCVKEIGNNMAIQYQYQLNSSELLYLTVHVKRLVKNLKE 279
           +CV +I    +  +Y Y LNSSELLYLTVHVKRLVK+LKE
Sbjct: 241 RCVLDISEEIKKRYNYHLNSSELLYLTVHVKRLVKHLKE 279
```

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

Example 2165

A DNA sequence (GBSx2282) was identified in *S.agalactiae* <SEQ ID 6691> which encodes the amino acid sequence <SEQ ID 6692>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1104(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 9335> which encodes amino acid sequence <SEQ ID 9336> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6693> which encodes the amino acid sequence <SEQ ID 6694>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3314(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 = 143/178 (80%), Positives = 161/178 (90%)

Query: 1   MTLHHDKHHATYVANANAALEKHPEIGEDLEALLADVSQIPEDIRQAVINNGGGHLNHAL 60
           MTLHHDKHHATYVAN NAALEKHPEIGE+LE LLADV++IPEDIRQ +INNGGGHLNHAL
Sbjct: 24  MTLHHDKHHATYVANTNAALEKHPEIGENLEELLADVTKIPEDIRQTLINNGGGHLNHAL 83

Query: 61  FWELMSPEETQISQELSEDINATFGSFEDFKAAFTAAATGRFGSGWAWLVVNAEGKLEVL 120
           FWEL+SPE+  ++ ++++ I+  FGSF+ FK  FTAAATGRFGSGWAWLVVN EG+LE+
Sbjct: 84  FWELLSPEKQDVTPDVAQAIDDAFGSFDAFKEQFTAAATGRFGSGWAWLVVNKEGQLEIT 143

Query: 121 STANQDTPIMEGKKPILGLDVWEHAYYLNYRNVRPNYIKAFFEIINWNKVNELYQAAK 178
           STANQDTPI EGKKPIL LDVWEHAYYLNYRNVRPNYIKAFFEI+NW KV+ELYQAAK
Sbjct: 144 STANQDTPISEGKKPILALDVWEHAYYLNYRNVRPNYIKAFFEIVNWKKVSELYQAAK 201
```

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

Example 2166

A DNA sequence (GBSx2283) was identified in *S.agalactiae* <SEQ ID 6695> which encodes the amino acid sequence <SEQ ID 6696>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3331(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 2167

A DNA sequence (GBSx2284) was identified in *S.agalactiae* <SEQ ID 6697> which encodes the amino acid sequence <SEQ ID 6698>. This protein is predicted to be DNA polymerase III delta subunit. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0511(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 9743> which encodes amino acid sequence <SEQ ID 9744> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6699> which encodes the amino acid sequence <SEQ ID 6700>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.22   Transmembrane  250 - 266 ( 249 - 266)

----- 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>
```

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

```
 Identities = 222/340 (65%), Positives = 282/340 (82%)

Query: 1   MIAIEEIGRITPDNLGLVTVLAGEDLGQYAQMKEKLFQVIGFNKDDLAYSYFDLSEEDYQ 60
           MIAIE+I +++ +NLGL+T++ G+D+GQY+Q+K +L + I F+KDDLAYSYFD+SE  YQ
Sbjct: 1   MIAIEKIEKLSKENLGLITLVTGDDIGQYSQLKSRLMEQIAFDKDDLAYSYFDMSEAAYQ 60

Query: 61  NAELDLESLPFLSDYKVVIFDQFQDITTDKKTYLDEQAMKRFEAYLQNPVDTTRLVICAP 120
           +AE+DL SLPF ++ KVVIFD   DITT+KK++L E+ +K FEAYL+NP++TTRL+I AP
Sbjct: 61  DAEMDLVSLPFFAEQKVVIFDHLLDITTNKKSFLKEKDLKAFEAYLENPLETTRLIIFAP 120

Query: 121 GKLDGKRRLVKLLKRDARVLEANTLKESDLKTYFQKYAHQEGLVFEAGVFDELLIKSNYD 180
           GKLD KRRLVKLLKRDA VLEAN LKE++L+TYFQKY+HQ GL FE+G FD+LL+KSN D
Sbjct: 121 GKLDSKRRLVKLLKRDALVLEANPLKEAELRTYFQKYSHQLGLGFESGAFDQLLLKSNDD 180

Query: 181 FSDTLTNIAFLKSYKTDGHISSNDVREAIPKSLQDNIFDLTQDVLLGRIDLARDLVRDLR 240
           FS  + N+AFLK+YK  G+IS  D+ +AIPKSLQDNIFDLT+ VL G+ID ARDL+ DLR
Sbjct: 181 FSQIMKNMAFLKAYKKTGNISLTDIEQAIPKSLQDNIFDLTRLVLGGKIDAARDLIHDLR 240

Query: 241 LQGEDEIKLIAIMLGQFRMFLQVKILASKGKSESQIVSELSHYIGRKINPYQVKFAVRDS 300
           L GED+IKLIAIMLGQFR+FLQ+ ILA   K+E Q+V  LS  +GR++NPYQVK+A++DS
Sbjct: 241 LSGEDDIKLIAIMLGQFRLFLQLTILARDVKNEQQLVISLSDILGRRVNPYQVKYALKDS 300

Query: 301 RNLPLAFLKEAIRILIETDYAIKRGTYDKDYLFDLALLKI 340
           R L LAFL  A++ LIETDY IK G Y+K YL D+ALLKI
Sbjct: 301 RTLSLAFLTGAVKTLIETDYQIKTGLYEKSYLVDIALLKI 340
```

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

Example 2168

A DNA sequence (GBSx2285) was identified in *S.agalactiae* <SEQ ID 6701> which encodes the amino acid sequence <SEQ ID 6702>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3071(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Example 2169

A DNA sequence (GBSx2286) was identified in *S.agalactiae* <SEQ ID 6703> which encodes the amino acid sequence <SEQ ID 6704>. This protein is predicted to be esterase. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -0.32   Transmembrane  175 - 191 ( 175 - 191)

----- 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:AAB17013 GB:L38252 esterase [Acinetobacter lwoffii]
 Identities = 63/218 (28%), Positives = 107/218 (48%), Gaps = 3/218 (1%)

Query: 105 KVIFYVHGGSYIHQASELQYIFVNKLAKKLDAKVVFPIYPKAPTYNYSDAIPKIKKLYQN 164
           ++IF++HGG++   +       +  LA +   +V+   YP AP + Y +AI  I  +YQ
Sbjct: 73  QLIFHIHGGAFFLGSLNTHRALMTDLAARTQMQVIHVDYPLAPEHPYPEAIDAIFDVYQA 132

Query: 165 TLASVTSPKQIILVGESAGGGLALGLADNLVTEHIKQPKEIILISPWLDIATNNPKIEKV 224
            L     PK II+ G+S G  LAL L   L  +    P  +IL+SP+LD+   +  +
Sbjct: 133 LLVQGIKPKDIIISGDSCGANLALALCLRLKQQPELMPSGLILMSPYLDLTLTSESLRFN 192

Query: 225 QKKDPLLKAWQLQQVAPYWANGKKNFKNPQVSPLYSSQFNKMAPISFFIGTHDIFYPDNQ 284
           QK D LL    LQ    ++       +P+VSPL+    + + P    +G+ +I   D++
Sbjct: 193 QKHDALLSIEALQAGIKHYLTDDIQPGDPRVSPLF-DDLDGLPPTLVQVGSKEILLDDSK 251

Query: 285 LLHQKLAKENIKHHYIVGQKMNHVYPVLP--IPEAETA 320
              +K  + ++K H+ +   M H + +     PEA+ A
Sbjct: 252 RFREKAEQADVKVHFKLYTGMWHNFQMFNAWFPEAKQA 289
```

There is also homology to SEQ ID 3498.

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

Example 2170

A DNA sequence (GBSx2287) was identified in *S.agalactiae* <SEQ ID 6705> which encodes the amino acid sequence <SEQ ID 6706>. This protein is predicted to be purine nucleotide synthesis repressor. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2970(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:CAB16124 GB:Z99124 similar to transcriptional regulator (LacI
          family) [Bacillus subtilis]
 Identities = 111/300 (37%), Positives = 175/300 (58%), Gaps = 4/300 (1%)

Query: 1   MTSISDIAKKAGVAKSTVSRVINHHPHVSDETRQKVMALITELDYIPNQLARDLSRGKTQ 60
           M +I +IA+ A V+ STVSRV+NHHP+VS+E R+ V  ++ ELDY PN+ A DL RGKT
Sbjct: 1   MANIKEIARLANVSVSTVSRVLNHHPYVSEEKRKLVHQVMKELDYTPNRTAIDLIRGKTH 60

Query: 61  KIGVVIPHTRHPYFTQLINGLLDAAKTTDYQLVMMPSDYNQELELSYLKQLKMEAIDALI 120
            +GV++P++ HP F +++NG+  AA   +Y   ++P++YN ++E+ YL+ L+ + ID LI
Sbjct: 61  TVGVILPYSDHPCFDKIVNGITKAAFQHEYATTLLPTNYNPDIEIKYLELLRTKKIDGLI 120

Query: 121 FTSRAISLDIIETYAKYGRIVVCEKLQEYNHLSSAYLDRYSSFLEAFSDMKLRGLEHLVL 180
            TSRA   D I  Y +YG ++ CE   + + +  A+ DR +++ E+F  +K RG E++
```

```
Sbjct: 121 ITSRANHWDSILAYQEYGPVIACEDTGDID-VPCAFNDRKTAYAESFRYLKSRGHENIAF 179

Query: 181 LFSRNNESSATYQSALLAYQEVYGQLSSPYMVVGNVHDFNDG-LNLSYQLVKEVSIDGIL 239
           R  + S +      AY+ V G+L   +M+ G  +D NDG L   +  +       I
Sbjct: 180 TCVREADRSPSTADKAAAYKAVCGRLEDRHMLSG-CNDMNDGELAAEHFYMSGRVPTAIY 238

Query: 240 ATSDEVAAGLIKGYEESRKKCPYIIGQECLLVGQLLKLPTIDHKSYYLGKLAFKQALAEK 299
           A SDEVAAG I  + +       IIG+    + ++L  P++D     LG  AF   L ++
Sbjct: 239 ANSDEVAAG-IHLFAKKNNWDVEIIGEGNTSISRVLGFPSLDLNLEQLGIAAFSLFLQDE 297
```

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

Example 2171

A DNA sequence (GBSx2288) was identified in *S.agalactiae* <SEQ ID 6707> which encodes the amino acid sequence <SEQ ID 6708>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3451(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:AAC21682 GB:U32686 conserved hypothetical protein [Haemophilus influenzae Rd]
 Identities = 79/264 (29%), Positives = 134/264 (49%), Gaps = 16/264 (6%)

Query: 1   MTIKRIFCDMDGTLLNSEGQVSKSNATLIREAA---IPVTLVSARAPMEMKDAVDALQLG 57
           M  K +F D +GTLL S+  +S     +I+      IP   +SAR+P+ +      L+
Sbjct: 1   MMYKAVFSDFNGTLLTSQHTISPRTVVVIKRLTANGIPFVPISARSPLGILPYWKQLETN 60

Query: 58  GVQVAFNGGLIYRIGDNNQVLPIHTQIIKKSTVKQLLRGIRFHFPQVSLSYYDLNNWYCD 117
            V VAF+G LI     N  + PI++  I+   + ++   +  H P + ++YY  N+ +
Sbjct: 61  NVLVAFSGALIL----NQNLEPIYSVQIEPKDILEINTVLAEH-PLLGVNYYTNNDCHAR 115

Query: 118 KID-EGIRYEHSLTQQCPTFIHNEDQFLEGHTNTFKIMMITFDEANMLELEKYLQSLELP 176
            ++ + + YE S+T+     IH  D+     T +   + I  +   ++E+E  L+  + P
Sbjct: 116 DVENKWVIYERSVTK---IEIHPFDEVA---TRSPHKIQIIGEAEEIIEIEVLLKE-KFP 168

Query: 177 EITIQRSGKAYLEITHLLAKKSKGIAYILQKEQLAREETAAFGDGHNDLPMLEMVGYPIV 236
            ++I RS   +LE+ H  A K   + ++     E  AFGD  NDL MLE VG  +
Sbjct: 169 HLSICRSHANFLEVMHKSATKGSAVRFLEDYFGVQTNEVIAFGDNFNDLDMLEHVGLGVA 228

Query: 237 MDNAFDDIKAIAYQLTKSNDEDGV 260
           M NA ++IK  A  +T +N+EDG+
Sbjct: 229 MGNAPNEIKQAANVVTATNNEDGL 252
```

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

Example 2172

A DNA sequence (GBSx2289) was identified in *S.agalactiae* <SEQ ID 6709> which encodes the amino acid sequence <SEQ ID 6710>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2854(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 2173

A DNA sequence (GBSx2290) was identified in *S.agalactiae* <SEQ ID 6711> which encodes the amino acid sequence <SEQ ID 6712>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.51   Transmembrane  392 - 408 ( 376 - 417)
   INTEGRAL    Likelihood = -9.92   Transmembrane  440 - 456 ( 433 - 461)
   INTEGRAL    Likelihood = -6.42   Transmembrane   52 -  68 (  51 -  70)
   INTEGRAL    Likelihood = -6.32   Transmembrane   29 -  45 (   9 -  48)
   INTEGRAL    Likelihood = -6.32   Transmembrane  309 - 325 ( 308 - 328)
   INTEGRAL    Likelihood = -4.46   Transmembrane   12 -  28 (   9 -  29)
   INTEGRAL    Likelihood = -3.29   Transmembrane  463 - 479 ( 462 - 479)
   INTEGRAL    Likelihood = -2.07   Transmembrane  353 - 369 ( 352 - 369)
   INTEGRAL    Likelihood = -1.17   Transmembrane  374 - 390 ( 374 - 390)
   INTEGRAL    Likelihood = -0.85   Transmembrane  247 - 263 ( 247 - 263)
   INTEGRAL    Likelihood = -0.06   Transmembrane  278 - 294 ( 278 - 294)

----- Final Results -----
              bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAC23742 GB:AF052208 competence protein [Streptococcus pneumoniae]
 Identities = 325/705 (46%), Positives = 478/705 (67%), Gaps = 3/705 (0%)

Query: 1   MLQLTKYFPLKPIYLALLVFQIYLLVFSWTMLGCAFLLFSFIFLIYQYDRETIFKTIAIV 60
           MLQ  K F +  IYL+ L+  +Y  +FS + L     +F  + L  Q+  ++  K + I
Sbjct: 1   MLQWIKNFSIPLIYLSFLLLWLYYAIFSASYLALLGFVFLLVCLFIQFPWKSAGKVLIIC 60

Query: 61  IFFLFYFLWQNHNMNVQYQRVPNHISQIKVRIDTISINGDVLSFQADASGNTYQAFYTLK 120
             F F+F++QN   +   Q + + + ++++  DT+ +NGD LSF+  A G  +Q +Y L+
Sbjct: 61  GIFGFWFVFQNWQQSQASQNLADSVERVRILPDTVKVNGDSLSFRGKADGRIFQVYYKLQ 120

Query: 121 NKSEKDYFQNLDNNIMIIADIKLEEAEERRHFNGFDYRQYLKRHGIYRIAKVTKIKQIRL 180
           ++ EK+ FQ L +   I  + KL E E +R+F GF+Y+ YLK  GIY+   + KI+ ++
Sbjct: 121 SEEEKEAFQALTDLHEIGLEGKLSEPEGQRNFGGFNYQAYLKTQGIYQTLNIKKIQSLQK 180

Query: 181 FQHRSFFALMSKWRRSAIVISQT-FPNPMRHYMSGLLFGYLDKTFDDMSDLYSSLGIIHL 239
                 +S  RR A+V  +T FP+PMR+YM+GLL G+LD  F++M++LYSSLGIIHL
Sbjct: 181 IGSWDIGENLSSLRRKAVVWIKTHFPDPMRNYMTGLLLGHLDTDFEEMNELYSSLGIIHL 240

Query: 240 FALSGMQVGFFLGIFRYICLRIGLRLDHVWLLQIPFSLIYAGLTGFSISVVRALIQSLLS 299
           FALSGMQVGFF+  F+ + LR+GL  + +  L  PFSLIYAGLTGFS SV+R+L+Q LL+
Sbjct: 241 FALSGMQVGFFMNGFKKLLLRLGLTQEKLKWLTYPFSLIYAGLTGFSASVIRSLLQKLLA 300

Query: 300 HSGVKKDENFALCLLICLISLPHSLLTTGGVLSFAYAFILTMTSFDHFSSIKKVAIESLT 359
             GVK  +N AL +L+  I +P+   T GGVLS AYAFILTM S +    +K VA ESL
Sbjct: 301 QHGVKGLDNCALTVLVLFIVMPNFFFTAGGVLSCAYAFILTMPSKEG-EGLKAVASESLV 359

Query: 360 VSVGILPILTYYFSGFQPISIILTALLSFAFDIIFLPLLTVIFVLSPIVKLSCINSLFEI 419
           +S+GILPIL++YF+ FQP SI+LT + SF FD+ FLPLL+++FVLS +  +  +N +FE
```

```
Sbjct: 360 ISLGILPILSFYFAEFQPWSILLTFVFSFLFDLTFLPLLSILFVLSFLYPVIQLNFIFEW 419

Query: 420 LEVLLKWTGQLFPRPLIFGKPSLFLLIVMIIILGLLYDYYHSKCFRYCSLLIIFTLFFIT 479
           LE +++   Q+  RPL+FG+P+ +LLI+++I L L+YD  +        L+I  LF +T
Sbjct: 420 LEGIIRLVSQVTSRPLVFGQPNTWLLILLLISLALVYDLRKNIKKLTVLCLLITGLFLLT 479

Query: 480 KNPITNEVAILDVGQGDSILVRDWLGKTILIDTGGRVR-FEQPEEWKQKVNQSNAKRTLI 538
           K+P+ NE+ +LDVGQG+SI +RD  GKTILID GG+   +++ ++W++K+  SNA+R+LI
Sbjct: 480 KHPLENEITMLDVGQGESIFLRDVTGKTILIDVGGKAESYKKIKKWQEKMTTSNAQRSLI 539

Query: 539 PYLKSRGISKIDDLVITHTDTDHMGDMEVISKHFKVARLITSSGSLTNSQYVKHLSKIGV 598
           PYLKSRG++KID L++T+TD +H+GD+  ++K F V  ++ S  SL   ++V  L
Sbjct: 540 PYLKSRGVAKIDQLILTNTDKEHVGDLSEMTKAFHVGEILVSKDSLKQKEFVAELQATQT 599

Query: 599 AVKSIEAGDKLAVMGSYLQVLYPWHKGDGKNNDSIVLYGHLLGKGFLFTGDLEEEGEKQL 658
            V+S+  G+ L + GS L+VL P   GDG ++D++VLYG  L K FLFTG+LEE+GEK L
Sbjct: 600 KVRSMIVGENLPIFGSQLEVLSPRKMGDGGHDDTLVLYGKFLDKQFLFTGNLEEKGEKDL 659

Query: 659 LEAYPNLSVDILKAGHHGSKGSSSLSFLKKLSPSVVLVSAGKNNR 703
           L+ YP+L V++LKA  HG+K SSS +FL+KL P + L+S GK+NR
Sbjct: 660 LKHYPDLKVNVLKASQHGNKKSSSPAFLEKLKPELTLISVGKSNR 704
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6713> which encodes the amino acid sequence <SEQ ID 6714>. Analysis of this protein sequence reveals the following:

```
     Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.19   Transmembrane  394 - 410 ( 380 - 422)
   INTEGRAL    Likelihood = -8.28   Transmembrane   54 -  70 (  52 -  72)
   INTEGRAL    Likelihood = -6.32   Transmembrane  356 - 372 ( 355 - 377)
   INTEGRAL    Likelihood = -4.73   Transmembrane    8 -  24 (   7 -  25)
   INTEGRAL    Likelihood = -4.30   Transmembrane   30 -  46 (  29 -  50)
   INTEGRAL    Likelihood = -3.88   Transmembrane  249 - 265 ( 249 - 267)
   INTEGRAL    Likelihood = -3.40   Transmembrane  467 - 483 ( 465 - 484)
   INTEGRAL    Likelihood = -2.39   Transmembrane  325 - 341 ( 325 - 347)
   INTEGRAL    Likelihood = -0.43   Transmembrane  441 - 457 ( 441 - 458)

----- 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 databases:

```
>GP:AAC23742 GB:AF052208 competence protein [Streptococcus pneumoniae]
 Identities = 311/706 (44%), Positives = 458/706 (64%), Gaps = 10/706 (1%)

Query: 5   WTKLVPLSKIQFAFLILVFFYQIHSPSWLTFL-LSLSLICLLVKRLSKK--EFLGVFAIL 61
           W K   +  I  +FL+L  +Y I S S+L  L     L+CL ++   K   + L +  I
Sbjct: 4   WIKNFSIPLIYLSFLLLWLYYAIFSASYLALLGFVFLLVCLFIQFPWKSAGKVLIICGIF 63

Query: 62  SFCALFLLYQKQQLVQKLEIQPVQITSVALVPDSIRINGDQLAVLGRHGKHSYQLFYRLK 121
            F  +F  +Q+ Q  Q L      +  V ++PD++++NGD L+  G+     +Q++Y+L+
Sbjct: 64  GFWFVFQNWQQSQASQNLADS---VERVRILPDTVKVNGDSLSFRGKADGRIFQVYYKLQ 120

Query: 122 SQAEAQLFKKEHRWLVMHAKVTLEKAEEVRNFKGFNYQTFLTYQGIYRIGKVEQIEQLEV 181
           S+ E + F+       +  +  L + E  RNF GFNYQ +L  QGIY+   +++I+ L+
Sbjct: 121 SEEEKEAFQALTDLHEIGLEGKLSEPEGQRNFGGFNYQAYLKTQGIYQTLNIKKIQSLQK 180

Query: 182 ISPESICDYLSSLRRRAIVHCQQHFPRPMSHYLTGLLFGYLDKSFGEMTDYYSQLGIIHL 241
           I    I + LSSLRR+A+V  + HFP PM +Y+TGLL G+LD  F EM + YS LGIIHL
Sbjct: 181 IGSWDIGENLSSLRRKAVVWIKTHFPDPMRNYMTGLLLGHLDTDFEEMNELYSSLGIIHL 240

Query: 242 FALSGMQVGFFLTCFRRVLLLLAVPLEWIKWIELPFACFYAALTGYSISVIRSLVQSQLR 301
           FALSGMQVGFF+  F+++LL L +  E +KW+  PF+  YA LTG+S SVIRSL+Q  L
Sbjct: 241 FALSGMQVGFFMNGFKKLLLRLGLTQEKLKWLTYPFSLIYAGLTGFSASVIRSLLQKLLA 300
```

```
Query: 302 HLGIKGLDNLACTFLLVFLWDAHFLMTVGGVLTFSYAFLLTVVTVEELSGAKRQLVQVLT 361
           G+KGLDN A T L++F+    +F . T GGVL+ +YAF+LT+ + +E  G K     + L
Sbjct: 301 QHGVKGLDNCALTVLVLFIVMPNFFFTAGGVLSCAYAFILTMPS-KEGEGLKAVASESLV 359

Query: 362 ISLGILPFLLFYFSSFNPMSMVLTGLLSYLFDLFILPLLCLVFCLSPLVTVSICNHLFIL 421
           ISLGILP L FYF+ F P S++LT + S+LFDL  LPLL ++F LS L  V   N +F
Sbjct: 360 ISLGILPILSFYFAEFQPWSILLTFVFSFLFDLTFLPLLSILFVLSFLYPVIQLNFIFEW 419

Query: 422 LEKVIQFLGNTFNSSLVFGSPTSWHLLILVISFAIFYDYRQ-VRQRVITCGLVIALTLLS 480
           LE +I+ +    +  LVFG P +W L++L+IS A+ YD R+ +++  + C L+  L LL+
Sbjct: 420 LEGIIRLVSQVTSRPLVFGQPNTWLLILLLISLALVYDLRKNIKKLTVLCLLITGLFLLT 479

Query: 481 VKYPLTNEVTFIDIGQGDSILVREWTGKNLLIDVGGR-PFFSSKEHWRRGHHVANAQKTL 539
            K+PL NE+T +D+GQG+SI +R+ TGK +LIDVGG+   +  W+      +NAQ++L
Sbjct: 480 -KHPLENEITMLDVGQGESIFLRDVTGKTILIDVGGKAESYKKIKKWQEKMTTSNAQRSL 538

Query: 540 IPYLKSRGIHTIDQLLVTHADTDHMGDIEVVAKAIRIKEILTSQGSLSHPSFVRRLRRLK 599
           IPYLKSRG+  IDQL++T+ D +H+GD+  + KA  + EIL S+ SL     FV  L+  +
Sbjct: 539 IPYLKSRGVAKIDQLILTNTDKEHVGDLSEMTKAFHVGEILVSKDSLKQKEFVAELQATQ 598

Query: 600 CHVRVLAAGDQLPIMGSVLQVLYPWQLGDGKNNDSLVLYGRLLNRTFLFTGDLEKEGENE 659
           VR +  G+ LPI GS L+VL P ++GDG ++D+LVLYG+ L++ FLFTG+LE++GE +
Sbjct: 599 TKVRSMIVGENLPIFGSQLEVLSPRKMGDGGHDDTLVLYGKFLDKQFLFTGNLEEKGEKD 658

Query: 660 IIKRYPQLRVDYLKAGHHGSNTSSSAAFLDHIQPKVAFISAGKNNR 705
           ++K YP L+V+ LKA  HG+  SSS AFL+ ++P++  IS GK+NR
Sbjct: 659 LLKHYPDLKVNVLKASQHGNKKSSSPAFLEKLKPELTLISVGKSNR 704
```

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

```
 Identities = 346/743 (46%), Positives = 491/743 (65%), Gaps = 3/743 (0%)

Query: 5   TKYFPLKPIYLALLVFQIYLLVFSWTMLGCAFLLFSFIFLIYQYDRETIFKTIAIVIFFL 64
           TK  PL  I  A L+   +  + S + L     L     L+ +  ++        AI+ F
Sbjct: 6   TKLVPLSKIQFAFLILVFFYQIHSPSWLTFLLSLSLICLLVKRLSKKEFLGVFAILSFCA 65

Query: 65  FYFLWQNHNMNVQYQRVPNHISQIKVRIDTISINGDVLSFQADASGNTYQAFYTLKNKSE 124
           + L+Q   +  + +  P  I+ + +  D+I INGD L+        ++YQ FY LK+++E
Sbjct: 66  LFLLYQKQQLVQKLEIQPVQITSVALVPDSIRINGDQLAVLGRHGKHSYQLFYRLKSQAE 125

Query: 125 KDYFQNLDNNIMIIADIKLEEAEERRHFNGFDYRQYLKRHGIYRIAKVTKIKQIRLFQHR 184
              F+     +++ A + LE+AEE R+F GF+Y+ +L   GIYRI KV +I+Q+ +
Sbjct: 126 AQLFKKEHRWLVMHAKVTLEKAEEVRNFKGFNYQTFLTYQGIYRIGKVEQIEQLEVISPE 185

Query: 185 SFFALMSKWRRSAIV-ISQTFPNPMRHYMSGLLFGYLDKTFDDMSDLYSSLGIIHLFALS 243
           S    +S  RR AIV   Q FP PM HY++GLLFGYLDK+F +M+D YS LGIIHLFALS
Sbjct: 186 SICDYLSSLRRRAIVHCQQHFPRPMSHYLTGLLFGYLDKSFGEMTDYYSQLGIIHLFALS 245

Query: 244 GMQVGFFLGIFRYICLRIGLRLDHVWLLQIPFSLIYAGLTGFSISVVRALIQSLLSHSGV 303
           GMQVGFFL  FR + L + + L+ +  +++PF+  YA LTG+SISV+R+L+QS L H G+
Sbjct: 246 GMQVGFFLTCFRRVLLLLAVPLEWIKWIELPFACFYAALTGYSISVIRSLVQSQLRHLGI 305

Query: 304 KKDENFALCLLICLISLPHSLLTTGGVLSFAYAFILIMTSFDHFSSIKKVAIESLTVSVG 363
           K  +N A   L+  +   H L+T GGVL+F+YAF+LT+TVVTVEELSGAKRQLVQVLTISLG
Sbjct: 306 KGLDNLACTFLLVFLWDAHFLMTVGGVLTFSYAFLLTVVTVEELSGAKRQLVQVLTISLG 365

Query: 364 ILPILTYYFSGFQPISIILTALLSFAFDIIFLPLLTVIFVLSPIVKLSCINSLFEILEVL 423
           ILP L +YFS F P+S++LT LLS+ FD+  LPLL ++F LSP+V +S  N LF +LE +
Sbjct: 366 ILPFLLFYFSSFNPMSMVLTGLLSYLFDLFILPLLCLVFCLSPLVTVSICNHLFILLEKV 425

Query: 424 LKWTGQLFPRPLIFGKPSLFLLIVMIIILGLLYDYYHSKC-FRYCSLLIIFTLFFITKNP 482
           +++ G  F   L+FG P+ + L++++I   + YDY  +     C L+I  TL  + K P
Sbjct: 426 IQFLGNTFNSSLVFGSPTSWHLLILVISFAIFYDYRQVRQRVITCGLVIALTLLSV-KYP 484

Query: 483 ITNEVAILDVGQGDSILVRDWLGKTILIDTGGRVRFEQPEEWKQKVNQSNAKRTLIPYLK 542
           +TNEV  +D+GQGDSILVR+W GK +LID GGR  F   E W++  + +NA++TLIPYLK
Sbjct: 485 LTNEVTFIDIGQGDSILVREWTGKNLLIDVGGRPFFSSKEHWRRGHHVANAQKTLIPYLK 544
```

```
Query: 543 SRGISKIDDLVITHTDTDHMGDMEVISKHFKVARLITSSGSLTNSQYVKHLSKIGVAVKS 602
           SRGI  ID L++TH DTDHMGD+EV++K  ++  ++TS GSL++  +V+ L ++   V+
Sbjct: 545 SRGIHTIDQLLVTHADTDHMGDIEVVAKAIRIKEILTSQGSLSHPSFVRRLRRLKCHVRV 604

Query: 603 IEAGDKLAVMGSYLQVLYPWHKGDGKNNDSIVLYGHLLGKGFLFTGDLEEEGEKQLLEAY 662
           + AGD+L +MGS LQVLYPW  GDGKNNDS+VLYG LL + FLFTGDLE+EGE ++++ Y
Sbjct: 605 LAAGDQLPIMGSVLQVLYPWQLGDGKNNDSLVLYGRLLNRTFLFTGDLEKEGENEIIKRY 664

Query: 663 PNLSVDILKAGHHGSKGSSSLSFLKKLSPSVVLVSAGKNNRYQHPHQETLQRFQKIKSKI 722
           P L VD LKAGHHGS  SSS +FL  + P V  +SAGKNNRYQHPH+ETL R +  +
Sbjct: 665 PQLRVDYLKAGHHGSNTSSSAAFLDHIQPKVAFISAGKNNRYQHPHRETLARLEDRQITY 724

Query: 723 FRTDQSGTIRLTGWWKWHIQTVR 745
           +RTD  G IRLTG   WH++TVR
Sbjct: 725 YRTDTQGAIRLTGRTSWHLETVR 747
```

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

Example 2174

A DNA sequence (GBSx2291) was identified in *S.agalactiae* <SEQ ID 6715> which encodes the amino acid sequence <SEQ ID 6716>. This protein is predicted to be competence protein (comEA). Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.77   Transmembrane   18 -  34 (  14 -  36)

----- 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>
```

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

```
>GP:AAC23741 GB:AF052208 competence protein [Streptococcus pneumoniae]
 Identities = 96/217 (44%), Positives = 138/217 (63%), Gaps = 4/217 (1%)

Query: 3   EIVLEKIKSHKWETTGIIVGLLLFGILGLNHFG-THHKEDNLNINLEK-KVSTITEKKVP 60
           E ++EKIK +K       +GLL+ G   L     T  KE NL   +       ++EK+V
Sbjct: 2   EAIIEKIKEYKIIVICTGLGLLVGGFFLLKPAPQTPVKETNLQAEVAAVSKDLVSEKEVN 61

Query: 61  MISHVKDKVSNQVTVDVKGAVNHPGVYSLPSQSRVTDAIKRAGGLSNLADSKSVNLAQKL 120
                +   + +TVDVKGAV  PG+Y LP  SR+ DA+++AGGL+  ADSKS+NLAQK+
Sbjct: 62  KEEKEEPLEQDLITVDVKGAVKSPGIYDLPVGSRINDAVQKAGGLTEQADSKSLNLAQKV 121

Query: 121 QDETVIYVAQKGEKITVVEEEKANNIATQGNSKGKINLNKADLSSLQTISGVGAKRAQDI 180
            DE ++YV  KGE+   V ++      A+  + + K+NLNKA L  L+ + G+G KRAQDI
Sbjct: 122 SDEALVYVPTKGEE--AVSQQTGLGTASSISKEKKVNLNKASLEELKQVKGLGGKRAQDI 179

Query: 181 LDYRDSQGGFKTIDDLKNVSGIGEKTLEKLRQDVTID 217
           +D+R++ G FK++D+LK VSGIG KT+EKL+  VT+D
Sbjct: 180 IDHREANGKFKSVDELKKVSGIGGKTIEKLKDYVTVD 216
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6717> which encodes the amino acid sequence <SEQ ID 6718>. Analysis of this protein sequence reveals the following:

```
     Possible site: 36

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.61   Transmembrane   22 -  38 (  16 -  42)

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

```
                bacterial membrane --- Certainty=0.4843(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:AAC23741 GB:AF052208 competence protein [Streptococcus pneumoniae]
 Identities = 82/179 (45%), Positives = 124/179 (68%), Gaps = 4/179 (2%)

Query: 42  NRQSKAAVPALREISPVKQQVSEEKKEIQEDSSILVDLKGAVQKEGVYKLTASSRVRDVI 101
           N Q++ A  + +++   K+   EEK+E  E   I VD+KGAV+  G+Y L   SR+ D +
Sbjct: 42  NLQAEVAAVS-KDLVSEKEVNKEEKEEPLEQDLITVDVKGAVKSPGIYDLPVGSRINDAV 100

Query: 102 ELAGGLTSEADKHAINFAEKLTDEQVVYVPKQGEEISVLPRSLVSGKKETASKDQSKVHI 161
           + AGGLT +AD  ++N A+K++DE +VYVP +GEE   + +    G   + SK++ KV++
Sbjct: 101 QKAGGLTEQADSKSLNLAQKVSDEALVYVPTKGEE--AVSQQTGLGTASSISKEK-KVNL 157

Query: 162 NKASLEELQHIPGIGAKRAQDIIDMRDKLGGFKALEDLRQVSGIGEKTLEKLKDDIFLD 220
           NKASLEEL+ + G+G KRAQDIID R+  G FK++++L++VSGIG KT+EKLKD + +D
Sbjct: 158 NKASLEELKQVKGLGGKRAQDIIDHREANGKFKSVDELKKVSGIGGKTIEKLKDYVTVD 216
```

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

```
 Identities = 81/166 (48%), Positives = 111/166 (66%), Gaps = 10/166 (6%)

Query: 62  ISHVKDKVSNQ---------VTVDVKGAVNHPGVYSLPSQSRVTDAIKRAGGLSNLADSK 112
           IS VK +VS +          + VD+KGAV   GVY L + SRV D I+ AGGL++ AD
Sbjct: 55  ISPVKQQVSEEKKEIQEDSSILVDLKGAVQKEGVYKLTASSRVRDVIELAGGLTSEADKH 114

Query: 113 SVNLAQKLQDETVIYVAQKGEKITVVEEEKANNIA-TQGNSKGKINLNKADLSSLQTISG 171
           ++N A+KL DE V+YV ++GE+I+V+     +    T    + K+++NKA L  LQ I G
Sbjct: 115 AINFAEKLTDEQVVYVPKQGEEISVLPRSLVSGKKETASKDQSKVHINKASLEELQHIPG 174

Query: 172 VGAKRAQDILDYRDSQGGFKTIDDLKNVSGIGEKTLEKLRQDVTID 217
           +GAKRAQDI+D RD  GGFK ++DL+ VSGIGEKTLEKL+ D+ +D
Sbjct: 175 IGAKRAQDIIDMRDKLGGFKALEDLRQVSGIGEKTLEKLKDDIFLD 220
```

A related GBS gene <SEQ ID 8989> and protein <SEQ ID 8990> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:       5.70
GvH: Signal Score (-7.5): -2.58
     Possible site: 38
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -3.77 threshold:  0.0
   INTEGRAL    Likelihood = -3.77   Transmembrane   18 -  34 (  14 -  36)
   PERIPHERAL  Likelihood = 10.40      73
 modified ALOM score:   1.25

*** Reasoning Step: 3

----- 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>
```

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

```
44.3/64.1% over 215aa

                                                                  Streptococcus
pneumoniae
  GP|3211753| competence protein Insert characterized

ORF01930(304 - 951 of 1014)
GP|3211753|gb|AAC23741.1||AF052208(1  -  216  of  216)  competence  protein  (Streptococcus
pneumoniae)
```

```
%Match = 25.0
%Identity = 44.2  %Similarity = 64.1
Matches = 96  Mismatches = 75  Conservative Sub.s = 43

90        120       150       180       210       240       270       300
DDGKKLNPLTYIYRLPLAIIAIVLLVLTLIFSYLASFVWDPQKHLK*GLHGNYLLFSK*FFWFLIGKSL*LRISKWRNIF

330       360       390       417       447       474       504       534
MFEIVLEKIKSHKWETTGIIVGLLLFGILGLNHFG-THHKEDNLNINLEK-KVSTITEKKVPMISHVKDKVSNQVTVDVK
  | ::|||| :|       :|||: | : |     |  || ||   :       ::||:|      :    : :|||||
 MEAIIEKIKEYKIIVICTGLGLLVGGFFLLKPAPQTPVKETNLQAEVAAVSKDLVSEKEVNKEEKEEPLEQDLITVDVK
          10        20        30        40        50        60        70

564       594       624       654       684       714       744       774
GAVNHPGVYSLPSQSRVTDAIKRAGGLSNLADSKSVNLAQKLQDETVIYVAQKGEKITVVXEEKANNIATQGNSKGKINL
|||  ||:| ||  ||: ||:::|||||:  |||||:|||||: || ::||  |||:  | ::     |:  : : |:||
GAVKSPGIYDLPVGSRINDAVQKAGGLTEQADSKSLNLAQKVSDEALVYVPTKGEE--AVSQQTGLGTASSISKEKKVNL
          90       100       110       120       130       140       150

804       834       864       894       924       954       984       1014
NKADLSSLQTISGVGAKRAQDILDYRDSQGGFKTIDDLKNVSGIGEKTLEKLRQDVTID*VFSSKTYLFSIVGLPNLLTS
||| |  |: : |:| ||||||:|:|:: | ||::|:|| ||||| ||:|||:  ||:|
NKASLEELKQVKGLGGKRAQDIIDHREANGKFKSVDELKKVSGIGGKTIEKLKDYVTVD
          170       180       190       200       210
```

SEQ ID 8990 (GBS129) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 4; MW 43.8kDa).

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

Example 2175

A DNA sequence (GBSx2292) was identified in *S.agalactiae* <SEQ ID 6719> which encodes the amino acid sequence <SEQ ID 6720>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-14.01   Transmembrane  215 - 231 ( 208 - 240)

----- 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 homology with the following sequences in the GENPEPT database.

```
>GP:CAB12793 GB:Z99109 similar to 1-acylglycerol-3-phosphate
          O-acyltransferase [Bacillus subtilis]
 Identities = 66/200 (33%), Positives = 111/200 (55%), Gaps = 10/200 (5%)

Query: 3   YTYLRTLVMFLIWVANGNAHYHNEDKMLKDDENYILVAPHRTFWDPVYMAFAARPKQFIF 62
           Y +    +  ++ +  G   Y+ E+  L  D  +++   H  + D + +     P Q  +
Sbjct: 2   YKFCANALKVILSLRGGVKVYNKEN--LPADSGFVIACTHSGWVDVITLGVGILPYQIHY 59

Query: 63  MAKKELFTNRLFGWWIKMCGAFPIDREKPGQDAIRYPVKMLKNSNRSLVMFPSGSRHSKD 122
           MAKKELF N+  G ++K   AFP+DRE PG  +I+ P+K+LK     + +FPSG+R S+D
Sbjct: 60  MAKKELFQNKWIGSFLKKIHAFPVDRENPGPSSIKTPIKLLK-EGEIVGIFPSGTRTSED 118

Query: 123 V--KGGVAVIAKMAKVRIMPAAYRGPMVFKNLLKGHRVDMNFGNPIDVSDIKRMDA-EGI 179
           V  K G   IA+M K  ++PAAY+GP   K L K  ++ +  G P+  +D    + + E +
Sbjct: 119 VPLKRGAVTIAQMGKAPLVPAAYQGPSSGKELFKKGKMKLIIGEPLHQADFAHLPSKERL 178

Query: 180 A----EVSRRIQEEFDRLDR 195
           A     +++RI+E  ++LD+
Sbjct: 179 AAMTEALNQRIKELENKLDQ 198
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6721> which encodes the amino acid sequence <SEQ ID 6722>. Analysis of this protein sequence reveals the following:

```
      Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.83   Transmembrane  241 - 257 ( 234 - 266)
   INTEGRAL    Likelihood = -4.41   Transmembrane   27 -  43 (  26 -  44)

----- Final Results -----
               bacterial membrane --- Certainty=0.5734(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB12793 GB:Z99109 similar to 1-acylglycerol-3-phosphate
          O-acyltransferase [Bacillus subtilis]
 Identities = 59/198 (29%), Positives = 104/198 (51%), Gaps = 6/198 (3%)

Query: 29  YAYLRGLVVFLLWVVNGNAHYHHEEKMLDASENYILVAPHRTFWDPVYMAFAARPKQFIF 88
           Y +    + +L +  G   Y+ E   L A   +++   H  + D + +     P Q  +
Sbjct: 2   YKFCANALKVILSLRGGVKVYNKEN--LPADSGFVIACTHSGWVDVITLGVGILPYQIHY 59

Query: 89  MAKKELFANRLFAWWIKMCGAFPIDRDKPSPDAIRYPVNMLKKSNRSLLMFPSGSRHSQE 148
           MAKKELF N+    ++K   AFP+DR+ P P +I+ P+ +LK+    + +FPSG+R S++
Sbjct: 60  MAKKELFQNKWIGSFLKKIHAFPVDRENPGPSSIKTPIKLLKE-GEIVGIFPSGTRTSED 118

Query: 149 V--KGGVAVIAKLAKVKIMPAAYQGPMSVKGLLAGERVDMTFGNPIDVSDIKRM-NDEGI 205
           V  K G   IA++ K  ++PAAYQGP S K L    ++ +  G P+  +D   + + E +
Sbjct: 119 VPLKRGAVTIAQMGKAPLVPAAYQGPSSGKELFKKGKMKLIIGEPLHQADFAHLPSKERL 178

Query: 206 AEVANRIQAEFDRIDDEL 223
           A +   +      ++++L
Sbjct: 179 AAMTEALNQRIKELENKL 196
```

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

```
 Identities = 186/244 (76%), Positives = 212/244 (86%)

Query: 1   MFYTYLRTLVMFLIWVANGNAHYHNEDKMLKDDENYILVAPHRTFWDPVYMAFAARPKQF 60
           +FY YLR LV+FL+WV NGNAHYH+E+KML   ENYILVAPHRTFWDPVYMAFAARPKQF
Sbjct: 27  VFYAYLRGLVVFLLWVVNGNAHYHHEEKMLDASENYILVAPHRTFWDPVYMAFAARPKQF 86

Query: 61  IFMAKKELFTNRLFGWWIKMCGAFPIDREKPGQDAIRYPVKMLKNSNRSLVMFPSGSRHS 120
           IFMAKKELF NRLF WWIKMCGAFPIDR+KP  DAIRYPV MLK SNRSL+MFPSGSRHS
Sbjct: 87  IFMAKKELFANRLFAWWIKMCGAFPIDRDKPSPDAIRYPVNMLKKSNRSLLMFPSGSRHS 146

Query: 121 KDVKGGVAVIAKMAKVRIMPAAYRGPMVFKNLLKGHRVDMNFGNPIDVSDIKRMDAEGIA 180
           ++VKGGVAVIAK+AKV+IMPAAY+GPM  K LL G RVDM FGNPIDVSDIKRM+ EGIA
Sbjct: 147 QEVKGGVAVIAKLAKVKIMPAAYQGPMSVKGLLAGERVDMTFGNPIDVSDIKRMNDEGIA 206

Query: 181 EVSRRIQEEFDRLDRENETYDDGKKLNPLTYIYRLPLAIIAIVLLVLTLIFSYLASFVWD 240
           EV+ RIQ EFDR+D E   +  GK  NPLTY+YRLPL ++ +V+L+LT++FSY+ASFVW+
Sbjct: 207 EVANRIQAEFDRIDDELAPFQPGKARNPLTYLYRLPLGLVLVVVLLLTMLFSYIASFVWN 266

Query: 241 PQKH 244
           P KH
Sbjct: 267 PDKH 270
```

SEQ ID 6720 (GBS171) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 36 (lane 2; MW 25kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 3; MW 49.8kDa).

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

Example 2176

A DNA sequence (GBSx2293) was identified in *S.agalactiae* <SEQ ID 6723> which encodes the amino acid sequence <SEQ ID 6724>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3268(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:CAB11810 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 113/244 (46%), Positives = 173/244 (70%), Gaps = 2/244 (0%)

Query: 6   LKENERIDQLFSTDVKIIQNKEVFSYSIDSVLLSRFPKLP-SRGLIVDLCSGNGAVGLFA 64
           L ++ER+D L + D+KIIQ+  VF++S+D+VLLS+F  +P  +G IVDLC+GNG V L
Sbjct: 4   LHDDERLDYLLAEDMKIIQSPTVFAFSLDAVLLSKFAYVPIQKGKIVDLCTGNGIVPLLL 63

Query: 65  STKTNATIIEIELQESLADMAKRSIKLNKLEKQVTMINDDLKNLLDHVQRSNVDLMLCNP 124
           ST++ A I+ +E+QE L DMA RS++ NKL+ Q+ +I+DDLKN+ + +  +  D++ CNP
Sbjct: 64  STRSKADILGVEIQERLHDMAVRSVEYNKLDDQIQIIHDDLKNMPEKLGHNRYDVVTCNP 123

Query: 125 PYFKASETSKKNLSPHYLLARHEITTNLREICQIAQHALKTKGRIAMVHRPDRFLEIIDT 184
           PYFK  + +++N++ H  +ARHEI   L ++  ++   LK  G+ A+VHRP R LEI +
Sbjct: 124 PYFKTPKQTEQNMNEHLRIARHEIHCTLEDVISVSSKLLKQGGKAALVHRPGRLLEIFEL 183

Query: 185 MRQFNLAPKRIQFVYPKLGKDANMLLIEAIKDGSTEGMKILPPLVVHQDNGDYTETIFDI 244
           M+ + + PKR+QFVYPK GK+AN +L+E IK G  + +KILPPL V+ +  +YT+ I  I
Sbjct: 184 MKAYQIEPKRVQFVYPKQGKEANTILVEGIKGGRPD-LKILPPLFVYDEQNEYTKEIRTI 242

Query: 245 YFGE 248
            +G+
Sbjct: 243 LYGD 246
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6725> which encodes the amino acid sequence <SEQ ID 6726>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2183(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 = 200/257 (77%), Positives = 228/257 (87%), Gaps = 3/257 (1%)

Query: 1   MIDTILKENERIDQLFSTDVKIIQNKEVFSYSIDSVLLSRFPKLPSRGLIVDLCSGNGAV 60
           MI  ILKE ERIDQLFS+DV IIQNK+VFSYSIDSVLLSRFPK+PS+GLIVDLCSGNGAV
Sbjct: 1   MIKAILKEGERIDQLFSSDVGIIQNKDVFSYSIDSVLLSRFPKMPSKGLIVDLCSGNGAV 60

Query: 61  GLFASTKTNATIIEIELQESLADMAKRSIKLNKLEKQVTMINDDLKNLLDHVQRSNVDLM 120
           GLFAST+T A I+E+ELQE LADM +RSI+LN+LE QVTMI DDLKNLL+HV RS VDLM
Sbjct: 61  GLFASTRTKAAIVEVELQERLADMGQRSIQLNQLEDQVTMICDDLKNLLNHVPRSGVDLM 120

Query: 121 LCNPPYFKASETSKKNLSPHYLLARHEITTNLREICQIAQHALKTKGRIAMVHRPDRFLE 180
```

```
            LCNPPYFK+ E+SKKN+S HYLLARHE+TTNL EICQ+A+HALK+ GR+AMVHRPDRFLE
Sbjct: 121 LCNPPYFKSHESSKKNVSEHYLLARHEVTTNLEEICQVARHALKSNGRLAMVHRPDRFLE 180

Query: 181 IIDTMRQFNLAPKRIQFVYPKLGKDANMLLIEAIKDGSTEGMKILPPLVVHQDNGDYTET 240
           IID++R   LAPKR+QFVYPKLGK ANMLLIEAIKDGS EGM ILPPLVVH++NG+YT+
Sbjct: 181 IIDSLRANGLAPKRVQFVYPKLGKSANMLLIEAIKDGSIEGMTILPPLVVHKENGEYTDH 240

Query: 241 IFDIYFGENGK---SHD 254
           IF+IYFG   K   +HD
Sbjct: 241 IFEIYFGAASKGKPNHD 257
```

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

Example 2177

A DNA sequence (GBSx2294) was identified in *S.agalactiae* <SEQ ID 6727> which encodes the amino acid sequence <SEQ ID 6728>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1512(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:CAB11811 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 40/82 (48%), Positives = 63/82 (76%)

Query: 7  YMYVLECSDGTLYTGYTTDVKRRLNTHNTGKGAKYTRARLPVKLLYSEAFNSKQEAMRAE 66
          + YV++C D + Y GYT D+ +R+ THN GKGAKYT+ R PV+L+++E+F++K+EAM+AE
Sbjct: 7  FFYVVKCKDNSWYAGYTNDLHKRVKTHNDGKGAKYTKVRRPVELIFAESFSTKREAMQAE 66

Query: 67 ALFKQKTRQAKLTYIKQHKNEQ 88
            FK+ TR+ K  YI++ +N +
Sbjct: 67 YYFKKLTRKKKELYIEEKRNSK 88
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6729> which encodes the amino acid sequence <SEQ ID 6730>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1838(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 = 60/84 (71%), Positives = 67/84 (79%), Gaps = 1/84 (1%)

Query: 6  AYMYVLECSDGTLYTGYTTDVKRRLNTHNTGKGAKYTRARLPVKLLYSEAFNSKQEAMRA 65
          AYMYVLEC D TLYTGYTTD+K+RL THN GKGAKYTR RLPV LLY E F+SK+ AM A
Sbjct: 6  AYMYVLECVDKTLYTGYTTDLKKRLATHNAGKGAKYTRYRLPVSLLYYEVFDSKEAAMSA 65

Query: 66 EALF-KQKTRQAKLTYIKQHKNEQ 88
          EALF K+KTR  KL YI  H+ E+
Sbjct: 66 EALFKKRKTRSQKLAYIATHQKEK 89
```

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

Example 2178

A DNA sequence (GBSx2295) was identified in *S.agalactiae* <SEQ ID 6731> which encodes the amino acid sequence <SEQ ID 6732>. This protein is predicted to be autoaggregation-mediating protein (deaD). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2287(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:AAD20136 GB:AF091502 autoaggregation-mediating protein
          [Lactobacillus reuteri]
 Identities = 289/504 (57%), Positives = 366/504 (72%), Gaps = 18/504 (3%)

Query: 1   MKFTELNLSQDILSAVEKAGFVEPSPIQEMTIPLALEGKDVIGQAQTGTGKTAAFGLPTL 60
           MKF+EL LS  +L A++++G+ E +PIQE TIP+ LEGKDVIGQAQTGTGKTAAFGLP +
Sbjct: 1   MKFSELGLSDSLLKAIKRSGYEEATPIQEQTIPMVLEGKDVIGQAQTGTGKTAAFGLPII 60

Query: 61  NKIHTEDNTIQALIIAPTRELAVQSQEELFRFGRDKGVKVRSVYGGSSIEKQIKALRSGA 120
             + TE+  IQA+II+PTRELA+Q+QEEL+R G+DK V+V+ VYGG+ I +QIK+L+
Sbjct: 61  ENVDTENPNIQAIIISPTRELAIQTQEELYRLGKDKHVRVQVVYGGADIRRQIKSLKQHP 120

Query: 121 HVVVGTPGRLLDLIKRKALKLNHIETLILDEADEMLNMGFLEDIEAIISRVPETRQTLLF 180
            ++VGTPGRL D I R  +KL+HI+TL+LDEADEMLNMGFLEDIE+II   P+ RQTLLF
Sbjct: 121 QILVGTPGRLRDHINRHTVKLDHIKTLVLDEADEMLNMGFLEDIESIIKETPDDRQTLLF 180

Query: 181 SATMPDPIKRIGVKFMKDPEHVKIKATELTNVNVDQYYVRVKENEKFDTMTRLMDVDQPE 240
           SATMP  IKRIGV+FM DPE V+IKA ELT   VDQYYVR ++ EKFD MTRL+DV  P+
Sbjct: 181 SATMPPEIKRIGVQFMSDPETVRIKAKELTTDLVDQYYVRARDYEKFDIMTRLIDVQDPD 240

Query: 241 LSIVFGRTKRRVDELTRGLKLRGFRAEGIHGDLDQNKRLRVIRDFKNDHIDILVATDVAA 300
           L+IVFGRTKRRVDEL++GL  RG+ A GIHGDL Q+KR +++  FKN+ +DILVATDVAA
Sbjct: 241 LTIVFGRTKRRVDELSKGLIARGYNAAGIHGDLTQDKRSKIMWKFKNNELDILVATDVAA 300

Query: 301 RGLDISGVTHVYNYDIPQDPESYVHRIGRTGRAGKSGQSITFVSPNEMGYLTIIENLTKK 360
           RGLDISGVTHVYNYDIP DP+SYVHRIGRTGRAG  G S+TFV+PNEM YL  IE LT+
Sbjct: 301 RGLDISGVTHVYNYDIPSDPDSYVHRIGRTGRAGHHGVSLTFVTPNEMDYLHEIEKLTRV 360

Query: 361 RMTGMKPATASEAFQAKKKVALKRIARDFED-QELVSK--FDKFKADALELATQYTPEEL 417
           RM  +KP TA EAF+         ++A  F D  EL+++   D+++  A +L   +   +L
Sbjct: 361 RMLPLKPPTAEEAFKG-------QVASAFNDIDELIAQDSTDRYEEAAEKLLETHNATDL 413

Query: 418 ALYVLSLTVQDPESLPEVEITREKPLPFKPSGGGFKGKGGRGNGRGGD--RRRNDRGDRR 475
             +L+   ++  S   V+IT E+PLP +         G R N  GG+  RR+N R  +
Sbjct: 414 VAALLNNMTKEAASEVPVKITPERPLPRRNKRN--NRNGNRNNSHGGNHYRRKNFRRHQH 471

Query: 476 GNRDRDDRG----SRCDFKRRDDK 495
           G+   D+ G     SR  F  R  K
Sbjct: 472 GSHRNDNHGKSHSSRHSFNIRHRK 495
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6733> which encodes the amino acid sequence <SEQ ID 6734>. Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1108(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 = 430/545 (78%), Positives = 463/545 (84%), Gaps = 24/545 (4%)

Query: 1   MKFTELNLSQDILSAVEKAGFVEPSPIQEMTIPLALEGKDVIGQAQTGTGKTAAFGLPTL 60
           +KFTE NLSQDI SAV  AGF + SPIQEMTIPLALEGKDVIGQAQTGTGKTAAFGLPTL
Sbjct: 1   LKFTEFNLSQDIQSAVVTAGFEKASPIQEMTIPLALEGKDVIGQAQTGTGKTAAFGLPTL 60

Query: 61  NKIHTEDNTIQALIIAPTRELAVQSQEELFRFGRDKGVKVRSVYGGSSIEKQIKALRSGA 120
           NKI T +N IQAL+IAPTRELAVQSQEELFRFGR+KGVKVRSVYGGSSIEKQIKAL+SGA
Sbjct: 61  NKIRTNENIIQALVIAPTRELAVQSQEELFRFGREKGVKVRSVYGGSSIEKQIKALKSGA 120

Query: 121 HVVVGTPGRLLDLIKRKALKLNHIETLILDEADEMLNMGFLEDIEAIISRVPETRQTLLF 180
           H+VVGTPGRLLDLIKRKAL L+H+ETLILDEADEMLNMGFLEDIEAIISRVP  RQTLLF
Sbjct: 121 HIVVGTPGRLLDLIKRKALILDHVETLILDEADEMLNMGFLEDIEAIISRVPADRQTLLF 180

Query: 181 SATMPDPIKRIGVKFMKDPEHVKIKATELTNVNVDQYYVRVKENEKFDTMTRLMDVDQPE 240
           SATMP PIK+IGVKFMKDPEHV+IK  ELTNVNVDQYYVRVKE EKFDTMTRLMDV+QPE
Sbjct: 181 SATMPAPIKQIGVKFMKDPEHVQIKNKELTNVNVDQYYVRVKEQEKFDTMTRLMDVNQPE 240

Query: 241 LSIVFGRTKRRVDELTRGLKLRGFRAEGIHGDLDQNKRLRVIRDFKNDHIDILVATDVAA 300
           LSIVFGRTKRRVDE+TRGLKLRGFRAEGIHGDLDQNKRLRVIRDFKND IDILVATDVAA
Sbjct: 241 LSIVFGRTKRRVDEITRGLKLRGFRAEGIHGDLDQNKRLRVIRDFKNDQIDILVATDVAA 300

Query: 301 RGLDISGVTHVYNYDIPQDPESYVHRIGRTGRAGKSGQSITFVSPNEMGYLTIIENLTKK 360
           RGLDISGVTHVYNYDI QDPESYVHRIGRTGRAGKSG+SITFVSPNEMGYL++IENLTKK
Sbjct: 301 RGLDISGVTHVYNYDITQDPESYVHRIGRTGRAGKSGESITFVSPNEMGYLSMIENLTKK 360

Query: 361 RMTGMKPATASEAFQAKKKVALKRIARDFEDQELVSKFDKFKADALELATQYTPEELALY 420
           +M  ++PATA EAFQAKKKVALK+I RDF D+ + S FDKFK DA++LA ++TPEELALY
Sbjct: 361 QMKPLRPATAEEAFQAKKKVALKKIERDFADETIRSNFDKFKGDAVQLAAEFTPEELALY 420

Query: 421 VLSLTVQDPESLPEVEITREKPLPFKPSGGGF---KGKGGRG--NGRGGDRRRNDRGDR- 474
           +LSLTVQDP+SLPEVEI REKPLPFK  GGG     GKGGRG  N   GDRR   RGDR
Sbjct: 421 ILSLTVQDPDSLPEVEIAREKPLPFKYVGGGHGNKNGKGGRGRDNRNRGDRRGGYRGDRN 480

Query: 475 ------------RGNRDRDDRGSRCDFKRRDDKFKKDNRRQENKKPHKNTSSEKQTGFVI 522
                       R  RD  D     DFKR+  +  KD   +E K       SS K TGFVI
Sbjct: 481 RDERDGDRRRQKRDKRDGHDGSGNRDFKRKSKRNSKDFFNKEKK------SSAKNTGFVI 534

Query: 523 RNKGD 527
           R+KG+
Sbjct: 535 RHKGE 539
```

A related GBS gene <SEQ ID 8991> and protein <SEQ ID 8992> were also identified. Analysis of this protein sequence reveals the following:

```
RGD motif 471-473
```

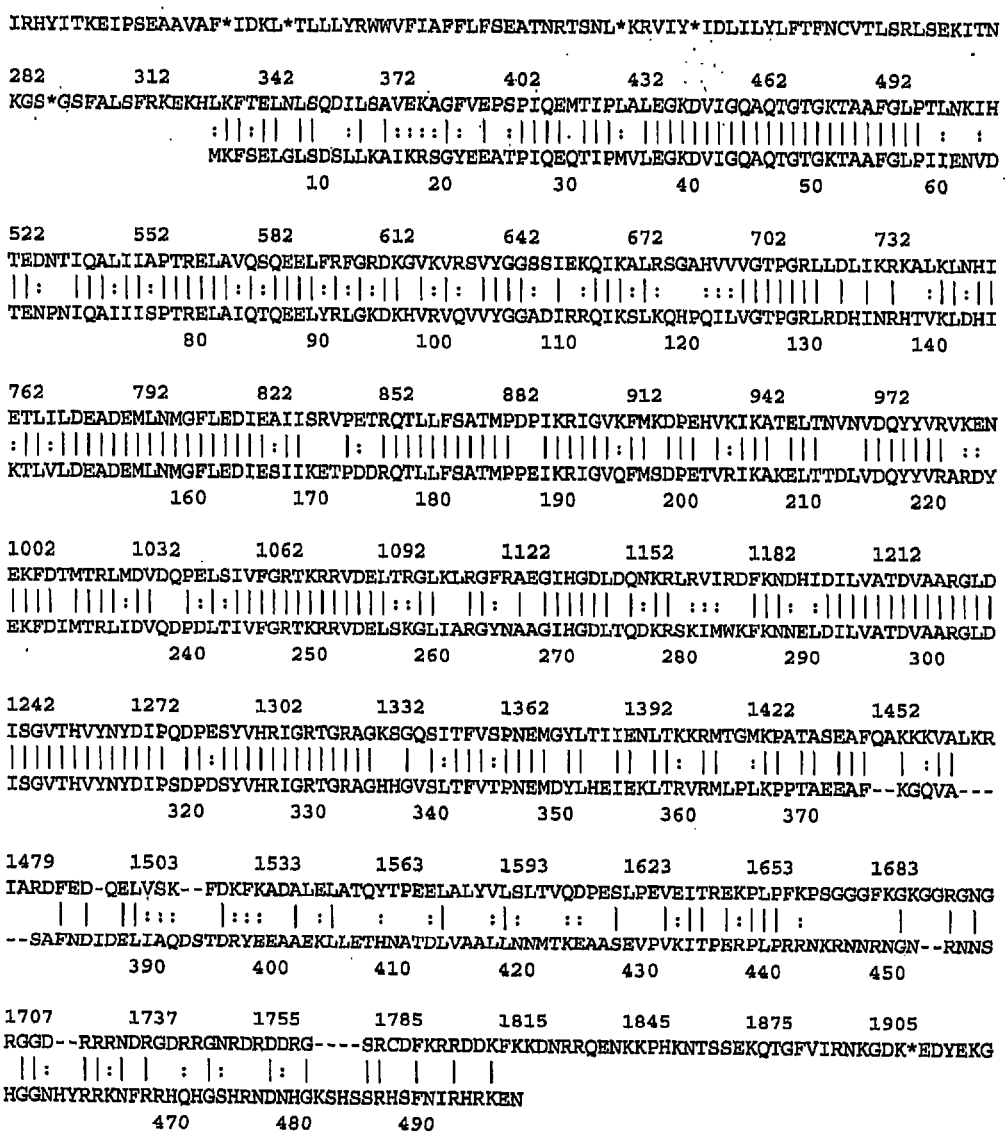
The protein has homology with the following sequences in the databases:

```
58.9/74.7% over 494aa

                                                                    Lactobacillus reuteri
  GP|4409804| autoaggregation-mediating protein Insert characterized

ORF01926(301 - 1785 of 2184)
GP|4409804|gb|AAD20136.1||AF091502(1  -  495  of  497)  autoaggregation-mediating  protein
(Lactobacillus reuteri)
%Match = 37.3
%Identity = 58.8  %Similarity = 74.6
Matches = 290  Mismatches = 118  Conservative Sub.s = 78

42        72        102       132       162       192       222       252
```



There is also homology to SEQ ID 4454.

SEQ ID 8992 (GBS307) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 7; MW 62kDa). 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 2; MW 86.7kDa).

The GBS307-GST fusion product was purified (Figure 208, lane 9; Figure 225, lane 10-11) and used to immunise mice. The resulting antiserum was used for FACS (Figure 272), 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 2179

A DNA sequence (GBSx2296) was identified in *S.agalactiae* <SEQ ID 6735> which encodes the amino acid sequence <SEQ ID 6736>. This protein is predicted to be outer membrane protein (yaeC). Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> 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:CAB73036 GB:AL139076 putative periplasmic protein [Campylobacter
          jejuni]
 Identities = 89/237 (37%), Positives = 132/237 (55%), Gaps = 3/237 (1%)

Query: 40  ITVATYSKPTSTFLDLVKDNVKEKGYTLKVVMVSDYIQANIALENKEHDANLLQHEFFMS 99
           IT+      P  + L+L+KD+ K KGY LK+V  SDYI  N ALE KE DANL QH+ F+
Sbjct: 23  ITIGATPNPFGSLLELMKDDFKNKGYELKIVEFSDYILPNRALEEKELDANLYQHKPFLE 82

Query: 100 IFNKENDGHLVSITPIYHSLAGFYGQHLKNIAELKDGAKVAIPSDPANMTRALLLLQEKK 159
            +N +   +L++ TP+  +  G Y + +KN+  LK+GA+VAIP+D  N +RAL LL++ K
Sbjct: 83  EYNLKKGSNLIATTPVLIAPVGVYSKKIKNLENLKEGARVAIPNDATNESRALELLEKAK 142

Query: 160 LITLKNTSKKTKAIEDIITNPKKLRIEPVALLNLNQAYFEYDLVFNFPGYVTKINLVPKR 219
           LI L   + KT    DI  NPKKL+   +    L +A  + D+       +     L P +
Sbjct: 143 LIELNKNTLKTPL--DINKNPKKLKFIELKAAQLPRALDDVDIAIINSNFALGAGLNPSK 200

Query: 220 DRLLYEKKPDIRFAGALVAREDNKNSDKIKVLKEVLTSKEIRHYITKEIPSEAAVAF 276
           D +  E K +  +   +V R + KNS+K KV+ E+L S + +  I +         AF
Sbjct: 201 DTIFREDK-NSPYVNYVVVRSEGKNSEKTKVIDEILRSDKFKAIINEHYKDILIPAF 256
```

SEQ ID 6736 (GBS126) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 7; MW 32kDa).

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

Example 2180

A DNA sequence (GBSx2297) was identified in *S.agalactiae* <SEQ ID 6737> which encodes the amino acid sequence <SEQ ID 6738>. This protein is predicted to be probable permease of ABC transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.99   Transmembrane  190 - 206 ( 187 - 215)
   INTEGRAL    Likelihood = -8.44   Transmembrane   25 -  41 (  16 -  45)
   INTEGRAL    Likelihood = -6.48   Transmembrane   69 -  85 (  68 -  90)
   INTEGRAL    Likelihood = -3.77   Transmembrane   90 - 106 (  88 - 109)
   INTEGRAL    Likelihood = -1.44   Transmembrane  145 - 161 ( 145 - 161)

----- Final Results -----
               bacterial membrane --- Certainty=0.5798(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:AAG08889 GB:AE004963 probable permease of ABC transporter
          [Pseudomonas aeruginosa]
 Identities = 80/206 (38%), Positives = 127/206 (60%), Gaps = 4/206 (1%)

Query: 15  SFWETNLMLGLTLILCFLIAFPTGILLFSLRKSYLIKHSLAYQLLNLFLGTLRSVPFLIF 74
           +FW    MLG +L+   ++  P G+LLF      + +    Y LL+L +  LRS+PF+I
Sbjct: 24  TFW----MLGGSLLFTVVLGLPLGVLLFLTGPRQMFEQKAVYTLLSLVVNILRSLPFIIL 79
```

```
Query: 75  IFILIPLNRLIFGTSFGTIAAILPLTLVSVSLYARYVEQALLNIPQVVVDRALSLGANKR 134
           + ++IPL  LI GTS G   AI PL + +   +AR VE AL  + + +++   ++GA+ R
Sbjct: 80  LIVMIPLTVLITGTSLGVAGAIPPLVVGATPFFARLVETALREVDKGIIEATQAMGASTR 139

Query: 135 QIIYYFLIPSIKIDLVLSFTATAISILGYSTIMGVIGAGGLGEYAYRFGYQEYDYPVMYL 194
           QII+  L+P  +  ++ + T TAI+++ Y+ + GV+GAGGLG+ A RFGYQ +   VM +
Sbjct: 140 QIIWNALLPEARPGIIAAITVTAITLVSYTAMAGVVGAGGLGDLAIRFGYQRFQTDVMVV 199

Query: 195 IVVLFIIYVFILQSLGYFIANRYSRK 220
           VV+ +I V ILQ++G  +   +SRK
Sbjct: 200 TVVMLLILVQILQTVGDKLVVHFSRK 225
```

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

Example 2181

A DNA sequence (GBSx2298) was identified in *S.agalactiae* <SEQ ID 6739> which encodes the amino acid sequence <SEQ ID 6740>. This protein is predicted to be ABC transporter, ATP-binding protein (oppF). Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5454(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 9333> which encodes amino acid sequence <SEQ ID 9334> was also identified.

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

```
>GP:AAC22280 GB:U32744 ABC transporter, ATP-binding protein
          [Haemophilus influenzae Rd]
 Identities = 62/174 (35%), Positives = 104/174 (59%), Gaps = 2/174 (1%)

Query: 1   MKMINGLIPYDKGNIYYQGKEVKSFSDNKLRQMRKDIAYIFQNHNLLAGESVYYHLALVY 60
           ++ +N L     G++   G E+   SD +L   R+ I  IFQ+ NLL+  +V+ ++AL
Sbjct: 48  IRCVNLLEKPTSGSVIVDGVELTKLSDRELVLARRQIGMIFQHFNLLSSRTVFENVALPL 107

Query: 61  KLNHQKVN--HDAINDILDFLGLMDLKQVKCHSLSGGQQQKVAIAMAVLQKPKLILCDEI 118
           +L  +      + I  +LD +GL + +     +LSGGQ+Q+VAIA A+    PK++LCDE
Sbjct: 108 ELESESKAKIQEKITALLDLVGLSEKRDAYPSNLSGGQKQRVAIARALASDPKVLLCDEA 167

Query: 119 SSALDTNSEKEIFNLLSDLREKYGISILMIAHHLSLLKQYCDRVMILDHQTIVD 172
           +SALD  + + I  LL ++    GI+IL+I H + ++KQ CD+V ++D   +V+
Sbjct: 168 TSALDPATTQSILKLLKEINRTLGITILLITHEMEVVKQICDQVAVIDQGRLVE 221
```

There is also homology to SEQ ID 76.

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

Example 2182

A DNA sequence (GBSx2299) was identified in *S.agalactiae* <SEQ ID 6741> which encodes the amino acid sequence <SEQ ID 6742>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
```

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

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

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 2183

A DNA sequence (GBSx2300) was identified in *S.agalactiae* <SEQ ID 6743> which encodes the amino acid sequence <SEQ ID 6744>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0904(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 9741> which encodes amino acid sequence <SEQ ID 9742> was also identified.

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

```
>GP:AAB87515 GB:AF034138 unknown [Bacillus subtilis]
 Identities = 74/125 (59%), Positives = 92/125 (73%)

Query: 5   MGIFSGLMGNASQMDTDKVENQLSDILISDEQVDLAYTLIRDLIVFTNYRLILVDKQGVT 64
           MG   GL+GNAS + T  V+ +L+ IL+  E+V+ A+ L+RDLIVFT+ RLILVDKQG+T
Sbjct: 1   MGFIDGLLGNASTLSTAAVQEELAHILLEGEKVEAAFKLVRDLIVFTDKRLILVDKQGIT 60

Query: 65  GKKVSYNSIPYASISRFTVETSGHFDLDAELKIWISSAIEPAEVLQFKNDRNIVSIQKAL 124
           GKK  + SIPY SISRF+VET+G FDLD+ELKIWIS A  PA   QFK D +I  IQK L
Sbjct: 61  GKKTEFQSIPYKSISRFSVETAGRFDLDSELKIWISGAELPAVSKQFKKDESIYDIQKVL 120

Query: 125 ATAVL 129
           A   +
Sbjct: 121 AAVCM 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 2184

A DNA sequence (GBSx2301) was identified in *S.agalactiae* <SEQ ID 6745> which encodes the amino acid sequence <SEQ ID 6746>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0921(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 9331> which encodes amino acid sequence <SEQ ID 9332> was also identified.

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

```
>GP:CAA74739 GB:Y14370 peptide chain release factor 3
          [Staphylococcus aureus]
 Identities = 274/462 (59%), Positives = 349/462 (75%), Gaps = 9/462 (1%)

Query: 1   MDIEKQRGISVTSSVMQFDYAGKRVNILDTPGHEDFSEDTYRTLMAVDAAVMVVDSAKGI 60
           M +E++RGISVTSSVMQFDY    +NILDTPGHEDFSEDTYRTLMAVD+AVMV+D AKG+
Sbjct: 57  MKVEQERGISVTSSVMQFDYDDYEINILDTPGHEDFSEDTYRTLMAVDSAVMVIDCAKGV 116

Query: 61  EAQTKKLFEVVKHRNIPVFTFINKLDRDGREPLDLLEELEEVLGIASYPMNWPIGMGKSF 120
           E  T KLF+V K R IP+FTFINKLDR G+EP +LL+E+EE L I +YPMNWPIGMG+SF
Sbjct: 117 EPPTLKLFKVCKMRGIPIFTFINKLDRVGKEPFELLDEIEETLNIETYPMNWPIGMGQSF 176

Query: 121 EGLYDLHNKRLELYKGDERFASIEDG-----DQLFANNPFYEQVKEDIELLQEAGNDFSE 175
            G+ D  +K +E ++ +E    + D      D    N+  +EQ  E++ L++EAG  F
Sbjct: 177 FGIIDRKSKTIEPFRDEENILHLNDDFELEEDHAITNDSDFEQAIEELMLVEEAGEAFDN 236

Query: 176 QAILDGDLTPVFFGSALTNFGVQTFLDTFLEFAPEPHGHKTTEGNVIDPLAKDFSGFVFK 235
           A+L GDLTPVFFGSAL NFGVQ FL+ +++FAP P+  +T E   + P    FSGF+FK
Sbjct: 237 DALLSGDLTPVFFGSALANFGVQNFLNAYVDFAPMPNARQTKENVEVSPFDDSFSGFIFK 296

Query: 236 IQANMDPRHRDRIAFVRIVSGEFERGMGVNLTRTGKGAKLSNVTQFMAES-RENVTNAVA 294
           IQANMDP+HRDRIAF+R+VSG FER + + L      +K S+V + + ++ ++ V +AVA
Sbjct: 297 IQANMDPKHRDRIAFMRVVSGAFER-VWMLLCNVLIKSKRSHVQRHLWQTIKKLVNHAVA 355

Query: 295 GDIIGVYDTGTYQVGDTLTVGKNKFEFEPLPTFTPELFMKVSAKNVMKQKSFHKGIEQLV 354
           GDIIG+YDTG YQ+GDTL  GK  + F+ LP FTPE+FMKVSAKNVMKQK FHKGIEQLV
Sbjct: 356 GDIIGLYDTGNYQIGDTLVGGKQTYSFQDLPQFTPEIFMKVSAKNVMKQKHFHKGIEQLV 415

Query: 355 QEGAIQLYKNYQTGEYMLGAVGQLQFEVFKHRMEGEYNAEVVMTPMGKKTVRW--INSDD 412
           QEGAIQ YK   T + +LGAVGQLQFEVF+HRM+ EYN +VVM P+G+K  RW   N D
Sbjct: 416 QEGAIQYYKTLHTNQIILGAVGQLQFEVFEHRMKNEYNVDVVMEPVGRKIARWDIENEDQ 475

Query: 413 LDERMSSSRNILAKDRFDQPVFLFENDFALRWFADKYPDVKL 454
           + ++M++SR+IL KDR+D  VFLFEN+FA RWF +K+P++KL
Sbjct: 476 ITDKMNTSRSILVKDRYDDLVFLFENEFATRWFEEKFPEIKL 517
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6747> which encodes the amino acid sequence <SEQ ID 6748>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2070(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 = 447/458 (97%), Positives = 455/458 (98%)

Query: 1   MDIEKQRGISVTSSVMQFDYAGKRVNILDTPGHEDFSEDTYRTLMAVDAAVMVVDSAKGI 60
           MDIEKQRGISVTSSVMQFDYAGKRVNILDTPGHEDFSEDTYRTLMAVDAAVMVVDSAKGI
Sbjct: 57  MDIEKQRGISVTSSVMQFDYAGKRVNILDTPGHEDFSEDTYRTLMAVDAAVMVVDSAKGI 116

Query: 61  EAQTKKLFEVVKHRNIPVFTFINKLDRDGREPLDLLEELEEVLGIASYPMNWPIGMGKSF 120
           EAQTKKLFEVVKHRNIPVFTFINKLDRDGREPL+LLEELEEVLGIASYPMNWPIGMG++F
Sbjct: 117 EAQTKKLFEVVKHRNIPVFTFINKLDRDGREPLELLEELEEVLGIASYPMNWPIGMGRAF 176
```

```
Query: 121 EGLYDLHNKRLELYKGDERFASIEDGDQLFANNPFYEQVKEDIELLQEAGNDFSEQAILD 180
           EGLYDLHNKRLELYKGDERFASIEDGDQLFANNPFYEQVKEDIELLQEAGNDFSEQAILD
Sbjct: 177 EGLYDLHNKRLELYKGDERFASIEDGDQLFANNPFYEQVKEDIELLQEAGNDFSEQAILD 236

Query: 181 GDLTPVFFGSALTNFGVQTFLDTFLEFAPEPHGHKTTEGNVIDPLAKDFSGFVFKIQANM 240
           GDLTPVFFGSALTNFGVQTFLDTFLEFAPEPHGHKTTEGNV+DPLAKDFSGFVFKIQANM
Sbjct: 237 GDLTPVFFGSALTNFGVQTFLDTFLEFAPEPHGHKTTEGNVVDPLAKDFSGFVFKIQANM 296

Query: 241 DPRHRDRIAFVRIVSGEFERGMGVNLTRTGKGAKLSNVTQFMAESRENVTNAVAGDIIGV 300
           DP+HRDRIAFVRIVSGEFERGMGVNLTRTGKGAKLSNVTQFMAESRENVTNAVAGDIIGV
Sbjct: 297 DPKHRDRIAFVRIVSGEFERGMGVNLTRTGKGAKLSNVTQFMAESRENVTNAVAGDIIGV 356

Query: 301 YDTGTYQVGDTLTVGKNKFEFEPLPTFTPELFMKVSAKNVMKQKSFHKGIEQLVQEGAIQ 360
           YDTGTYQVGDTLTVGKNKFEFEPLPTFTPE+FMKVS KNVMKQKSFHKGIEQLVQEGAIQ
Sbjct: 357 YDTGTYQVGDTLTVGKNKFEFEPLPTFTPEIFMKVSPKNVMKQKSFHKGIEQLVQEGAIQ 416

Query: 361 LYKNYQTGEYMLGAVGQLQFEVFKHRMEGEYNAEVVMTPMGKKTVRWINSDDLDERMSSS 420
           LYKNYQTGEYMLGAVGQLQFEVFKHRMEGEYNAEVVMTPMGKKTVRWI+ DDLD+RMSSS
Sbjct: 417 LYKNYQTGEYMLGAVGQLQFEVFKHRMEGEYNAEVVMTPMGKKTVRWISEDDLDQRMSSS 476

Query: 421 RNILAKDRFDQPVFLFENDFALRWFADKYPDVKLEEKM 458
           RNILAKDRFDQPVFLFENDFALRWFADKYPDV LEEKM
Sbjct: 477 RNILAKDRFDQPVFLFENDFALRWFADKYPDVTLEEKM 514
```

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

Example 2185

A DNA sequence (GBSx2302) was identified in *S.agalactiae* <SEQ ID 6749> which encodes the amino acid sequence <SEQ ID 6750>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3061(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:AAC38046 GB:AF000954 No definition line found [Streptococcus mutans]
 Identities = 122/142 (85%), Positives = 138/142 (96%)

Query: 1   MLEFAAQKTGKENKEMAVTFVTNERSHELNLEYRDTDRPTDVISLEYKPEVDISFDEEDL 60
           +LEFAAQKTGKE+KEMAVTFVTNERSHELNL+YRDT+RPTDVISLEYKPE  +SFDEEDL
Sbjct: 23  ILEFAAQKTGKEDKEMAVTFVTNERSHELNLKYRDTNRPTDVISLEYKPESSLSFDEEDL 82

Query: 61  AENPELAEMLEDFDSYIGELFISIDKAKEQAEEYGHSYEREMGFLAVHGFLHINGYDHYT 120
           A++P+LAE+L +FD+YIGELFIS+DKA+EQA+EYGHS+EREMGFLAVHGFLHINGYDHYT
Sbjct: 83  ADDPDLAEVLTEFDAYIGELFISVDKAREQAQEYGHSFEREMGFLAVHGFLHINGYDHYT 142

Query: 121 PEEEKEMFSLQEEILTAYGLKR 142
           P+EEKEMFSLQEEIL AYGLKR
Sbjct: 143 PQEEKEMFSLQEEILDAYGLKR 164
```

There is also homology to SEQ ID 120.

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

Example 2186

A DNA sequence (GBSx2303) was identified in *S.agalactiae* <SEQ ID 6751> which encodes the amino acid sequence <SEQ ID 6752>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-15.39   Transmembrane  108 - 124 ( 100 - 131)
   INTEGRAL    Likelihood = -8.92   Transmembrane   61 -  77 (  52 -  82)
   INTEGRAL    Likelihood = -5.36   Transmembrane   41 -  57 (  40 -  60)

----- Final Results -----
               bacterial membrane --- Certainty=0.7156(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:AAC38047 GB:AF000954 diacylglycerol kinase [Streptococcus mutans]
 Identities = 107/133 (80%), Positives = 121/133 (90%), Gaps = 2/133 (1%)

Query: 1   MDLNDN--NHKKWKNRTLTSSMEFAVTGIFTAFKEERNMRKHLVSAILVILAGLTFQVSM 58
           MDL DN  + KKWKNRTLTSS+EFA+TGIFTAFKEERNM+KH VSA+L ++AGL F+VS+
Sbjct: 3   MDLRDNKQSQKKWKNRTLTSSLEFALTGIFTAFKEERNMKKHAVSALLAVIAGLVFKVSV 62

Query: 59  VEWLFLLLSIFLVITFEIINSAIENVVDLASNYHFSMLAKNAKDMAAGAVLVVSLFAVLV 118
           +EWLFLLLSIFLVITFEI+NSAIENVVDLAS+YHFSMLAKNAKDMAAGAVLV+S FA L
Sbjct: 63  IEWLFLLLSIFLVITFEIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVLVISGFAALT 122

Query: 119 GLIIFIPKILALL 131
           GLIIF+PKI  LL
Sbjct: 123 GLIIFVPKIWFLL 135
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6753> which encodes the amino acid sequence <SEQ ID 6754>. Analysis of this protein sequence reveals the following:

```
     Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.67   Transmembrane   63 -  79 (  41 -  84)
   INTEGRAL    Likelihood = -7.32   Transmembrane  110 - 126 ( 105 - 129)
   INTEGRAL    Likelihood = -5.41   Transmembrane   43 -  59 (  41 -  62)

----- 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 databases:

```
>GP:AAC38047 GB:AF000954 diacylglycerol kinase [Streptococcus mutans]
 Identities = 104/135 (77%), Positives = 119/135 (88%)

Query: 1   MALHDNNTTKRKWKNRTITSSLEFALTGVFTAFKEERNLRSHLLSACLACVAGLFFSISA 60
           M L DN  +++KWKNRT+TSSLEFALTG+FTAFKEERN++ H +SA LA +AGL F +S
Sbjct: 3   MDLRDNKQSQKKWKNRTLTSSLEFALTGIFTAFKEERNMKKHAVSALLAVIAGLVFKVSV 62

Query: 61  IEWLFLLLAIFLVITLEIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVLMISGYAVLT 120
           IEWLFLLL+IFLVIT EIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVL+ISG+A LT
Sbjct: 63  IEWLFLLLSIFLVITFEIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVLVISGFAALT 122

Query: 121 GLIIFIPKIWNIFVH 135
           GLIIF+PKIW +  H
Sbjct: 123 GLIIFVPKIWFLLFH 137
```

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

```
 Identities = 98/129 (75%), Positives = 115/129 (88%), Gaps = 2/129 (1%)

Query: 1   MDLNDNN--HKKWKNRTLTSSMEFAVTGIFTAFKEERNMRKHLVSAILVILAGLTFQVSM 58
           M L+DNN   +KWKNRT+TSS+EFA+TG+FTAFKEERN+R HL+SA L  +AGL F +S
Sbjct: 1   MALHDNNTTKRKWKNRTITSSLEFALTGVFTAFKEERNLRSHLLSACLACVAGLFFSISA 60

Query: 59  VEWLFLLLLSIFLVITFEIINSAIENVVDLASNYHFSMLAKNAKDMAAGAVLVVSLFAVLV 118
           +EWLFLLL+IFLVIT EI+NSAIENVVDLAS+YHFSMLAKNAKDMAAGAVL++S +AVL
Sbjct: 61  IEWLFLLLAIFLVITLEIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVLMISGYAVLT 120

Query: 119 GLIIFIPKI 127
           GLIIFIPKI
Sbjct: 121 GLIIFIPKI 129
```

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

Example 2187

A DNA sequence (GBSx2304) was identified in *S.agalactiae* <SEQ ID 6755> which encodes the amino acid sequence <SEQ ID 6756>. This protein is predicted to be GTPase Era (era). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1871(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 10017> which encodes amino acid sequence <SEQ ID 10018> was also identified.

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

```
>GP:AAD41632 GB:AF072811 GTPase Era [Streptococcus pneumoniae]
 Identities = 273/299 (91%), Positives = 290/299 (96%)

Query: 16  MTFKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTTRNKIMGIYTTETEQIVFIDT 75
           MTFKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTTRNKIMGIYTT+ EQIVFIDT
Sbjct: 1   MTFKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTTRNKIMGIYTTDKEQIVFIDT 60

Query: 76  PGIHKPKTALGDFMVESAYSTLREVETVLFMVPADEKRGKGDDMIIERLKAAKIPVILVI 135
           PGIHKPKTALGDFMVESAYSTLREV+TVLFMVPADE RGKGDDMIIERLKAAK+PVILV+
Sbjct: 61  PGIHKPKTALGDFMVESAYSTLREVDTVLFMVPADEARGKGDDMIIERLKAAKVPVILVV 120

Query: 136 NKIDKVHPDQLLEQIDDFRSQMDFKEVVPISALQGNNVPTLIKLLTDNLEEGFQYFPEDQ 195
           NKIDKVHPDQLL QIDDFR+QMDFKE+VPISALQGNNV  L+ +L++NL+EGFQYFP DQ
Sbjct: 121 NKIDKVHPDQLLSQIDDFRNQMDFKEIVPISALQGNNVSRLVDILSENLDEGFQYFPSDQ 180

Query: 196 ITDHPERFLVSEMVREKVLHLTQQEVPHSVAVVVESMKRDEETDKVHIRATIMVERDSQK 255
           ITDHPERFLVSEMVREKVLHLT++E+PHSVAVVV+SMKRDEETDKVHIRATIMVERDSQK
Sbjct: 181 ITDHPERFLVSEMVREKVLHLTREEIPHSVAVVVDSMKRDEETDKVHIRATIMVERDSQK 240

Query: 256 GIIIGKQGAMLKKIGKMARRDIELMLGDKVYLETWVKVKKNWRDKKLDLADFGYNEKEY 314
           GIIIGK GAMLKKIG MARRDIELMLGDKV+LETWVKVKKNWRDKKLDLADFGYNE+EY
Sbjct: 241 GIIIGKGGAMLKKIGSMARRDIELMLGDKVFLETWVKVKKNWRDKKLDLADFGYNEREY 299
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6757> which encodes the amino acid sequence <SEQ ID 6758>. Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1088(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 = 295/297 (99%), Positives = 296/297 (99%)

Query: 18  FKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTTRNKIMGIYTTETEQIVFIDTPG 77
           FKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTTRNKIMGIYTTETEQIVFIDTPG
Sbjct: 2   FKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTTRNKIMGIYTTETEQIVFIDTPG 61

Query: 78  IHKPKTALGDFMVESAYSTLREVETVLFMVPADEKRGKGDDMIIERLKAAKIPVILVINK 137
           IHKPKTALGDFMVESAYSTLREVETVLFMVPADEKRGKGDDMIIERLKAAKIPVILVINK
Sbjct: 62  IHKPKTALGDFMVESAYSTLREVETVLFMVPADEKRGKGDDMIIERLKAAKIPVILVINK 121

Query: 138 IDKVHPDQLLEQIDDFRSQMDFKEVVPISALQGNNVPTLIKLLTDNLEEGFQYFPEDQIT 197
           IDKVHPDQLLEQIDDF SQMDFKEVVPISAL+GNNVPTLIKLLTDNLEEGFQYFPEDQIT
Sbjct: 122 IDKVHPDQLLEQIDDFHSQMDFKEVVPISALEGNNVPTLIKLLTDNLEEGFQYFPEDQIT 181

Query: 198 DHPERFLVSEMVREKVLHLTQQEVPHSVAVVVESMKRDEETDKVHIRATIMVERDSQKGI 257
           DHPERFLVSEMVREKVLHLTQQEVPHSVAVVVESMKRDEETDKVHIRATIMVERDSQKGI
Sbjct: 182 DHPERFLVSEMVREKVLHLTQQEVPHSVAVVVESMKRDEETDKVHIRATIMVERDSQKGI 241

Query: 258 IIGKQGAMLKKIGKMARRDIELMLGDKVYLETWVKVKKNWRDKKLDLADFGYNEKEY 314
           IIGKQGAMLKKIGKMARRDIELMLGDKVYLETWVKVKKNWRDKKLDLADFGYNEKEY
Sbjct: 242 IIGKQGAMLKKIGKMARRDIELMLGDKVYLETWVKVKKNWRDKKLDLADFGYNEKEY 298
```

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

Example 2188

A DNA sequence (GBSx2305) was identified in *S.agalactiae* <SEQ ID 6759> which encodes the amino acid sequence <SEQ ID 6760>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2679(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 2189

A DNA sequence (GBSx2306) was identified in *S.agalactiae* <SEQ ID 6761> which encodes the amino acid sequence <SEQ ID 6762>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
```

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

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

```
>GP:BAA16793 GB:D90900 hypothetical protein [Synechocystis sp.]
 Identities = 36/119 (30%), Positives = 57/119 (47%), Gaps = 15/119 (12%)

Query: 390 TSDYEKAKVIHDHLVNNYTYATEELATTRETASGISIHAPEALYKDKRGVCQAFAVMFKD 449
           ++D+E+A++ +  +  N  Y    +A TR     I    PE +      +C  ++ +++
Sbjct: 153 SNDWEEARLAYSWITQNIAYDVP-MAETRN----IDDLRPETVLARGETICSGYSNLYQA 207

Query: 450 MAATAGLSVWYVTGQAGGG----------NHAWNIVTINGVKYYVDTTWDNNIKSNKYF 498
           +A   GL V  + G A GG          NHAWN V I+G  Y +DTTW   I S+  F
Sbjct: 208 LAKELGLDVVIIEGFAKGGDVIVGDDPDVNHAWNGVKIDGQWYLLDTTWGAGIVSDGKF 266
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6763> which encodes the amino acid sequence <SEQ ID 6764>. Analysis of this protein sequence reveals the following:

```
      Possible site: 23

>>> 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.

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

```
 Identities = 41/181 (22%), Positives = 79/181 (42%), Gaps = 17/181 (9%)

Query: 355 ITITYTLKGDMVGLHKEYKQFVDSFVKENITNKNITSDYEKAKVIHDHLVNNYTYATE-- 412
           + +T+ +  D   ++++  Q + + +  N  +K+    YE+ K  ++ ++ +  Y  +
Sbjct: 124 VFVTFPIPEDAKNIYQDL-QAIGNDIVANTPSKD---RYEQVKYFYEVIIRDTDYNKKAF 179

Query: 413 ELATTRETASGISIHAPEALYKDKRGVCQAFAVMFKDMAATAGLSVWYVTGQAGGGN--- 469
           E   +   A   S    ++++ D   VC  +A  F+ +   AG+ V Y+ G
Sbjct: 180 EAYQSGSQAQVASNQDIKSVFIDHLSVCNGYAQAFQFLCQKAGIPVAYIRGTGTSQQPQQ 239

Query: 470 ---HAWNIVTINGVKYYVDTTW-----DNNIKSNKYFLVGKTIMDADHLLDSQYNALAKDI 522
              HAWN V IN   Y VD TW     DN++   K   +  +     L +  +  +KDI
Sbjct: 240 SFAHAWNAVQINNTYYGVDVTWGDPVFDNHLSHQKQGTINYSFLCLPDYLMALSHQPSKDI 300
```

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

Example 2190

A DNA sequence (GBSx2307) was identified in *S.agalactiae* <SEQ ID 6765> which encodes the amino acid sequence <SEQ ID 6766>. This protein is predicted to be rgg protein. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16   Transmembrane  187 - 203 ( 187 - 203)

----- 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 10015> which encodes amino acid sequence <SEQ ID 10016> was also identified.

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

```
>GP:AAA26968 GB:M89776 rgg [Streptococcus gordonii]
 Identities = 71/273 (26%), Positives = 140/273 (51%), Gaps = 16/273 (5%)

Query: 8   KELGKTLRRLRKGKKVSISSLADEHLSKSQISRFERGESEITCSRLLNILDKLNITIDEF 67
           K  GK L+ +R+ K +S+  +A   +S +Q+SR+ERG S +T     + L  +++++ EF
Sbjct: 5   KSSGKILKIIRESKNMSLKEVAAGDISVAQLSRYERGISSLTVDSFYSCLRNMSVSLAEF 64

Query: 68  VSI-HSKAHTHFFILLNRVRKYCAEKNVTKLVALL----------EDHNHKDYEKIMIK 115
             + H+       +L  ++ +   E N+ KL ++L          E  N+K    I+I+
Sbjct: 65  QYVYHNYREADDVVLSQKLSEAQRENNIVKLESILAGSEAMAQEFPEKKNYK-LNTIVIR 123

Query: 116 ALIFSIDQSIEPNQEELARLTDYLFTVEQWGYYEIILLGNCSRLINYNTLFLLTKEMVNS 175
           A + S +   + ++ ++  LTDYLF+VE+WG YE+ L  N   L+   TL     EM+N
Sbjct: 124 ATLTSCNPDYQVSKGDIEFLTDYLFSVEEWGRYELWLFTNSVNLLTLETLETFASEMINR 183

Query: 176 FAYSEQNKTNKILVTQLAINCLIISIDHSYFEHSHYLIDKVRSLLQDEVNFYEKTVFLYV 235
             +      N+  + ++ +N +   I++++ + +   ++ + +    E + Y++ +  Y
Sbjct: 184 TQFYNNLPENRRRIIKMLLNVVSACIENNHLQVAMKFLNYIDNTKIPETDLYDRVLIKYH 243

Query: 236 TGYYHLKLGDTSSGKEDMRKALQIFKYLGEDSF 268
              Y  K+G+  + + D+ + L  F+YL  DSF
Sbjct: 244 KALYSYKVGNPHA-RHDIEQCLSTFEYL--DSF 273
```

There is also homology to SEQ ID 628.

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

Example 2191

A DNA sequence (GBSx2308) was identified in *S.agalactiae* <SEQ ID 6767> which encodes the amino acid sequence <SEQ ID 6768>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3234(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:BAA05066 GB:D26071 formamidopyrimidine-DNA glycosylase
           [Streptococcus mutans]
 Identities = 182/271 (67%), Positives = 217/271 (79%)

Query: 1   MPELPEVETVRKGLERLVVNQEIASITIKVPKMVKTDLNDFMISLPGKTIQQVLRRGKYL 60
           MPELPEVETVR+GLE L+V ++I S+ ++VPKMVKT + DF + + G+T + + RRGKYL
Sbjct: 1   MPELPEVETVRRGLEHLIVGKKIVSVEVRVPKMVKTGVEDFQLDILGQTFESIGRRGKYL 60

Query: 61  LFDFGEMVMVSHLRMEGKYLLFPNKVPDNKHFHLYFKLTNGSTLVYQDVRKFGTFELVRK 120
           L +     ++SHLRMEGKYLLF ++VPDNKHFHL+F L  GSTLVYQDVRKFGTFEL+ K
Sbjct: 61  LLNLNRQTIISHLRMEGKYLLFEDEVPDNKHFHLFFGLDGGSTLVYQDVRKFGTFELLPK 120

Query: 121 SSLKDYFTQKKLGPEPTADTFQFEPFSKGLANSKKPIKPLLLDQRLVAGLGNIYVDEVLW 180
           S ++ YF QKK+GPEP A  F+ +PF +GLA S K IK LLLDQ LVAGLGNIYVDEVLW
Sbjct: 121 SQVEAYFVQKKIGPEPNAKDFKLKPFEEGLAKSHKVIKTLLLDQHLVAGLGNIYVDEVLW 180

Query: 181 AAKIHPQRLANQLTESETSLLHKEIIRILTLGIEKGGSTIRTYKNALGEDGTMQKYLQVY 240
```

```
          AAK+ P+RLA+QL  SE   +H E IRIL L IEKGGSTIR+YKN+LGEDG+MQ  LQVY
Sbjct: 181 AAKVDPERLASQLKTSEIKRIHDETIRILQLAIEKGGSTIRSYKNSLGEDGSMQDCLQVY 240

Query: 241 GKTGQPCPRCGCLIKKIKVGGRGTHYCPRCQ 271
           GKT QPC RC   I+KIKVGGRGTH+CP CQ
Sbjct: 241 GKTDQPCARCATPIEKIKVGGRGTHFCPSCQ 271
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6769> which encodes the amino acid sequence <SEQ ID 6770>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2068(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 = 190/271 (70%), Positives = 229/271 (84%)

Query: 1   MPELPEVETVRKGLERLVVNQEIASITIKVPKMVKTDLNDFMISLPGKTIQQVLRRGKYL 60
           MPELPEVETVR+GLE LV+ QEI ++T+KVPKMVKTDL  F ++LPG+ IQ V RRGKYL
Sbjct: 1   MPELPEVETVRRGLETLVLGQEIVAVTLKVPKMVKTDLETFALTLPGQIIQSVGRRGKYL 60

Query: 61  LFDFGEMVMVSHLRMEGKYLLFPNKVPDNKHFHLYFKLTNGSTLVYQDVRKFGTFELVRK 120
           L D G++V+VSHLRMEGKYLLFP++VPDNKHFH++F+L NGSTLVYQDVRKFGTF+L+ K
Sbjct: 61  LIDLGQLVLVSHLRMEGKYLLFPDEVPDNKHFHVFFELKNGSTLVYQDVRKFGTFDLIAK 120

Query: 121 SSLKDYFTQKKLGPEPTADTFQFEPFSKGLANSKKPIKPLLLDQRLVAGLGNIYVDEVLW 180
           S L  +F ++KLGPEP  +TF+ + F   L +S+KPIKP LLDQ LVAGLGNIYVDEVLW
Sbjct: 121 SQLSAFFAKRKLGPEPKKETFKLKTFEAALLSSQKPIKPHLLDQTLVAGLGNIYVDEVLW 180

Query: 181 AAKIHPQRLANQLTESETSLLHKEIIRILTLGIEKGGSTIRTYKNALGEDGTMQKYLQVY 240
           AAK+HP+  +++L ++E   LH E IRIL LGIEKGGST+RTY+NALG DGTMQ YLQVY
Sbjct: 181 AAKVHPETASSRLNKAEIKRLHDETIRILALGIEKGGSTVRTYRNALGADGTMQDYLQVY 240

Query: 241 GKTGQPCPRCGCLIKKIKVGGRGTHYCPRCQ 271
           G+TG+PCPRCG  I K+KVGGRGTH CP+CQ
Sbjct: 241 GQTGKPCPRCGQAIVKLKVGGRGTHICPKCQ 271
```

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

Example 2192

A DNA sequence (GBSx2309) was identified in *S.agalactiae* <SEQ ID 6771> which encodes the amino acid sequence <SEQ ID 6772>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0797(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 10013> which encodes amino acid sequence <SEQ ID 10014> was also identified.

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

```
>GP:AAC00353 GB:AF008220 YtaG [Bacillus subtilis]
```

```
 Identities = 80/189 (42%), Positives = 113/189 (59%), Gaps = 1/189 (0%)

Query: 8   MTKIIGLTGGIASGKSTVTKIIRESGFKVIDADQVVHKLQAKGGKLYQALLEWLGPEILD 67
           MT +IGLTGGIASGKSTV  ++ E G  VIDAD +  +   KG   Y+ +++  G +IL
Sbjct: 1   MTLVIGLTGGIASGKSTVANMLIEKGITVIDADIIAKQAVEKGMPAYRQIIDEFGEDILL 60

Query: 68  ADGELDRPKLSQMIFANPDNMKTSARLQNSIIRQELACQRDQLKQTEEIF-FMDIPLLIE 126
           ++G++DR KL  ++F N          + +  +RQE+  +RD+     E F  +DIPLL E
Sbjct: 61  SNGDIDRKKLGALVFTNEQKRLALNAIVHPAVRQEMLNRRDEAVANREAFVVLDIPLLFE 120

Query: 127 EKYIKWFDEIWLVFVDKEKQLQRLMARNNYSREEAELRLSHQMPLTDKKSFASLIIDNNG 186
            K     D+I +V V KE QL+RLM RN  + EEA  R+  QMPL +K + A  +IDN+G
Sbjct: 121 SKLESLVDKIIVVSVTKELQLERLMKRNQLTEEEAVSRIRSQMPLEEKTARADQVIDNSG 180

Query: 187 DLITLKEQI 195
           L   K Q+
Sbjct: 181 TLEETKRQL 189
```

A related sequence was also identified in GAS <SEQ ID 9111> which encodes the amino acid sequence <SEQ ID 9112>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty= 0.101(Affirmative) < succ>
               bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
                bacterial outside --- Certainty= 0.000(Not Clear) < succ>
```

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

```
 Identities = 118/191 (61%), Positives = 153/191 (79%)

Query: 9   TKIIGLTGGIASGKSTVTKIIRESGFKVIDADQVVHKLQAKGGKLYQALLEWLGPEILDA 68
           T IIG+TGGIASGKSTV K+IR++G++VIDADQVVH LQ KGG+LY+AL E  G +IL A
Sbjct: 9   TMIIGITGGIASGKSTVVKVIRKAGYQVIDADQVVHDLQEKGGRLYEALREAFGNQILKA 68

Query: 69  DGELDRPKLSQMIFANPDNMKTSARLQNSIIRQELACQRDQLKQTEEIFFMDIPLLIEEK 128
           DGELDR KLS+M+F+NPDNM TS+ +QN II++ELA +RD L Q++ IFFMDIPLL+E
Sbjct: 69  DGELDRTKLSEMLFSNPDNMATSSAIQNQIIKEELAAKRDHLAQSQAIFFMDIPLLMELG 128

Query: 129 YIKWFDEIWLVFVDKEKQLQRLMARNNYSREEAELRLSHQMPLTDKKSFASLIIDNNGDL 188
           Y  WFD IWLV+VD + QLQRLMARN   + +A  R++ Q+P+ +KK +ASL+IDN+GD+
Sbjct: 129 YQDWFDAIWLVYVDAQTQLQRLMARNRLDKGKARQRIASQLPIEEKKPYASLVIDNSGDI 188

Query: 189 ITLKEQILDAL 199
             L +Q+  AL
Sbjct: 189 AALIKQVQSAL 199
```

A related GBS gene <SEQ ID 8993> and protein <SEQ ID 8994> were also identified. Analysis of this protein sequence reveals a signal peptide at residues 1-16.

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

```
42.2/60.6% over 189aa

  OMNI|NT01BS3382| Insert characterized

ORF02237(319 - 885 of 1206)
OMNI|NT01BS3382(3 - 192 of 200) ()
%Match = 17.0
%Identity = 42.1  %Similarity = 60.5
Matches = 80  Mismatches = 74  Conservative Sub.s = 35

78        108        138        168        198        228        258        288
KNSPTAFG*SIDRI*NKLITQGNYSHFNFRHRKRWLHD*NI*ECSWRGRYDAKVFTGLW*NWATVSKVWLFN*EDKSRRE
```

```
318       348       378       408       438       468       498       528
RDALLPSVSMLMTKIIGLTGGIASGKSTVTKIIRESGFKVIDADQVVHKLQAKGGKLYQALLEWLGPEILDADGELDRPK
          |:| :|||||||||||||||  :: | |  |||| :  :   ||   |: ::: :| :|| ::|::|| |
VDLLTLVIGLTGGIASGKSTVANMLIEKGITVIDADIIAKQAVEKGMPAYRQIIDEFGEDILLSNGDIDRKK
                10        20        30        40        50        60        70

558       588       618       648       675       705       735       765
LSQMIFANPDNMKTSARLQNSIIRQELACQRDQLKQTEEIFF-MDIPLLIEEKYIKWFDEIWLVFVDKEKQLQRLMARNN
|  ::| |        : :  :|||:  :||:     | |  :|||||| |      |:| :| | || ||:||| ||
LGALVFTNEQKRLALNAIVHPAVRQEMLNRRDEAVANREAFVVLDIPLLFESKLESLVDKIIVVSVTKELQLERLMKRNQ
          .       90        100       110       120       130       140       150

795       825       855       885       915       945       975       1005
YSREEAELRLSHQMPLTDKKSFASLIIDNNGDLITLKEQILDALQRL*NY*MDNVFIHFLSLLH*F*KTCD*TTVIVQ*Y
 : |||  |:  ||||  :| : |  :|||:| |   | |: : :
LTEEEAVSRIRSQMPLEEKTARADQVIDNSGTLEETKRQLDEIMNSWA
               170       180       190       200
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6773> which encodes amino acid sequence <SEQ ID 6774>. An alignment of the GAS and GBS sequences follows:

```
Score =  218 bits (550), Expect = 4e-59
Identities = 104/175 (59%), Positives = 138/175 (78%)

Query: 25  VVKVIRKAGYQVIDADQVVHDLQEKGGRLYEALREAFGNQILKADGELDRTKLSEMLFSN 84
           V K+IR++G++VIDADQVVH LQ KGG+LY+AL E  G +IL ADGELDR KLS+M+F+N
Sbjct: 20  VTKIIRESGFKVIDADQVVHKLQAKGGKLYQALLEWLGPEILDADGELDRPKLSQMIFAN 79

Query: 85  PDNMATSSAIQNQIIKEELAAKRDHLAQSQAIFFMDIPLLMELGYQDWFDAIWLVYVDAQ 144
           PDNM TS+ +QN II++ELA +RD L Q++ IFFMDIPLL+E  Y  WFD IWLV+VD +
Sbjct: 80  PDNMKTSARLQNSIIRQELACQRDQLKQTEEIFFMDIPLLIEEKYIKWFDEIWLVFVDKE 139

Query: 145 TQLQRLMARNRLDKGKARQRIASQLPIEEKKPYASLVIDNSGDIAALIKQVQSAL 199
            QLQRLMARN   + +A  R++ Q+P+ +KK +ASL+IDN+GD+  L +Q+  AL
Sbjct: 140 KQLQRLMARNNYSREEAELRLSHQMPLTDKKSFASLIIDNNGDLITLKEQILDAL 194
```

SEQ ID 8994 (GBS245) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 6; MW 23.7kDa). It was also expressed in *E.coli* as a GST-fusion product, and purified GBS245-GST is shown in Figure 211, lane 6.

The purified GST fusion product was used to immunise mice ands the resulting antiserum was used for FACS (Figure 278). This 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 2193

A DNA sequence (GBSx2310) was identified in *S.agalactiae* <SEQ ID 6775> which encodes the amino acid sequence <SEQ ID 6776>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4073(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:BAA30330 GB:AP000005 253aa long hypothetical ATP-binding
```

```
            transport protein [Pyrococcus horikoshii]
 Identities = 78/240 (32%), Positives = 130/240 (53%), Gaps = 13/240 (5%)

Query: 3   LVIRDIRKRFQETEVLRGASYRFYSGKITGVLGRNGAGKTTLFNILYGDLAADNGTICLL 62
           +++ ++RK+F   EVL+G ++    G+I G+LG NG+GK+T   IL G +    G + +
Sbjct: 2   IIVENLRKKFGSKEVLKGINFTVNDGEIYGLLGPNGSGKSTTMRILSGIITDFEGKVMVA 61

Query: 63  -KDNHEYPLTDKDI-GIVYSENYLPEFLTGYEFVKFYMDLH--PSDDL-MTIDDYLDFME 117
             D    P+  K+I G V     L E LT  EF  F   +   P D L   +   +D
Sbjct: 62  GVDVSRDPMKVKEIVGYVPETPALYESLTPAEFFSFIGGVRRIPQDILEERVKRLVDAFG 121

Query: 118 IGQTERHRIIKGYSDGMKSKLSLICLMISKPKVILLDEPLTAVDVVSSIAIKRLLLELSE 177
           IG+   +++I   S G K K+SLI   ++  P+V++LDE +  +D  S+   + LL E  E
Sbjct: 122 IGK-YMNQLIGTLSFGTKQKISLISALLHDPQVLILDEAMNGLDPKSARIFRELLFEFKE 180

Query: 178 D-HIIILSTHIMALAEDLCDIVAVLDKGKL---QTLDIDR---KHEQFEERLLQVLKGDE 230
           +   I+ STHI+ALAE +CD + ++ +G++    T+D  R   + E+ E+  L++ +  E
Sbjct: 181 EGKSIVFSTHILALAEVMCDRIGIIYEGRIVAEGTIDELREIAREEKLEDIFLKLTQAKE 240
```

There is also homology to SEQ ID 2876.

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

Example 2194

A DNA sequence (GBSx2311) was identified in *S.agalactiae* <SEQ ID 6777> which encodes the amino acid sequence <SEQ ID 6778>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.6138(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 2195

A DNA sequence (GBSx2312) was identified in *S.agalactiae* <SEQ ID 6779> which encodes the amino acid sequence <SEQ ID 6780>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-15.34   Transmembrane  526 - 542 ( 511 - 546)
   INTEGRAL    Likelihood = -9.61   Transmembrane  340 - 356 ( 335 - 359)
   INTEGRAL    Likelihood = -8.17   Transmembrane  455 - 471 ( 451 - 476)
   INTEGRAL    Likelihood = -8.01   Transmembrane   97 - 113 (  95 - 121)
   INTEGRAL    Likelihood = -8.01   Transmembrane  216 - 232 ( 207 - 236)
   INTEGRAL    Likelihood = -3.40   Transmembrane   50 -  66 (  46 -  67)
   INTEGRAL    Likelihood = -1.33   Transmembrane  178 - 194 ( 178 - 194)

----- Final Results -----
            bacterial membrane --- Certainty=0.7135(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 10011> which encodes amino acid sequence <SEQ ID 10012> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 376.

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 2196

A DNA sequence (GBSx2314) was identified in *S.agalactiae* <SEQ ID 6781> which encodes the amino acid sequence <SEQ ID 6782>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.17   Transmembrane  140 - 156 ( 134 - 160)
   INTEGRAL    Likelihood = -6.64   Transmembrane  255 - 271 ( 253 - 274)
   INTEGRAL    Likelihood = -5.79   Transmembrane  345 - 361 ( 343 - 363)
   INTEGRAL    Likelihood = -3.29   Transmembrane  184 - 200 ( 183 - 202)
   INTEGRAL    Likelihood = -2.34   Transmembrane   66 -  82 (  65 -  83)
   INTEGRAL    Likelihood = -1.65   Transmembrane  221 - 237 ( 221 - 239)
   INTEGRAL    Likelihood = -0.00   Transmembrane  121 - 137 ( 121 - 137)

----- Final Results -----
               bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9401> which encodes amino acid sequence <SEQ ID 9402> was also identified.

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

```
>GP:CAA07482 GB:AJ007367 multi-drug resistance efflux pump
           [Streptococcus pneumoniae]
 Identities = 213/372 (57%), Positives = 295/372 (79%)

Query: 1   MPFMVLYVEQLGAPSNKVEWYAGLSVSLSALSSALVAPLWGRLADKYGRKPMMVRAGLMM 60
           +PFM ++VE LG  S +V +YAGL++S+SA+S+AL +P+WG LADKYGRKPMM+RAGL M
Sbjct: 28  VPFMPIFVENLGVGSQQVAFYAGLAISVSAISAALFSPIWGILADKYGRKPMMIRAGLAM 87

Query: 61  TFTMGGLAFIHSVTGLLILRILNGIFAGYVPNSTALIASQAPQEESGYALGTLATGVTGG 120
           T TMGGLAF+ ++  L+ LR+LNG+FAG+VPN+TALIASQ P+E+SG ALGTL+TGV  G
Sbjct: 88  TITMGGLAFVPNIYWLIFLRLLNGVFAGFVPNATALIASQVPKEKSGSALGTLSTGVVAG 147

Query: 121 MLIGPLLGGLLAEWFGIREVFLLVGTILLISTLMTIFMVKEDFKPISNEETMPTTEVFKS 180
            L GP +GG +AE FGIR VFLLVG+ L ++ ++TI  +KEDF+P++ E+ +PT E+F S
Sbjct: 148 TLTGPFIGGFIAELFGIRTVFLLVGSFLFLAAILTICFIKEDFQPVAKEKAIPTKELFTS 207

Query: 181 VKSLQILIGLFVTSMIIQISAQSIAPILTLYIRHLGQTENLMFVSGLIVSGMGFSSILSS 240
           VK   +L+ LF+TS +IQ SAQSI PIL LY+R LGQTENL+FVSGLIVS MGFSS++S+
Sbjct: 208 VKYPYLLLLNLFLTSFVIQFSAQSIGPILALYVRDLGQTENLLFVSGLIVSSMGFSSMMSA 267

Query: 241 PKLGRIGDRIGNHRLLLLALLYSFLMYVLCSLAQTSLQLGVIRFLYGFGTGALMPSINSI 300
             +G++GD++GNHRLL++A  YS ++Y+LC+ A + LQLG+ RFL+G GTGAL+P +N++
Sbjct: 268 GVMGKLGDKVGNHRLLVVAQFYSVIIYLLCANASSPLQLGLYRFLFGLGTGALIPGVNAL 327

Query: 301 LTKIAPRQGLSRIFSYNQMFSNLGQVLGPFVGSAVSIHLGFRWVFFVTSFIVLANFVWCF 360
           L+K+ P+ G+SR+F++NQ+F  LG V+GP  GSAV+   G+  VF+ TS  V  + ++
```

```
Sbjct: 328 LSKMTPKAGISRVFAFNQVFFYLGGVVGPMAGSAVAGQFGYHAVFYATSLCVAFSCLFNL 387

Query: 361 INFRKYIRVKEI 372
           I FR  ++VKEI
Sbjct: 388 IQFRTLLKVKEI 399
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6783> which encodes the amino acid sequence <SEQ ID 6784>. Analysis of this protein sequence reveals the following:

```
      Possible site: 58
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.14   Transmembrane  165 - 181 ( 150 - 185)
   INTEGRAL    Likelihood = -7.43   Transmembrane  371 - 387 ( 367 - 391)
   INTEGRAL    Likelihood = -3.88   Transmembrane   90 - 106 (  86 - 109)
   INTEGRAL    Likelihood = -3.35   Transmembrane  145 - 161 ( 143 - 162)
   INTEGRAL    Likelihood = -1.70   Transmembrane  279 - 295 ( 279 - 297)
   INTEGRAL    Likelihood = -0.85   Transmembrane  209 - 225 ( 209 - 226)
   INTEGRAL    Likelihood = -0.27   Transmembrane  347 - 363 ( 347 - 363)

----- Final Results -----
               bacterial membrane --- Certainty=0.5055(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:CAA07482 GB:AJ007367 multi-drug resistance efflux pump
           [Streptococcus pneumoniae]
 Identities = 236/396 (59%), Positives = 309/396 (77%)

Query: 1   VNWRQNLKVAWLGNFFTGASFSLVMPFMALYVENLGTPTELVEYYAGLAVAVTALASALF 60
           +NW+ NL++AW GNF TGAS SLV+PFM ++VENLG  ++ V +YAGLA++V+A+++ALF
Sbjct: 4   INWKDNLRIAWFGNFLTGASISLVVPFMPIFVENLGVGSQQVAFYAGLAISVSAISAALF 63

Query: 61  APVWGKLADRYGRKPMMLRASFVMTFTMGGLAIIPNVFWLLILRLLTGVSAGYVPNATAL 120
           +P+WG LAD+YGRKPMM+RA   MT TMGGLA +PN++WL+ LRLL GV AG+VPNATAL
Sbjct: 64  SPIWGILADKYGRKPMMIRAGLAMTITMGGLAFVPNIYWLIFLRLLNGVFAGFVPNATAL 123

Query: 121 IASQAPKEESGYALGTLATGVTAGALIGPLLGGILAELLGIRQVFLLVGVILFLCSLMTA 180
           IASQ PKE+SG ALGTL+TGV AG L GP +GG +AEL GIR VFLLVG  LFL +++T
Sbjct: 124 IASQVPKEKSGSALGTLSTGVVAGTLTGPFIGGFIAELFGIRTVFLLVGSFLFLAAILTI 183

Query: 181 VYVKEEFKPVRRFEMIPTKVILKQVKSPQIMLGLFVTSMIIQISAQSVAPILSLYIRHLG 240
            ++KE+F+PV + + IPTK +   VK P ++L LF+TS +IQ SAQS+ PIL+LY+R LG
Sbjct: 184 CFIKEDFQPVAKEKAIPTKELFTSVKYPYLLLNLFLTSFVIQFSAQSIGPILALYVRDLG 243

Query: 241 QTHNLMFTSGLVVSAMGFSSLFSSSYLGKLGDRFGNHRLLLAALCYSFIMYFSSALAQTS 300
           QT NL+F SGL+VS+MGFSS+ S+  +GKLGD+ GNHRLL+ A  YS I+Y   A A +
Sbjct: 244 QTENLLFVSGLIVSSMGFSSMMSAGVMGKLGDKVGNHRLLVVAQFYSVIIYLLCANASSP 303

Query: 301 FQLGVLRFAYGFGVGALMPSINSLLTKLTPKEGISRVFAYNQMFSNLGQVIGPFIGSNVA 360
            QLG+ RF +G G GAL+P +N+LL+K+TPK GISRVFA+NQ+F  LG V+GP  GS VA
Sbjct: 304 LQLGLYRFLFGLGTGALIPGVNALLSKMTPKAGISRVFAFNQVFFYLGGVVGPMAGSAVA 363

Query: 361 VVLGYRSVFYVTSLIVFVNLIWSLIIFRKYIKVKDI 396
             GY +VFY TSL V  + +++LI FR  +KVK+I
Sbjct: 364 GQFGYHAVFYATSLCVAFSCLFNLIQFRTLLKVKEI 399
```

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

```
 Identities = 262/373 (70%), Positives = 314/373 (83%)

Query: 1   MPFMVLYVEQLGAPSNKVEWYAGLSVSLSALSSALVAPLWGRLADKYGRKPMMVRAGLMM 60
           MPFM LYVE LG P+  VE+YAGL+V+++AL+SAL AP+WG+LAD+YGRKPMM+RA  +M
Sbjct: 25  MPFMALYVENLGTPTELVEYYAGLAVAVTALASALFAPVWGKLADRYGRKPMMLRASFVM 84

Query: 61  TFTMGGLAFIHSVTGLLILRILNGIFAGYVPNSTALIASQAPQEESGYALGTLATGVTGG 120
```

```
            TFTMGGLA I +V  LLILR+L G+ AGYVPN+TALIASQAP+EESGYALGTLATGVT G
Sbjct: 85   TFTMGGLAIIPNVFWLLILRLLTGVSAGYVPNATALIASQAPKEESGYALGTLATGVTAG 144

Query: 121 MLIGPLLGGLLAEWFGIREVFLLVGTILLISTLMTIFMVKEDFKPISNEETMPTTEVFKS 180
            LIGPLLGG+LAE  GIR+VFLLVG IL + +LMT   VKE+FKP+   E +PT  + K
Sbjct: 145 ALIGPLLGGILAELLGIRQVFLLVGVILFLCSLMTAVYVKEEFKPVRRFEMIPTKVILKQ 204

Query: 181 VKSLQILIGLFVTSMIIQISAQSIAPILTLYIRHLGQTENLMFVSGLIVSGMGFSSILSS 240
            VKS QI++GLFVTSMIIQISAQS+APIL+LYIRHLGQT NLMF SGL+VS MGFSS+ SS
Sbjct: 205 VKSPQIMLGLFVTSMIIQISAQSVAPILSLYIRHLGQTHNLMFTSGLVVSAMGFSSLFSS 264

Query: 241 PKLGRIGDRIGNHRLLLLALLYSFLMYVLCSLAQTSLQLGVIRFLYGFGTGALMPSINSI 300
             LG++GDR GNHRLLL AL YSF+MY   +LAQTS QLGV+RF YGFG GALMPSINS+
Sbjct: 265 SYLGKLGDRFGNHRLLLAALCYSFIMYFSSALAQTSFQLGVLRFAYGFGVGALMPSINSL 324

Query: 301 LTKIAPRQGLSRIFSYNQMFSNLGQVLGPFVGSAVSIHLGFRWVFFVTSFIVLANFVWCF 360
            LTK+ P++G+SR+F+YNQMFSNLGQV+GPF+GS V++ LG+R VF+VTS IV  N +W
Sbjct: 325 LTKLTPKEGISRVFAYNQMFSNLGQVIGPFIGSNVAVVLGYRSVFYVTSLIVFVNLIWSL 384

Query: 361 INFRKYIRVKEIV 373
            I FRKYI+VK+IV
Sbjct: 385 IIFRKYIKVKDIV 397
```

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

Example 2197

A DNA sequence (GBSx2315) was identified in *S.agalactiae* <SEQ ID 6785> which encodes the amino acid sequence <SEQ ID 6786>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2343(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:AAB69986 GB:U94356 glycerol kinase [Enterococcus faecalis]
 Identities = 156/186 (83%), Positives = 167/186 (88%), Gaps = 1/186 (0%)

Query: 3   SEEKYIMAIDQGTTSSRAIIFNKKGEKIASSQKEFPQIFPQAGWVEHNANQIWNSVQSVI 62
            +EEKYIMAIDQGTTSSRAIIF+KKG KI SSQKEF Q FP AGWVEHNAN+IWNSVQSVI
Sbjct: 2   AEEKYIMAIDQGTTSSRAIIFDKKGNKIGSSQKEFTQYFPNAGWVEHNANEIWNSVQSVI 61

Query: 63  AGAFIESSIKPGQIEAIGITNQRETTVVWDKKTGLPIYNAIVWQSRQTAPIADQLKQEGH 122
            AG+ IES +KP  I  IGITNQRETTVVWDK TGLPIYNAIVWQSRQT PIADQLK++G+
Sbjct: 62  AGSLIESGVKPTDIAGIGITNQRETTVVWDKATGLPIYNAIVWQSRQTTPIADQLKEDGY 121

Query: 123 TNMIHEKTGLVIDAYFSATKVRWILDHVPGAQERAEKGELLFGTIDTWLVWKLTDGLVHV 182
            + MIHEKTGL+IDAYFSATKVRWILDHV GAQERAE GEL+FGTIDTWLVWKLT G  HV
Sbjct: 122 SEMIHEKTGLIIDAYFSATKVRWILDHVEGAQERAENGELMFGTIDTWLVWKLT-GDTHV 180

Query: 183 TDYSNA 188
            TDYSNA
Sbjct: 181 TDYSNA 186
```

There is also high homology to SEQ ID 2844:

```
Identities = 174/186 (93%), Positives = 182/186 (97%)

Query: 3   SEEKYIMAIDQGTTSSRAIIFNKKGEKIASSQKEFPQIFPQAGWVEHNANQIWNSVQSVI 62
            S+EKYIMAIDQGTTSSRAIIFN+KGEK++SSQKEFPQIFP AGWVEHNANQIWNSVQSVI
```

```
Sbjct: 2   SQEKYIMAIDQGTTSSRAIIFNQKGEKVSSSQKEFPQIFPHAGWVEHNANQIWNSVQSVI 61

Query: 63  AGAFIESSIKPGQIEAIGITNQRETTVVWDKKTGLPIYNAIVWQSRQTAPIADQLKQEGH 122
           AGAFIESSIKP QIEAIGITNQRETTVVWDKKTG+PIYNAIVWQSRQTAPIA+QLKQ+GH
Sbjct: 62  AGAFIESSIKPSQIEAIGITNQRETTVVWDKKTGVPIYNAIVWQSRQTAPIAEQLKQDGH 121

Query: 123 TNMIHEKTGLVIDAYFSATKVRWILDHVPGAQERAEKGELLFGTIDTWLVWKLTDGLVHV 182
           T MIHEKTGLVIDAYFSATK+RWILDHVPGAQERAEKGELLFGTIDTWLVWKLTDG VHV
Sbjct: 122 TKMIHEKTGLVIDAYFSATKIRWILDHVPGAQERAEKGELLFGTIDTWLVWKLTDGAVHV 181

Query: 183 TDYSNA 188
           TDYSNA
Sbjct: 182 TDYSNA 187
```

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

Example 2198

A DNA sequence (GBSx2317) was identified in *S.agalactiae* <SEQ ID 6787> which encodes the amino acid sequence <SEQ ID 6788>. This protein is predicted to be glycyl-tRNA synthetase beta chain (glyS). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2933(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:CAB14468 GB:Z99117 glycyl-tRNA synthetase (beta subunit)
        [Bacillus subtilis]
 Identities = 315/687 (45%), Positives = 447/687 (64%), Gaps = 21/687 (3%)

Query: 3   KDLLLELGLEELPAYVVTPSEKQLGQKMVKFLEDHRLSFETVQIFSTPRRLAVRVKGLAD 62
           +DLLLE+GLEE+PA  +  S  QLG K+  +L++  ++   V++F+TPRRLAV VK +A+
Sbjct: 4   QDLLLEIGLEEMPARFLNESMVQLGDKLTGWLKEKNITHGEVKLFNTPRRLAVFVKDVAE 63

Query: 63  QQTDLTEDFKGPSKKIALDAEGNFSKAAQGFVRGKGLSVDDIEFREVKGEEYVYVTKHET 122
           +Q D+ E+ KGP+KKIALDA+GN++KAA GF +G+G +V+D+   +EVKG EYV+V K +
Sbjct: 64  KQDDIKEEAKGPAKKIALDADGNWTKAAIGFSKGQGANVEDLYIKEVKGIEYVFVQKFQA 123

Query: 123 GKSAIDVLASVTEVLTELTFPVNMHWANNSFEYIRPVHTLVVLLDDQALELDFLDIHSGR 182
           G+    +L  ++ ++T L FP NM W N    YIRP+  +V L     +  ++ SGR
Sbjct: 124 GQETKSLLPELSGLITSLHFPKNMRWGNEDLRYIRPIKWIVALFGQDVIPFSITNVESGR 183

Query: 183 ISRGHRFLGSDTEISSASSYEDDLRQQFVIADAKERQQMIVNQIHAIEEKKNISVEIDED 242
            ++GHRFLG +  I S S+YE+ L+ Q VIAD   R+QMI +Q+  +  + N S+ +DED
Sbjct: 184 TTQGHRFLGHEVSIESPSAYEEQLKGQHVIADPSVRKQMIQSQLETMAAENNWSIPVDED 243

Query: 243 LLNEVLNLVEYPTAFLGSFDEKYLDVPEEVLVTSMKNHQRYFVVRDRDGKLLPNFISVRN 302
           LL+EV +LVEYPTA  GSF+ ++L +PEEVLVT+MK HQRYF V+D++G LLP+FI+VRN
Sbjct: 244 LLDEVNHLVEYPTALYGSFESEFLSIPEEVLVTTMKEHQRYFPVKDKNGDLLPHFITVRN 303

Query: 303 GNAEHIENVIKGNEKVLVARLEDGEFFWQEDQKLNIADLVEKLKQVTFHEKIGSLYEHMD 362
           GN+  IENV +GNEKVL ARL D  FF++EDQKLNI   V+KL+ + FHE++GSL + +
Sbjct: 304 GNSHAIENVARGNEKVLRARLSDASFFYKEDQKLNIDANVKKLENIVFHEELGSLADKVR 363

Query: 363 RVKVISQYLAEKADLSDEEKLAVLRAASIYKFDLLTGMVDEFDELQGIMGEKYALLAGEQ 422
           RV  I++ LA +    ++    V RAA I KFDL+T M+ EF ELQGIMGEKYA + GE
Sbjct: 364 RVTSIAEKLAVRLQADEDTLKHVKRAAEISKFDLVTHMIYEFPELQGIMGEKYARMLGED 423
```

```
Query: 423 PAVAAAIREHYMPTSADGELPETRVGAILALADKFDTLLSFFSVGLIPSGSNDPYALRRA 482
           AVAAA+ EHYMP SA GE P T  GA++A+ADK DT+ SFFS+G+IP+GS DPY L R
Sbjct: 424 EAVAAAVNEHYMPRSAGGETPSTFTGAVVAMADKLDTIASFFSIGVIPTGSQDPYGLPRQ 483

Query: 483 TQGIVRILEAFGWDIPLDELVTNLYGLSFASLDYANQKEVMAFISARIEKMIGS-KVPKD 541
             GIV IL    W I  +EL+T      F   D  N  E++ F + R++ ++ + ++  D
Sbjct: 484 ASGIVAILLDRNWGISFEELLT------FVQTDKEN--ELLDFFTQRLKYVLNAEQIRHD 535

Query: 542 IREAVLESDTYIVSLILEASQALVQKSKDAQYKVSVESLSRAFNLAEKVTHSVLVDSSLF 601
           + +AVLES        L  +Q L QK     +K + E+L R  ++++K      +   LF
Sbjct: 536 VIDAVLESSELEPYSALHKAQVLEQKLGAPGFKETAEALGRVISISKKGVRGD-IQPDLF 594

Query: 602 ENNQEKALYQAILSLELTEDMHDNLDK---------LFALSPIINDFFDNTMVMTDDEKM 652
           EN  E  L+ A  + +  E++ +N  K          L AL   I+ +FD+TMV+ D+E +
Sbjct: 595 ENEYEAKLFDAYQTAK--ENLQENFSKKDYEAALASLAALKEPIDAYFDHTMVIADNESL 652

Query: 653 KQNRLAILNSLVAKARTVAAFNLLNTK 679
           K NRLA + SL  + ++ A  N L  K
Sbjct: 653 KANRLAQMVSLADEIKSFANMNALIVK 679
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2835> which encodes the amino acid sequence <SEQ ID 2836>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.96   Transmembrane  450 - 466 ( 450 - 466)

----- 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>
```

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

```
 Identities = 505/679 (74%), Positives = 578/679 (84%)

Query: 1   MTKDLLLELGLEELPAYVVTPSEKQLGQKMVKFLEDHRLSFETVQIFSTPRRLAVRVKGL 60
           M+K+LL+ELGLEELPAYVVTPSEKQLG+++   FL ++RLSFE +Q FSTPRRLAVRV GL
Sbjct: 1   MSKNLLIELGLEELPAYVVTPSEKQLGERLATFLTENRLSFEDIQTFSTPRRLAVRVSGL 60

Query: 61  ADQQTDLTEDFKGPSKKIALDAEGNFSKAAQGFVRGKGLSVDDIEFREVKGEEYVYVTKH 120
           ADQQTDLTEDFKGP+KKIALDA+GNFSKAAQGFVRGKGL+ D IEFREVKGEEYVYVTKH
Sbjct: 61  ADQQTDLTEDFKGPAKKIALDADGNFSKAAQGFVRGKGLTTDAIEFREVKGEEYVYVTKH 120

Query: 121 ETGKSAIDVLASVTEVLTELTFPVNMHWANNSFEYIRPVHTLVVLLDDQALELDFLDIHS 180
           E GK A +VL  VTEVL+ +TFPV+MHWANNSFEYIRPVHTL VLL+D+ALELDFLDIHS
Sbjct: 121 EAGKPAKEVLLGVTEVLSAMTFPVSMHWANNSFEYIRPVHTLTVLLNDEALELDFLDIHS 180

Query: 181 GRISRGHRFLGSDTEISSASSYEDDLRQQFVIADAKERQQMIVNQIHAIEEKKNISVEID 240
           GR+SRGHRFLG++T I+SA SYE DLR QFVIADAKERQ+MIV QI  +E ++ + V+ID
Sbjct: 181 GRVSRGHRFLGTETTITSADSYEADLRSQFVIADAKERQEMIVEQIKTLEVEQGVQVDID 240

Query: 241 EDLLNEVLNLVEYPTAFLGSFDEKYLDVPEEVLVTSMKNHQRYFVVRDRDGKLLPNFISV 300
           EDLLNEVLNLVE+PTAF+GSF+ KYLDVPEEVLVTSMKNHQRYFVVRD+ G L+PNF+SV
Sbjct: 241 EDLLNEVLNLVEFPTAFMGSFEAKYLDVPEEVLVTSMKNHQRYFVVRDQAGHLMPNFVSV 300

Query: 301 RNGNAEHIENVIKGNEKVLVARLEDGEFFWQEDQKLNIADLVEKLKQVTFHEKIGSLYEH 360
           RNGN + IENVIKGNEKVLVARLEDGEFFW+EDQKL IADLV KL  VTFHEKIGSL EH
Sbjct: 301 RNGNDQAIENVIKGNEKVLVARLEDGEFFWREDQKLQIADLVAKLTNVTFHEKIGSLAEH 360

Query: 361 MDRVKVISQYLAEKADLSDEEKLAVLRAASIYKFDLLTGMVDEFDELQGIMGEKYALLAG 420
           MDR +VI+  LA++A+LS EE  AV RAA IYKFDLLTGMV EFDELQGIMGEKYALLAG
Sbjct: 361 MDRTRVIAASLAKEANLSAEEVTAVDRAAQIYKFDLLTGMVGEFDELQGIMGEKYALLAG 420

Query: 421 EQPAVAAAIREHYMPTSADGELPETRVGAILALADKFDTLLSFFSVGLIPSGSNDPYALR 480
           E  AVA AIREHY+P +A G LPET+VGA+LALA K DTLLSFFSVGLIPSGSNDPYALR
Sbjct: 421 EDAAVATAIREHYLPDAAGGALPETKVGAVLALAAKLDTLLSFFSVGLIPSGSNDPYALR 480
```

```
Query: 481 RATQGIVRILEAFGWDIPLDELVTNLYGLSFASLDYANQKEVMAFISARIEKMIGSKVPK 540
           RATQGIVRIL+ FGW IP+D+LV +LY LSF SL YAN+ +VM FI AR++KM+G   PK
Sbjct: 481 RATQGIVRILDHFGWRIPMDKLVDSLYDLSFDSLTYANKADVMNFIRARVDKMMGKAAPK 540

Query: 541 DIREAVLESDTYIVSLILEASQALVQKSKDAQYKVSVESLSRAFNLAEKVTHSVLVDSSL 600
           DIREA+L S T++V  +L A++ALV+ S     YK +VESLSRAFNLAEK   SV VD SL
Sbjct: 541 DIREAILASSTFVVPEMLAAAEALVKASHTENYKPAVESLSRAFNLAEKADASVQVDPSL 600

Query: 601 FENNQEKALYQAILSLELTEDMHDNLDKLFALSPIINDFFDNTMVMTDDEKMKQNRLAIL 660
           FEN QE  L+ AI  L L       L+++FALSP+INDFFDNTMVM  D+ +K NRLAIL
Sbjct: 601 FENEQENTLFAAIQGLTLAGSAAQQLEQVFALSPVINDFFDNTMVMAGDQALKNNRLAIL 660

Query: 661 NSLVAKARTVAAFNLLNTK 679
           + LV+KA+T+ AFN LNTK
Sbjct: 661 SDLVSKAKTIVAFNQLNTK 679
```

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

Example 2199

A DNA sequence (GBSx2318) was identified in *S.agalactiae* <SEQ ID 6789> which encodes the amino acid sequence <SEQ ID 6790>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2182(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:AAD24436 GB:AF112858 NAD(P)H dehydrogenase [Bacillus
          stearothermophilus]
 Identities = 64/174 (36%), Positives = 98/174 (55%), Gaps = 6/174 (3%)

Query: 2   NTLIVNSHPDFSNPYSFTTILQEKFIELYNEHFPNHQLSILNLYDCVLPEITKEVLLSIW 61
           N L + +HP   +  S++  + + FI+ Y +  P+H++  L+LY   +PEI  +V  S W
Sbjct: 3   NVLYITAHPH-DDTQSYSMAVGKAFIDTYKQVHPDHEVIHLDLYKEYIPEIDVDVF-SGW 60

Query: 62  SKQRKGL---ELTADEIVQAKISKDLLEQFKSHHRIVFVSPMHNYNVTARAKTYIDNIFI 118
            K R G    EL+ +E  +     +L EQF S  + VFV+PM N++     K YID + +
Sbjct: 61  GKLRSGKSFEELSDEEKAKVGRMNELCEQFISADKYVFVTPMWNFSFPPVLKAYIDAVAV 120

Query: 119 AGETFKYTENGSVGLMTDDYRLLMLESAGSIYSKGQYSPYEFPVHYLKAIFKDF 172
           AG+TFKYTE G VGL+TD  + L +++ G  YS+G  +  E    YL  I + F
Sbjct: 121 AGKTFKYTEQGPVGLLTDK-KALHIQARGGFYSEGPAAEMEMGHRYLSVIMQFF 173
```

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 2200

A DNA sequence (GBSx2319) was identified in *S.agalactiae* <SEQ ID 6791> which encodes the amino acid sequence <SEQ ID 6792>. This protein is predicted to be glycyl-tRNA synthetase (glyQ). Analysis of this protein sequence reveals the following:

```
Possible site: 56
```

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1364(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 9521> which encodes amino acid sequence <SEQ ID 9522> was also identified.

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

```
>GP:BAB05089 GB:AP001511 glycyl-tRNA synthetase (alpha subunit)
          [Bacillus halodurans]
 Identities = 222/287 (77%), Positives = 250/287 (86%)

Query: 6   LTFQEIILTLQQFWNDQGCMLMQAYDNEKGAGTMSPYTFLRAIGPEPWNAAYVEPSRRPA 65
            +  Q +ILTLQ++W+ Q C+L+QAYD EKGAGTMSPYT LR IGPEPWN AYVEPSRRPA
Sbjct: 1   MNVQTMILTLQEYWSKQNCILLQAYDTEKGAGTMSPYTMLRTIGPEPWNVAYVEPSRRPA 60

Query: 66  DGRYGENPNRLYQHHQFQVVMKPSPSNIQELYLKSLELLGINPLEHDIRFVEDNWENPST 125
            DGRYGENPNRLYQHHQFQV+MKPSP+NIQELYL SL  LGINPLEHDIRFVEDNWENPS
Sbjct: 61  DGRYGENPNRLYQHHQFQVIMKPSPTNIQELYLDSLRALGINPLEHDIRFVEDNWENPSL 120

Query: 126 GSAGLGWEVWLDGMEITQFTYFQQVGGLQTGPVTSEVTYGLERLASYIQEVDSVYDIEWA 185
            G AGLGWEVWLDGMEITQFTYFQQVGGL+  PV++E+TYGLERLASYIQ+ ++V+D+EW
Sbjct: 121 GCAGLGWEVWLDGMEITQFTYFQQVGGLEANPVSAEITYGLERLASYIQDKENVFDLEWV 180

Query: 186 PGVKYGEIFTQPEYEHSKYSFEISDQVMLLENFEKFEREAKRALEEGLVHPAYDYVLKCS 245
             G  YG+IFTQPEYEHSKY+FE+SD  ML E F  +E+EA RALEE LV PAYDYVLKCS
Sbjct: 181 EGFTYGDIFTQPEYEHSKYTFEVSDSAMLFELFSTYEKEADRALEENLVFPAYDYVLKCS 240

Query: 246 HTFNLLDARGAVSVTERAGYIARIRNLARVVAKTFVAERKKLGFPLL 292
            HTFNLLDARGA+SVTER GYI R+RNLAR  AK +  ER+KLGFP+L
Sbjct: 241 HTFNLLDARGAISVTERTGYIGRVRNLARKCAKKYYEEREKLGFPML 287
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6793> which encodes the amino acid sequence <SEQ ID 6794>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2081(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 = 290/304 (95%), Positives = 294/304 (96%)

Query: 2   MSKKLTFQEIILTLQQFWNDQGCMLMQAYDNEKGAGTMSPYTFLRAIGPEPWNAAYVEPS 61
            MSKKLTFQEIILTLQQ+WNDQGCMLMQAYDNEKGAGTMSPYTFLRAIGPEPWNAAYVEPS
Sbjct: 1   MSKKLTFQEIILTLQQYWNDQGCMLMQAYDNEKGAGTMSPYTFLRAIGPEPWNAAYVEPS 60

Query: 62  RRPADGRYGENPNRLYQHHQFQVVMKPSPSNIQELYLKSLELLGINPLEHDIRFVEDNWE 121
            RRPADGRYGENPNRLYQHHQFQVVMKPSPSNIQELYL SLE LGINPLEHDIRFVEDNWE
Sbjct: 61  RRPADGRYGENPNRLYQHHQFQVVMKPSPSNIQELYLASLEKLGINPLEHDIRFVEDNWE 120

Query: 122 NPSTGSAGLGWEVWLDGMEITQFTYFQQVGGLQTGPVTSEVTYGLERLASYIQEVDSVYD 181
            NPSTGSAGLGWEVWLDGMEITQFTYFQQVGGL T PVT+EVTYGLERLASYIQEVDSVYD
Sbjct: 121 NPSTGSAGLGWEVWLDGMEITQFTYFQQVGGLATSPVTAEVTYGLERLASYIQEVDSVYD 180

Query: 182 IEWAPGVKYGEIFTQPEYEHSKYSFEISDQVMLLENFEKFEREAKRALEEGLVHPAYDYV 241
            IEWAPGVKYGEIF QPEYEHSKYSFEISDQ MLLENFEKFE+EA RALEEGLVHPAYDYV
Sbjct: 181 IEWAPGVKYGEIFLQPEYEHSKYSFEISDQDMLLENFEKFEKEASRALEEGLVHPAYDYV 240
```

```
Query: 242 LKCSHTFNLLDARGAVSVTERAGYIARIRNLARVVAKTFVAERKKLGFPLLDEETRIKLL 301
           LKCSHTFNLLDARGAVSVTERAGYIARIRNLARVVAKTFVAERKKLGFPLLDE TR  LL
Sbjct: 241 LKCSHTFNLLDARGAVSVTERAGYIARIRNLARVVAKTFVAERKKLGFPLLDEATRAILL 300

Query: 302 AEED 305
           AE+D
Sbjct: 301 AEDD 304
```

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

Example 2201

A DNA sequence (GBSx2320) was identified in *S.agalactiae* <SEQ ID 6795> which encodes the amino acid sequence <SEQ ID 6796>. This protein is predicted to be vacB protein (vacB). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2966(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 9399> which encodes amino acid sequence <SEQ ID 9400> was also identified.

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

```
>GP:CAB15366 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 338/780 (43%), Positives = 485/780 (61%), Gaps = 47/780 (6%)

Query: 4   AKAFPKLIKTISNLESHRQL---RFDDNGSLSLQKKEAKKKEITVRGLFRANKAGFGFL- 59
           A+ F +L+K +  LE    +   R D  G          +K   ++G   A+  GF FL
Sbjct: 36  AEEFKELVKALVALEEKGLIVRTRSDRYG--------IPEKMNLIKGKISAHAKGFAFLL 87

Query: 60  SIDQDEDDMFIGKNDIAYAIDGDTVEAVVKKPADRLNGTAAEARVVNIVERSLKTLVGKF 119
             D    D+FI  N++  A++GD V   +   +G+  E  V+ I+ER+++ +VG +
Sbjct: 88  PEDTSLSDVFIPPNELNTAMNGDIVMVRLNSQS---SGSRQEGTVIRILERAIQRVVGTY 144

Query: 120 VLDDERPKYAGYIKSKNQKINQKIYIRKEPV--VLDGTEIIKVDIDKYPTRGHDYFVASV 177
                + G++   ++KI   I+I K       +G +++ V +  YP  G       V
Sbjct: 145 T----ETRNFGFVIPDDKKITSDIFIPKNGKNGAAEGHKVV-VKLTSYP-EGRMNAEGEV 198

Query: 178 RDIVGHQGDVGIDVLEVLESMDIVSEFPEDVIAEANAIPDAPTEKDLIGRVDLRQEVTFT 237
             I+GH+ D GID+L V+    +  EFP D + +A++ PD   EKDL  R DLR +V  T
Sbjct: 199 ETILGHKNDPGIDILSVIHKHGLPGEFPADAMEQASSTPDTIDEKDLKDRRDLRDQVIVT 258

Query: 238 IDGADAKDLDDAVHIKLLDNGHFELGVHIADVSYYVTEGSALNREALSRGTSVYVTDRVV 297
           IDGADAKDLDDAV +  LD+G ++LGVHIADVS+YVTE S +++EAL RGTSVY+ DRV+
Sbjct: 259 IDGADAKDLDDAVTVTKLDDGSYKLGVHIADVSHYVTENSPIDKEALERGTSVYLVDRVI 318

Query: 298 PMLPERLSNGICSLNPNLDRLTQSCIMEIDQNGRVVNHQITQSVINTTYRMTYTAVNDII 357
           PM+P RLSNGICSLNP +DRLT SC M I+  G+V  H+I QSVI TT RMTY+ VN I+
Sbjct: 319 PMIPHRLSNGICSLNPKVDRLTLSCEMTINSQGQVTEHEIFQSVIKTTERMTYSDVNKIL 378

Query: 358 A-GDEEICSEYESIVSSVQHMVTLHHTLEAMRTRRGALNFDTSEAKIMVNDKGMPVDIVI 416
             DEE+  +YE +V   + M  L   L   R  RGA++FD  EAK++V+D+G   D+VI
Sbjct: 379 VDDDEELKQKYEPLVPMFKDMERLAQILRDKRMDRGAVDFDFKEAKVLVDDEGAVKDVVI 438

Query: 417 RNRGIAERMIESFMLAANETVAEHYARLKLPFIYRIHEEPKAEKLQKFIDYASVFGVQIQ 476
           R R +AE++IE FML ANETVAEH+  + +PFIYRIHEEP AEKLQKF+++ + FG  ++
```

```
Sbjct: 439 RERSVAEKLIEEFMLVANETVAEHFHWMNVPFIYRIHEEPNAEKLQKFLEFVTTFGYVVK 498

Query: 477 GTATKITQSALQDFMKKVQGQPGSEVLSMMLLRSMQQARYSEHNHGHYGLAAEYYTHFTS 536
           GTA  I   ALQ  +  V+ +P . V+S ++LRSM+QA+Y   + GH+GL+ E+YTHFTS
Sbjct: 499 GTAGNIHPRALQSILDAVRDRPEETVISTVMLRSMKQAKYDPQSLGHFGLSTEFYTHFTS 558

Query: 537 PIRRYPDLLVHRMIRDY-DDKAMDKA--DHFANLIPEIATQTSSLERRAIDAERIVEAMK 593
           PIRRYPDL+VHR+IR Y  +  +D+A  + +A  +P+IA  TSS+ERRA+DAER  + +K
Sbjct: 559 PIRRYPDLIVHRLIRTYLINGKVDEATQEKWAERLPDIAEHTSSMERRAVDAERETDDLK 618

Query: 594 KAEYMEEYVGEEFEGVVASVVKFGMFVELPNTIEGLIHVTTL-PEYYHFNERTLTLQGEK 652
           KAEYM + +GEEF+G+++SV  FGMFVELPNTIEGL+HV+ +  +YY F+E+   + GE+
Sbjct: 619 KAEYMLDKIGEEFDGMISSVTNFGMFVELPNTIEGLVHVSFMTDDYYRFDEQHFAMIGER 678

Query: 653 SGKVFRVGQQIKVKLIRSDKETGDIDFDYLPSDFDIVEKVSKSSREGRPNRSSKREHQHR 712
           +G VFR+G +I VK++  +K+  +IDF+ +                +G P R  + +
Sbjct: 679 TGNVFRIGDEITVKVVDVNKDERNIDFEIV-------------GMKGTPRRPRELD---- 721

Query: 713 ISDRDNKNKNTSKKKASRKPKRNSDSKSHHHKDDRTTGSTKKKTKKPFYKGVAKKGQKRK 772
            S R  K    ++K++       S + S   K +  T   KKK K+ F     +K +K+K
Sbjct: 722 -SSRSRKRGKPARKRVQSTNTPVSPAPS-EEKGEWFTKPKKKKKKRGFQNAPKQKRKKKK 779
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6797> which encodes the amino acid sequence <SEQ ID 6798>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0811(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 = 579/773 (74%), Positives = 664/773 (84%), Gaps = 22/773 (2%)

Query: 1   MAGAKAFPKLIKTISNLESHRQLRFDDNGSLSLQKKEAKKKEITVRGLFRANKAGFGFLS 60
           MAGAK FP LIKTIS +ES   LRF D+GSL+L+K+  KKKE TV+G+FRANKAGFGFL
Sbjct: 27  MAGAKHFPSLIKTISKMESQSLLRFSDDGSLALRKEREKKKEPTVQGVFRANKAGFGFLH 86

Query: 61  IDQDEDDMFIGKNDIAYAIDGDTVEAVVKKPADRLNGTAAEARVVNIVERSLKTLVGKFV 120
           +D++EDDMFIG+ND+ YAIDGDTVE VVKKPADRL GTAAEA+VV IV+RSLKT VG F+
Sbjct: 87  VDENEDDMFIGRNDVGYAIDGDTVEVVVKKPADRLKGTAAEAKVVAIVDRSLKTAVGTFI 146

Query: 121 LDDERPKYAGYIKSKNQKINQKIYIRKEPVVLDGTEIIKVDIDKYPTRGHDYFVASVRDI 180
           LDD++PKYAGYI+SKNQKI QKIYI+KEPVVL GTEIIKVDIDKYP RGHDYFVASVRDI
Sbjct: 147 LDDDKPKYAGYIRSKNQKIQQKIYIKKEPVVLKGTEIIKVDIDKYPIRGHDYFVASVRDI 206

Query: 181 VGHQGDVGIDVLEVLESMDIVSEFPEDVIAEANAIPDAPTEKDLIGRVDLRQEVTFTIDG 240
           VGHQGDVGIDVLEVLESMDIVSEFP +V+AEANAI +APT KDLIGRVDLRQE T TIDG
Sbjct: 207 VGHQGDVGIDVLEVLESMDIVSEFPAEVLAEANAISEAPTAKDLIGRVDLRQETTITIDG 266

Query: 241 ADAKDLDDAVHIKLLDNGHFELGVHIADVSYYVTEGSALNREALSRGTSVYVTDRVVPML 300
           ADAKDLDDA+HIKLLDNG++ELGVHIADVSYYVTEGSAL++EA++RGTSVYVTDRVVPML
Sbjct: 267 ADAKDLDDAIHIKLLDNGNYELGVHIADVSYYVTEGSALDKEAIARGTSVYVTDRVVPML 326

Query: 301 PERLSNGICSLNPNLDRLTQSCIMEIDQNGRVVNHQITQSVINTTYRMTYTAVNDIIAGD 360
           PERLSNGICSLNPN+DRLTQS +MEI+  G VVN+QI QSVI TTYRMTY+ VND+IAGD
Sbjct: 327 PERLSNGICSLNPNIDRLTQSALMEINSQGHVVNYQICQSVIKTTYRMTYSTVNDMIAGD 386

Query: 361 EEICSEYESIVSSVQHMVTLHHTLEAMRTRRGALNFDTSEAKIMVNDKGMPVDIVIRNRG 420
           EE   E+ SI   V  MV LH  LEAMR++RGALNFDT EAKI+VNDKGMPVD+V+R RG
Sbjct: 387 EEALQEFASIADDVTLMVALHRILEAMRSKRGALNFDTQEAKIIVNDKGMPVDVVLRQRG 446

Query: 421 IAERMIESFMLAANETVAEHYARLKLPFIYRIHEEPKAEKLQKFIDYASVFGVQIQGTAT 480
           IAERMIESFMLAANE VAEH+A+ KLPFIYRIHEEPKAEKLQ+FIDYAS FG+ IQGTA
```

```
Sbjct: 447 IAERMIESFMLAANECVAEHFAKAKLPFIYRIHEEPKAEKLQQFIDYASTFGIHIQGTAN 506

Query: 481 KITQSALQDFMKKVQGQPGSEVLSMMLLRSMQQARYSEHNHGHYGLAAEYYTHFTSPIRR 540
           KI+Q ALQ FM KV+GQPG+EVL+MMLLRSMQQARYSEHNHGHYGLAAEYYTHFTSPIRR
Sbjct: 507 KISQEALQAFMAKVEGQPGAEVLNMMLLRSMQQARYSEHNHGHYGLAAEYYTHFTSPIRR 566

Query: 541 YPDLLVHRMIRDYDDKAMDKADHFANLIPEIATQTSSLERRAIDAERIVEAMKKAEYMEE 600
           YPDLLVHRM+R+Y+  + +K DHFA +IPE+AT +S LERRAIDAER+VEAMKKAEYM E
Sbjct: 567 YPDLLVHRMVREYNQPSQEKRDHFAQIIPELATSSSQLERRAIDAERVVEAMKKAEYMAE 626

Query: 601 YVGEEFEGVVASVVKFGMFVELPNTIEGLIHVTTLPEYYHFNERTLTLQGEKSGKVFRVG 660
           YVGEEF+G+V+SVVKFG FVELPNTIEGL+H+T+LPEYYHFNERTL+LQGEKSGKVF+VG
Sbjct: 627 YVGEEFDGIVSSVVKFGFFVELPNTIEGLVHITSLPEYYHFNERTLSLQGEKSGKVFKVG 686

Query: 661 QQIKVKLIRSDKETGDIDFDYLPSDFDIVEKVSKSSREGRPNRSSKREHQHRISDRDNKN 720
           Q I+VKL+++DKETGDIDF+YLPSDFD+VEK+  S +  R +R              K+
Sbjct: 687 QPIRVKLVKADKETGDIDFEYLPSDFDVVEKIKMSDKASRRDR--------------RKS 732

Query: 721 KNTSKKKASRKPKRNSDSKSHHHKDDRTTGSTKKKTKKPFYKGVAKKGQKRKS 773
             +SK    ++PK  + +K        T G TKK +KKPFYK  AKK  +++S
Sbjct: 733 KSSSKGTKKKEPKEVAKAK--------TKGKTKKGSKKPFYKEQAKKKSRKRS 777
```

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

Example 2202

A DNA sequence (GBSx2321) was identified in *S.agalactiae* <SEQ ID 6799> which encodes the amino acid sequence <SEQ ID 6800>. This protein is predicted to be VacB homolog (smpB). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2988(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:AAC23745 GB:AF052209 VacB homolog [Streptococcus pneumoniae]
 Identities = 121/155 (78%), Positives = 139/155 (89%)

Query: 1   MVKGQGNVVAQNKKAHHDYTIVETIEAGIVLTGTEIKSVRAARITLKDGYAQIKNGEAWL 60
           M KG+G VVAQNKKA HDYTIV+T+EAG+VLTGTEIKSVRAARI LKDG+AQ+KNGE WL
Sbjct: 1   MAKGEGKVVAQNKKARHDYTIVDTLEAGMVLTGTEIKSVRAARINLKDGFAQVKNGEVWL 60

Query: 61  INVHITPYDQGNIWNQDPDRTRKLLLKKREIEKISNELKGTGMTLVPLKVYLKDGFAKVL 120
            NVHI PY++GNIWNQ+P+R RKLLL K++I+K+  E KGTGMTLVPLKVY+KDG+AK+L
Sbjct: 61  SNVHIAPYEEGNIWNQEPERRRKLLLHKKQIQKLEQETKGTGMTLVPLKVYIKDGYAKLL 120

Query: 121 LGLAKGKHDYDKRESIKRREQNRDIARQLKNYNSR 155
           LGLAKGKHDYDKRESIKRREQNRDIAR +K  N R
Sbjct: 121 LGLAKGKHDYDKRESIKRREQNRDIARVMKAVNQR 155
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6801> which encodes the amino acid sequence <SEQ ID 6802>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2918(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 = 124/155 (80%), Positives = 145/155 (93%)

Query: 1   MVKGQGNVVAQNKKAHHDYTIVETIEAGIVLTGTEIKSVRAARITLKDGYAQIKNGEAWL 60
           M KG+G+++AQNKKA HDY IVET+EAGIVLTGTEIKSVRAARI LKDG+AQIKNGEAWL
Sbjct: 1   MAKGEGHILAQNKKARHDYHIVETVEAGIVLTGTEIKSVRAARIQLKDGFAQIKNGEAWL 60

Query: 61  INVHITPYDQGNIWNQDPDRTRKLLLKKREIEKISNELKGTGMTLVPLKVYLKDGFAKVL 120
           +NVHI P++QGNIWN DP+RTRKLLLKKREI  ++NELKG+GMTLVPLKVYLKDGFAKVL
Sbjct: 61  VNVHIAPFEQGNIWNADPERTRKLLLKKREITHLANELKGSGMTLVPLKVYLKDGFAKVL 120

Query: 121 LGLAKGKHDYDKRESIKRREQNRDIARQLKNYNSR 155
           +GLAKGKH+YDKRE+IKRR+Q RDI +Q+K+YN+R
Sbjct: 121 IGLAKGKHEYDKRETIKRRDQERDIKKQMKHYNAR 155
```

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

Example 2203

A DNA sequence (GBSx2322) was identified in *S.agalactiae* <SEQ ID 6803> which encodes the amino acid sequence <SEQ ID 6804>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.6876(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 2204

A DNA sequence (GBSx2323) was identified in *S.agalactiae* <SEQ ID 6805> which encodes the amino acid sequence <SEQ ID 6806>. This protein is predicted to be d-serine/d-alanine/glycine transporter (cycA). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.02   Transmembrane   71 -  87 (  62 -  90)
   INTEGRAL    Likelihood = -8.92   Transmembrane  320 - 336 ( 316 - 344)
   INTEGRAL    Likelihood = -8.33   Transmembrane  254 - 270 ( 251 - 275)
   INTEGRAL    Likelihood = -6.00   Transmembrane  158 - 174 ( 154 - 175)
   INTEGRAL    Likelihood = -2.76   Transmembrane  197 - 213 ( 196 - 213)
   INTEGRAL    Likelihood = -2.50   Transmembrane  117 - 133 ( 116 - 136)
   INTEGRAL    Likelihood = -1.38   Transmembrane  282 - 298 ( 279 - 298)
   INTEGRAL    Likelihood = -0.32   Transmembrane  342 - 358 ( 342 - 360)

----- Final Results -----
              bacterial membrane --- Certainty=0.4609(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 9397> which encodes amino acid sequence <SEQ ID 9398> was also identified.

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

```
>GP:CAB14651 GB:Z99117 amino acid permease [Bacillus subtilis]
 Identities = 165/361 (45%), Positives = 227/361 (62%), Gaps = 17/361 (4%)

Query: 1   MGIFLT-LSYWISLIFIGMAEITAVGEYVQFWFPEWPSWIIQIVFLAILSSINLIAVKAF 59
            M  F+T  +YW   I + MA++TAVG Y Q+W P+ P W+  ++ L IL  +NL  VK F
Sbjct: 95  MAAFITGWTYWFCWISLAMADLTAVGIYTQYWLPDVPQWLPGLLALIILLIMNLATVKLF 154

Query: 60  GETEFWFAMIKVIAILGLIATGIFMVLTNFDTGHGYHASISNITNHFEWFPKGKLNFFMA 119
            GE EFWFA+IKVIAIL LI TGI ++   F    G  AS++N+ +H   FP G   F ++
Sbjct: 155 GELEFWFALIKVIAILALIVTGILLIAKGFSAASG-PASLNNLWSHGGMFPNGWHGFILS 213

Query: 120 FQMVFFAYLAIEFVGVTTSETANPRKVLPKAIQEIPMRIILFYAGSLLAIMAIFPWQQLP 179
            FQMV FA++ IE VG+T  ET NP+KV+PKAI +IP+RI+LFY G+L  IM I+PW  L
Sbjct: 214 FQMVVFAFVGIELVGLTAGETENPQKVIPKAINQIPVRILLFYVGALFVIMCIYPWNVLN 273

Query: 180 VNESPFVTVFKLAGIKWAAALINFVVLTSAASALNSTLYSTGRHLFQLANE--SPNALTK 237
            NESPFV VF   GI  AA+LINFVVLTSAASA NS L+ST R ++ LA +  +P  L K
Sbjct: 274 PNESPFVQVFSAVGIVVAASLINFVVLTSAASAANSALFSTSRMVYSLAKDHHAPGLLKK 333

Query: 238 ALKLDQLSRQSVPSRAIIAS--AVIVGASALISVLPGISDAFSLITASSSGVYISIYVLI 295
                L+  +VPS A+  S  A+++G S L  ++P      F+LIT+ S+  +I I+ +
Sbjct: 334 ------LTSSNVPSNALFFSSIAILIGVS-LNYLMP--EQVFTLITSVSTICFIFIWGIT 384

Query: 296 MIAHWKYRKS--PDFMEDGYKMPAYKILSPITLLFFLFVFVSLFLQDSTYIGAIGATIWII 354
            +I H KYRK+   +   + +KMP Y + + +TL F  F+ V L L + T I      +W +
Sbjct: 385 VICHLKYRKTRQHEAKANKFKMPFYPLSNYLTLAFLAFILVILALANDTRIALFVTPVWFV 445
```

There is also homology to SEQ ID 4070:

```
Identities = 286/364 (78%), Positives = 322/364 (87%), Gaps = 1/364 (0%)

Query: 2   GIFLTLSYWISLIFIGMAEITAVGEYVQFWFPEWPSWIIQIVFLAILSSINLIAVKAFGE 61
            G F  LSYWISLIFIGMAEITAVG YVQFWFP WP+W+IQ+VFL +LSSINLIAV+ FGE
Sbjct: 101 GYFSGLSYWISLIFIGMAEITAVGAYVQFWFPSWPAWLIQLVFLVLLSSINLIAVRVFGE 160

Query: 62  TEFWFAMIKVIAILGLIATGIFMVLTNFDTGHGYHASISNITNHFEWFPKGKLNFFMAFQ 121
            TEFWFAMIK++AIL LIAT IFMVLT F+T H  HAS+SNI +HF  FP GKL FFMAFQ
Sbjct: 161 TEFWFAMIKILAILALIATAIFMVLTGFET-HTGHASLSNIFDHFSMFPNGKLKFFMAFQ 219

Query: 122 MVFFAYLAIEFVGVTTSETANPRKVLPKAIQEIPMRIILFYAGSLLAIMAIFPWQQLPVN 181
            MVFFAY AIEFVG+TTSETANPRKVLPKAIQEIP RI++FY G+L++IMAI PW QLPV+
Sbjct: 220 MVFFAYQAIEFVGITTSETANPRKVLPKAIQEIPTRIVIFYVGALVSIMAIVPWHQLPVD 279

Query: 182 ESPFVTVFKLAGIKWAAALINFVVLTSAASALNSTLYSTGRHLFQLANESPNALTKALKL 241
            ESPFV VFKL GIKWAAALINFVVLTSAASALNSTLYSTGRHL+Q+ANE+PNALT  LK+
Sbjct: 280 ESPFVMVFKLIGIKWAAALINFVVLTSAASALNSTLYSTGRHLYQIANETPNALTNRLKI 339

Query: 242 DQLSRQSVPSRAIIASAVIVGASALISVLPGISDAFSLITASSSGVYISIYVLIMIAHWK 301
            + LSRQ VPSRAIIASAV+VG SALI++LPG++DAFSLITASSSGVYI+IY L MIAHWK
Sbjct: 340 NTLSRQGVPSRAIIASAVVVGISALINILPGVADAFSLITASSSGVYIAIYALTMIAHWK 399

Query: 302 YRKSPDFMEDGYKMPAYKILSPITLLFFLFVFVSLFLQDSTYIGAIGATIWIIGFGLYSH 361
            YR+S DFM DGY MP YK+ +P+TL FF FVF+SLFLQ+STYIGAIGATIWII FG+YS+
Sbjct: 400 YRQSKDFMADGYLMPKYKVTTPLTLAFFAFVFISLFLQESTYIGAIGATIWIIIFGIYSN 459

Query: 362 FKHK 365
            K K
Sbjct: 460 VKFK 463
```

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

Example 2205

A DNA sequence (GBSx2324) was identified in *S.agalactiae* <SEQ ID 6807> which encodes the amino acid sequence <SEQ ID 6808>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.33   Transmembrane  194 - 210 ( 191 - 215)
   INTEGRAL    Likelihood = -5.47   Transmembrane   17 -  33 (  14 -  38)
   INTEGRAL    Likelihood = -5.15   Transmembrane  125 - 141 ( 119 - 144)
   INTEGRAL    Likelihood = -3.88   Transmembrane  155 - 171 ( 153 - 176)
   INTEGRAL    Likelihood = -1.38   Transmembrane   96 - 112 (  94 - 114)
   INTEGRAL    Likelihood = -0.43   Transmembrane   49 -  65 (  49 -  65)

----- 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 homology with the following sequences in the GENPEPT database.

```
>GP:AAC95438 GB:AF068901 unknown [Streptococcus pneumoniae]
 Identities = 80/214 (37%), Positives = 122/214 (56%), Gaps = 3/214 (1%)

Query: 4   FFSNIRTEIPQMPLLIHSLILSVLPFLMWLTLVNRDKPLYKTIWSILLGLQLITIYTWFF 63
           FF+   T+ P+  L  +  + ++L    +     R+K +Y+  + IL  +QLI +Y W++
Sbjct: 7   FFTTQATKPPKFDLFWYVSLFTLLALTFYTAHRYREKKVYQRFFQILQTVQLILLYGWYW 66

Query: 64  WAKLPLSESLPLYHCRIGMFVVLLARPGI--LKDYFALLGVVGGVLAMIHPDFYPYQFLH 121
              +PLSESLP YHCR+ MFVVLL  PG    K YFALLG  G + A ++P    Y F H
Sbjct: 67  VNHMPLSESLPFYHCRMAMFVVLLL-PGQSKYKQYFALLGTFGTLAAFVYPVPDAYPFPH 125

Query: 122 VTNIFFFIGHFALFVLSLLHLMTQSNLDKLNPKLIIQLTLLINMSLIFINLLTGGNYGFM 181
           +T + F  GH AL   SL++L+ Q N   L+ K I  +T  +N  +  +NL+TGG+YGF+
Sbjct: 126 ITILSFIFGHLALLGNSLVYLLRQYNARLLDVKGIFLMTFALNALIFVVNLVTGGDYGFL 185

Query: 182 MKTPILGITNPFLNLFIVTTLLSFLVLFVKQIFQ 215
            K P++G      N  +V+ +L    +   K+I +
Sbjct: 186 TKPPLVGDHGLVANYLLVSIVLVATISLTKKILE 219
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6809> which encodes the amino acid sequence <SEQ ID 6810>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.25   Transmembrane   16 -  32 (  11 -  39)
   INTEGRAL    Likelihood = -3.45   Transmembrane  154 - 170 ( 153 - 173)
   INTEGRAL    Likelihood = -3.08   Transmembrane   96 - 112 (  94 - 112)
   INTEGRAL    Likelihood = -1.91   Transmembrane  191 - 207 ( 191 - 209)
   INTEGRAL    Likelihood = -1.12   Transmembrane   71 -  87 (  71 -  87)

----- 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:

```
>GP:AAC95438 GB:AF068901 unknown [Streptococcus pneumoniae]
 Identities = 90/231 (38%), Positives = 128/231 (54%), Gaps = 7/231 (3%)

Query: 3   FFAIDPIGLPHTSLIFYLSSLLIALLLVFLTFQAYRLKS-HRYFFLFLQLSQVIGLYTWY 61
```

```
           FF       P   L +Y+S L   L L F T   YR K  ++ FF  LQ  Q+I LY WY
Sbjct: 7   FFTTQATKPPKFDLFWYVS-LFTLLALTFYTAHRYREKKVYQRFFQILQTVQLILLYGWY 65

Query: 62  VLRGFPLDEALPLYHCRIAMLAIFFLPDRNKFKQLFMVLGIGGTFLALL--SPDLYPFRL 119
            +   PL E+LP YHCR+AM  +  LP ++K+KQ F +LG  GT  A +   PD YPF
Sbjct: 66  WVNHMPLSESLPFYHCRMAMFVVLLLPGQSKYKQYFALLGTFGTLAAFVYPVPDAYPFP- 124

Query: 120 WHVANVSFYFGHVALLVNGLIYLLRFYDASQLRLLSVVRYLATVNFLLLLVSLATKGNYG 179
            H+  +SF FGH ALL N L+YLLR Y+A  L +  +       +N L+ +V+L T G+YG
Sbjct: 125 -HITILSFIFGHLALLGNSLVYLLRQYNARLLDVKGIFLMTFALNALIFVVNLVTGGDYG 183

Query: 180 FVMDIPVIHTRHLLLNFVIVTSGLTFMVKITEYFYLKFGEAQQLALAFSKE 230
           F+   P++    L+ N+++V+  L   + +T+   L+F  AQ+       KE
Sbjct: 184 FLTKPPLVGDHGLVANYLLVSIVLVATISLTKKI-LEFFLAQEAEKMIVKE 233
```

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

```
 Identities = 70/216 (32%), Positives = 117/216 (53%), Gaps = 1/216 (0%)

Query: 2   IEFFSNIRTEIPQMPLLIHSLILSVLPFLMWLTLVNRDKPLYKTIWSILLGLQLITIYTW 61
           ++FF+     +P   L+ +   L +   L++LT         ++  +  L   Q+I +YTW
Sbjct: 1   MDFFAIDPIGLPHTSLIFYLSSLLIALLLVFLTFQAYRLKSHRYFFLFLQLSQVIGLYTW 60

Query: 62  FFWAKLPLSESLPLYHCRIGMFVVL-LARPGILKDYFALLGVVGGVLAMIHPDFYPYQFL 120
           +     PL E+LPLYHCRI M  +  L       K  F +LG+ G   LA++ PD YP++
Sbjct: 61  YVLRGFPLDEALPLYHCRIAMLAIFFLPDRNKFKQLFMVLGIGGTFLALLSPDLYPFRLW 120

Query: 121 HVTNIFFFIGHFALFVLSLLHLMTQSNLDKLNPKLIIQLTLLINMSLIFINLLTGGNYGF 180
           HV N+ F+ GH+AL V  L++L+   +  +L    +++     +N  L+ ++L T GNYGF
Sbjct: 121 HVANVSFYFGHYALLVNGLIYLLRFYDASQLRLLSVVRYLATVNFLLLLVSLATKGNYGF 180

Query: 181 MMKTPILGITNPFLNLFIVTTLLSFLVLFVKQIFQK 216
           +M  P++   + LN  IVT+ L+F+V   + + K
Sbjct: 181 VMDIPVIHTRHLLLNFVIVTSGLTFMVKITEYFYLK 216
```

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

Example 2206

A DNA sequence (GBSx2325) was identified in *S.agalactiae* <SEQ ID 6811> which encodes the amino acid sequence <SEQ ID 6812>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3297(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 2207

A DNA sequence (GBSx2326) was identified in *S.agalactiae* <SEQ ID 6813> which encodes the amino acid sequence <SEQ ID 6814>. This protein is predicted to be oxalate:formate antiporter (oxlT-2). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.80   Transmembrane  380 - 396 ( 376 - 399)
   INTEGRAL    Likelihood = -7.43   Transmembrane  291 - 307 ( 284 - 310)
   INTEGRAL    Likelihood = -5.63   Transmembrane  169 - 185 ( 163 - 186)
   INTEGRAL    Likelihood = -4.99   Transmembrane  226 - 242 ( 223 - 245)
   INTEGRAL    Likelihood = -4.19   Transmembrane   46 -  62 (  39 -  63)
   INTEGRAL    Likelihood = -4.09   Transmembrane  311 - 327 ( 308 - 329)
   INTEGRAL    Likelihood = -1.49   Transmembrane  261 - 277 ( 260 - 278)
   INTEGRAL    Likelihood = -1.06   Transmembrane  133 - 149 ( 133 - 150)
   INTEGRAL    Likelihood = -0.85   Transmembrane   98 - 114 (  98 - 114)
   INTEGRAL    Likelihood = -0.06   Transmembrane   77 -  93 (  77 -  93)

----- 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:AAF36228 GB:AF168363 oxalate:formate antiporter [Lactococcus
          lactis]
 Identities = 220/398 (55%), Positives = 306/398 (76%), Gaps = 3/398 (0%)

Query: 5   NRYVVAVSGVVLHLMLGSTYAWSVFRNPIISETGWDISSVSFAFSLAIFCLGMSAAFMGH 64
           NRYVVA +GV+ HLM+GS YAWSVF NPI  + GW  SSV+ AFS+AI+ LGMSAAFMG
Sbjct: 4   NRYVVAFAGVMFHLMIGSVYAWSVFTNPIAKQNGWAESSVALAFSIAIYFLGMSAAFMGK 63

Query: 65  LVERFGPRIMGMISAILYGAGNVLTGLAIETQQLWLLYVAYGILGGIGLGSGYITPVSTI 124
           +VE+ GPR+ G I++ LYG G ++TG AI     +WLLY++YG++GG+GLG+GY+TPVSTI
Sbjct: 64  VVEKIGPRLTGTIASFLYGTGTIMTGWAIHQNSIWLLYLSYGVIGGLGLGAGYVTPVSTI 123

Query: 125 IKWFPDRRGLATGFAIMGFGFASLVTSPLAQSLMIRIGVGKTFYILGLVYFFVMMIASQF 184
           IKWFPD+RGLATG AIMGFGFA+++T P+AQ LM  +G+ +TFY+LG  YF +M++A+QF
Sbjct: 124 IKWFPDKRGLATGLAIMGFGFAAMLTGPVAQQLMASVGLEQTFYLLGTFYFVIMLLAAQF 183

Query: 185 IKQPPQEKITILTHDGKKNAMNSQIITG--LKANAAIKSKTFYIIWLTLFINISCGLGLI 242
           I + P   ++  T +       +++  G  L AN A+K+K+F  +W+  FINI+CG+GL+
Sbjct: 184 IVR-PNLALSSTTENSISQKKGTRLTRGPELTANQALKTKSFTFLWIMFFINITCGIGLV 242

Query: 243 SAASPMAQDLAGYSAESAALLVGVLGIFNGFGRLLWASLSDYIGRPLTFIILFIVNFIMT 302
           SAASPMAQ + G S ++AA++VG++G+FNGFGRL+WA+LSDYIGRP TF  +FI++ +M
Sbjct: 243 SAASPMAQSMTGMSVQTAAIMVGIIGLFNGFGRLIWATLSDYIGRPATFSAIFILDIVML 302

Query: 303 SSLFLSFNAIVFAIAMSILMTCYGAGFSLLPAYLSDIFGTKELATLHGYSLTAWAIAGLF 362
           S++ +    ++F IA+ +LM+CYGAGFS++PAYL D+FGTKEL  +HGY LTAWA AG+
Sbjct: 303 SAILIFKLPLLFVIALCLLMSCYGAGFSVIPAYLGDVFGTKELGAVHGYVLTAWAAAGVV 362

Query: 363 GPLLLSKTYSWGNSYQLTLMVFGFLFLFGLLLSLYLRK 400
           GPLLLS T+   ++Y LTL  F  + L  LL+S ++++
Sbjct: 363 GPLLLSLTHQLFHNYTLTLAAFILIDLLALLISFWIQR 400
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6815> which encodes the amino acid sequence <SEQ ID 6816>. Analysis of this protein sequence reveals the following:

```
     Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.95   Transmembrane  289 - 305 ( 282 - 321)
   INTEGRAL    Likelihood =-11.83   Transmembrane  376 - 392 ( 372 - 397)
   INTEGRAL    Likelihood = -8.55   Transmembrane  163 - 179 ( 160 - 189)
   INTEGRAL    Likelihood = -7.75   Transmembrane  227 - 243 ( 221 - 247)
   INTEGRAL    Likelihood = -5.89   Transmembrane   44 -  60 (  41 -  67)
   INTEGRAL    Likelihood = -1.38   Transmembrane  310 - 326 ( 309 - 327)
   INTEGRAL    Likelihood = -0.90   Transmembrane  353 - 369 ( 353 - 369)
   INTEGRAL    Likelihood = -0.37   Transmembrane  138 - 154 ( 138 - 154)
   INTEGRAL    Likelihood = -0.06   Transmembrane   98 - 114 (  98 - 114)
   INTEGRAL    Likelihood = -0.00   Transmembrane  259 - 275 ( 259 - 275)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.6180(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:AAF36228 GB:AF168363 oxalate:formate antiporter [Lactococcus
           lactis]
 Identities = 222/399 (55%), Positives = 305/399 (75%), Gaps = 3/399 (0%)

Query: 3   KTKRYIIATAGILLHLMLGSTYAWSVYRNPILQETGWDQAPVAFAFSLAIFCLGLSAAFM 62
           KT RY++A AG++ HLM+GS YAWSV+ NPI ++ GW ++ VA AFS+AI+ LG+SAAFM
Sbjct: 2   KTNRYVVAFAGVMFHLMIGSVYAWSVFTNPIAKQNGWAESSVALAFSIAIYFLGMSAAFM 61

Query: 63  GNLVEQYGPRLTGTVSAILYASGNMLTGLAIDRKEIWLLYIGYGVIGGLGLGAGYITPIS 122
           G +VE+ GPRLTGT+++ LY +G ++TG AI +  IWLLY+ YGVIGGLGLGAGY+TP+S
Sbjct: 62  GKVVEKIGPRLTGTIASFLYGTGTIMTGWAIHQNSIWLLYLSYGVIGGLGLGAGYVTPVS 121

Query: 123 TIIKWFPDKRGMATGFAIMGFGFASLLTSPIAQWLIETEGLVATFYLLGLIYLIVMLFAS 182
           TIIKWFPDKRG+ATG AIMGFGFA++LT P+AQ L+ + GL  TFYLLG  Y ++ML A+
Sbjct: 122 TIIKWFPDKRGLATGLAIMGFGFAAMLTGPVAQQLMASVGLEQTFYLLGTFYFVIMLLAA 181

Query: 183 QLIIKPTAAEIAILDKKRLQ-NNSYLIEG--MTAKEALKTKSFYCLWVILFINITCGLGL 239
           Q I++P  A  +  +    Q   + L  G  +TA +ALKTKSF  LW++ FINITCG+GL
Sbjct: 182 QFIVRPNLALSSTTENSISQKKGTRLTRGPELTANQALKTKSFTFLWIMFFINITCGIGL 241

Query: 240 ISVVAPMAQDLTGMSPEMSAIVVGAMGIFNGFGRLVWASLSDYIGRRVTVILLFLVSIIM 299
           +S  +PMAQ +TGMS + +AI+VG +G+FNGFGRL+WA+LSDYIGR  T   +F++ I+M
Sbjct: 242 VSAASPMAQSMTGMSVQTAAIMVGIIGLFNGFGRLIWATLSDYIGRPATFSAIFILDIVM 301

Query: 300 TISLIFAHSSLIFMISIATLMTCYGAGFSLIPPYLSDLFGAKELATLHGYILTAWAIAAL 359
             +++     L+F+I++  LM+CYGAGFS+IP YL D+FG KEL  +HGY+LTAWA A +
Sbjct: 302 LSAILIFKLPLLFVIALCLLMSCYGAGFSVIPAYLGDVFGTKELGAVHGYVLTAWAAAGV 361

Query: 360 TGPMLLSITVEWTHNYLLTLCVFIVLYILGLMVALRLKK 398
            GP+LLS+T +  HNY LTL  FI++ +L L+++  +++
Sbjct: 362 VGPLLLSLTHQLFHNYTLTLAAFILIDLLALLISFWIQR 400
```

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

```
 Identities = 252/400 (63%), Positives = 329/400 (82%), Gaps = 2/400 (0%)

Query: 1   MKNLNRYVVAVSGVVLHLMLGSTYAWSVFRNPIISETGWDISSVSFAFSLAIFCLGMSAA 60
           M+   RY++A +G++LHLMLGSTYAWSV+RNPI+ ETGWD + V+FAFSLAIFCLG+SAA
Sbjct: 1   MEKTKRYIIATAGILLHLMLGSTYAWSVYRNPILQETGWDQAPVAFAFSLAIFCLGLSAA 60

Query: 61  FMGHLVERFGPRIMGMISAILYGAGNVLTGLAIETQQLWLLYVAYGILGGIGLGSGYITP 120
           FMG+LVE++GPR+ G +SAILY GN+LTGLAI+ +++WLLY+ YG++GG+GLG+GYITP
Sbjct: 61  FMGNLVEQYGPRLTGTVSAILYASGNMLTGLAIDRKEIWLLYIGYGVIGGLGLGAGYITP 120

Query: 121 VSTIIKWFPDRRGLATGFAIMGFGFASLVTSPLAQSLMIRIGVGKTFYILGLVYFFVMMI 180
           +STIIKWFPD+RG+ATGFAIMGFGFASL+TSP+AQ L+   G+  TFY+LGL+Y  VM+
Sbjct: 121 ISTIIKWFPDKRGMATGFAIMGFGFASLLTSPIAQWLIETEGLVATFYLLGLIYLIVMLF 180

Query: 181 ASQFIKQPPQEKITILTHDGKKNAMNSQIITGLKANAAIKSKTFYIIWLTLFINISCGLG 240
           ASQ I +P   +I IL  D K+   NS +I G+ A  A+K+K+FY +W+ LFINI+CGLG
Sbjct: 181 ASQLIIKPTAAEIAIL--DKKRLQNNSYLIEGMTAKEALKTKSFYCLWVILFINITCGLG 238

Query: 241 LISAASPMAQDLAGYSAESAALLVGVLGIFNGFGRLLWASLSDYIGRPLTFIILFIVNFI 300
           LIS  +PMAQDL G S E +A++VG +GIFNGFGRL+WASLSDYIGR +T I+LF+V+ I
Sbjct: 239 LISVVAPMAQDLTGMSPEMSAIVVGAMGIFNGFGRLVWASLSDYIGRRVTVILLFLVSII 298

Query: 301 MTSSLFLSFNAIVFAIAMSILMTCYGAGFSLLPAYLSDIFGTKELATLHGYSLTAWAIAG 360
           MT SL  + ++++F I+++ LMTCYGAGFSL+P YLSD+FG KELATLHGY LTAWAIA
Sbjct: 299 MTISLIFAHSSLIFMISIATLMTCYGAGFSLIPPYLSDLFGAKELATLHGYILTAWAIAA 358
```

```
Query: 361 LFGPLLLSKTYSWGNSYQLTLMVFGFLFLFGLLLSLYLRK 400
          L GP+LLS T  W ++Y LTL VF  L++ GL+++L L+K
Sbjct: 359 LTGPMLLSITVEWTHNYLLTLCVFIVLYILGLMVALRLKK 398
```

A related GBS gene <SEQ ID 8995> and protein <SEQ ID 8996> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:      5.06
GvH: Signal Score (-7.5): 4.38
     Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 10 value:  -7.80 threshold:  0.0
   INTEGRAL    Likelihood = -7.80   Transmembrane  380 - 396 ( 376 - 399)
   INTEGRAL    Likelihood = -7.43   Transmembrane  291 - 307 ( 284 - 310)
   INTEGRAL    Likelihood = -5.63   Transmembrane  169 - 185 ( 163 - 186)
   INTEGRAL    Likelihood = -4.99   Transmembrane  226 - 242 ( 223 - 245)
   INTEGRAL    Likelihood = -4.19   Transmembrane   46 -  62 (  39 -  63)
   INTEGRAL    Likelihood = -4.09   Transmembrane  311 - 327 ( 308 - 329)
   INTEGRAL    Likelihood = -1.49   Transmembrane  261 - 277 ( 260 - 278)
   INTEGRAL    Likelihood = -1.06   Transmembrane  133 - 149 ( 133 - 150)
   INTEGRAL    Likelihood = -0.85   Transmembrane   98 - 114 (  98 - 114)
   INTEGRAL    Likelihood = -0.06   Transmembrane   77 -  93 (  77 -  93)
   PERIPHERAL  Likelihood =  0.42     352
 modified ALOM score:   2.06

*** Reasoning Step: 3

----- 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 databases:

```
ORF02272(313 - 1500 of 1818)
GP|7107009|gb|AAF36228.1|AF168363_4|AF168363(4  -  400  of  421)  oxalate:formate  antiporter
{Lactococcus lactis}
%Match = 38.5
%Identity = 55.4  %Similarity = 79.1
Matches = 220  Mismatches = 81  Conservative Sub.s = 94

216       246       276       306       336       366       396       426
GK*IC*AENW*YIQFFDNLFITNYIFKNKT*VRF*EDCLKNLNRYVVAVSGVVLHLMLGSTYAWSVFRNPIISETGWDIS
                                          |||||| :||::|||:|| |||||| |||  : ||  |
                                        MKTNRYVVAFAGVMFHLMIGSVYAWSVFTNPIAKQNGWAES
                                                 10        20        30        40

456       486       516       546       576       606       636       666
SVSFAFSLAIFCLGMSAAFMGHLVERFGPRIMGMISAILYGAGNVLTGLAIETQQLWLLYVAYGILGGIGLGSGYITPVS
||::|||:||: |||||||||| :||: |||: | |:: ||| | ::|| ||    :||||::||::||:|||:||:|||
SVALAFSIAIYFLGMSAAFMGKVVEKIGPRLTGTIASFLYGTGTIMTGWAIHQNSIWLLYLSYGVIGGLGLGAGYVTPVS
                  60        70        80        90        100       110       120

696       726       756       786       816       846       876       906
TIIKWFPDRRGLATGFAIMGFGFASLVTSPLAQSLMIRIGVGKTFYILGLVYFFVMMIASQFIKQPPQEKITILTHDGKK
||||||||:||||||:||||||||:::| |:|| ||  :|: :|||:||  || :|::|:||| :|    ::  | :
TIIKWFPDKRGLATGLAIMGFGFAAMLTGPVAQQLMASVGLEQTFYLLGTFYFVIMLLAAQFIVRP-NLALSSTTENSIS
                  140       150       160       170       180       190       200

936       960       990       1020      1050      1080      1110      1140
NAMNSQIITG--LKANAAIKSKTFYIIWLTLFINISCGLGLISAASPMAQDLAGYSAESAALLVGVLGIFNGFGRLLWAS
    :::  |  | || |:|:|:|  :|: :||||:||:||:|||||||| : | | ::||::||::|:|||||||:||:
QKKGTRLTRGPELTANQALKTKSFTFLWIMFFINITCGIGLVSAASPMAQSMTGMSVQTAAIMVGIIGLFNGFGRLIWAT
       210       220       230       240       250       260       270       280

1170      1200      1230      1260      1290      1320      1350      1380
```

```
LSDYIGRPLTFIILFIVNFIMTSSLFLSFNAIVFAIAMSILMTCYGAGFSLLPAYLSDIFGTKELATLHGYSLTAWAIAG
||||||| ||  :||:: :| |::::    ::| ||: :||:|||||||::|||| |:|||||  :||| |||||:||
LSDYIGRPATFSAIFILDIVMLSAILIFKLPLLFVIALCLLMSCYGAGFSVIPAYLGDVFGTKELGAVHGYVLTAWAAAG
       290       300       310       320       330       340       350       360

1410      1440      1470      1500      1530      1560      1590      1620
LFGPLLLSKTYSWGNSYQLTLMVFGFLFLFGLLLSLYLRKLTTKVV*YISNLKFFGFTKEFFL*KIVLSYSK*FDILSI*
: |||||| |:   ::| ||| | :: |: ||:|:::::
VVGPLLLSLTHQLFHNYTLTLAAFILIDLLALLISFWIQRDFIKASKLIKKQIIKNYFKAH
       370       380       390       400       410       420
```

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

Example 2208

A DNA sequence (GBSx2327) was identified in *S.agalactiae* <SEQ ID 6817> which encodes the amino acid sequence <SEQ ID 6818>. This protein is predicted to be D-Ala-D-Ala adding enzyme (murF). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1311(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 9739> which encodes amino acid sequence <SEQ ID 9740> was also identified.

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

```
>GP:AAC95436 GB:AF068901 D-Ala-D-Ala adding enzyme [Streptococcus pneumoniae]
 Identities = 313/453 (69%), Positives = 375/453 (82%)

Query: 32  MKLSLHEVAKVVGAKNQVSEFEDVPLGNIEFDSRNISEGDLFLPLKGARDGHEFIEMAFD 91
           MKL++HE+A+VVGAKN +S FED  L   EFDSR I  GDLF+PLKGARDGH+FIE AF+
Sbjct: 1   MKLTIHEIAQVVGAKNDISIFEDTQLEKAEFDSRLIGTGDLFVPLKGARDGHDFIETAFE 60

Query: 92  NGAIATISEKEIEGHPYLLVSDALKAFQVLAQYYIEKMNVDVIAVTGSNGKTTTKDMIAA 151
           NGA  T+SEKE+  HPY+LV D L AFQ LA YY+EK  VDV AVTGSNGKTTTKDM+A
Sbjct: 61  NGAAVTLSEKEVSNHPYILVDDVLTAFQSLASYYLEKTTVDVFAVTGSNGKTTTKDMLAH 120

Query: 152 ILSTTYKTYKTQGNYNNEIGLPYTVLHMPEDTEKIILEMGQDHLGDIHVLSEIAKPRIAV 211
           +LST YKTYKTQGNYNNEIGLPYTVLHMPE TEK++LEMGQDHLGDIH+LSE+A+P+ A+
Sbjct: 121 LLSTRYKTYKTQGNYNNEIGLPYTVLHMPEGTEKLVLEMGQDHLGDIHLLSELARPKTAI 180

Query: 212 VTLIGEAHLEFFGSREKIAEGKMQITDGMSSDGILIAPGDPIIDPYLPANQMTIRFGHDQ 271
           VTL+GEAHL FF  R +IA+GKMQI DGM+S  +L+AP DPI++ YLP ++  +RFG
Sbjct: 181 VTLVGEAHLAFFKDRSEIAKGKMQIADGMASGSLLLAPADPIVEDYLPTDKKVVRFGQGA 240

Query: 272 ELQVTELKEEKHSLTFKTNALEHQLRIPVPGKYNATNAMVAAYVGKLLAVAEEDIVDALE 331
           EL++T+L E K SLTFK N LE  L +PV GKYNATNAM+A+YV      V+EE I  A +
Sbjct: 241 ELEITDLVERKDSLTFKANFLEQVLDLPVTGKYNATNAMIASYVALQEGVSEEQIHQAFQ 300

Query: 332 NLQLTRNRTEWKKSANGADILSDVYNANPTAMRLILETFSAIPNNDGGKKIALLADMKEL 391
           +L+LTRNRTEWKK+ANGADILSDVYNANPTAM+LILETFSAIP N+GGKKIA+LADMKEL
Sbjct: 301 DLELTRNRTEWKKAANGADILSDVYNANPTAMKLILETFSAIPANEGGKKIAVLADMKEL 360

Query: 392 GEQSVDLHNQMIMSIRPDSIDTLICYGQDIEGLAQLASQMFPIGKVYFFKKNQEVDQFDQ 451
           G QSV LHNQMI+S+ PD +DT+I YG+DI  LAQLASQMFPIG VY+FKK ++ DQF+
Sbjct: 361 GNQSVQLHNQMILSLSPDVLDTVIFYGEDIAELAQLASQMFPIGHVYYFKKTEDQDQFED 420
```

```
Query: 452 LLAKVKDTLKEKDQILLKGSNSMNLSKIVDILE 484
           L+ +VK++L   DQILLKGSNSMNL+ +V+ LE
Sbjct: 421 LVKQVKESLSANDQILLKGSNSMNLAMLVESLE 453
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6819> which encodes the amino acid sequence <SEQ ID 6820>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3299(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 = 323/452 (71%), Positives = 387/452 (85%)

Query: 32  MKLSLHEVAKVVGAKNQVSEFEDVPLGNIEFDSRNISEGDLFLPLKGARDGHEFIEMAFD 91
           MKL+LHEVAK+V A+N VS+ +DVPL +IEFDSR I++GDLFLPLKG RDGHEFI++AF
Sbjct: 1   MKLTLHEVAKIVDAQNNVSDLDDVPLHHIEFDSRKITKGDLFLPLKGQRDGHEFIDLAFQ 60

Query: 92  NGAIATISEKEIEGHPYLLVSDALKAFQVLAQYYIEKMNVDVIAVTGSNGKTTTKDMIAA 151
           NGA+AT SEKE+ G P+LLV D LKAFQ LA YYI+KM VDVIAVTGSNGKT+TKDMI A
Sbjct: 61  NGAVATFSEKELPGKPHLLVEDCLKAFQKLAHYYIDKMRVDVIAVTGSNGKTSTKDMIGA 120

Query: 152 ILSTTYKTYKTQGNYNNEIGLPYTVLHMPEDTEKIILEMGQDHLGDIHVLSEIAKPRIAV 211
           +LSTTYKTYKTQGNYNNEIGLPYTVLHMP+DTEKI+LEMGQDH+GDI +LSEIA+PRIAV
Sbjct: 121 VLSTTYKTYKTQGNYNNEIGLPYTVLHMPDDTEKIVLEMGQDHMGDIRLLSEIARPRIAV 180

Query: 212 VTLIGEAHLEFFGSREKIAEGKMQITDGMSSDGILIAPGDPIIDPYLPANQMTIRFGHDQ 271
           +TL+GEAHLE+FGSR+KIA+GKMQI DGM+SDGILIAPGDPIIDPYLP NQM IRFG+ Q
Sbjct: 181 LTLVGEAHLEYFGSRDKIAQGKMQIVDGMNSDGILIAPGDPIIDPYLPENQMVIRFGNQQ 240

Query: 272 ELQVTELKEEKHSLTFKTNALEHQLRIPVPGKYNATNAMVAAYVGKLLAVAEEDIVDALE 331
           E+ VT ++E+K SLTF TN L    + +P+PGKYNATNAMVAAYVGKLLAV +EDI+ AL+
Sbjct: 241 EIDVTGIQEDKDSLTFTTNVLATPVSLPLPGKYNATNAMVAAYVGKLLAVTDEDIIAALQ 300

Query: 332 NLQLTRNRTEWKKSANGADILSDVYNANPTAMRLILETFSAIPNNDGGKKIALLADMKEL 391
            + LT NRTEWKK+ANGADILSDVYNANPTAMRLILETF+ I  N GGKKIA+LADMKEL
Sbjct: 301 TVTLTGNRTEWKKAANGADILSDVYNANPTAMRLILETFANIAKNPGGKKIAVLADMKEL 360

Query: 392 GEQSVDLHNQMIMSIRPDSIDTLICYGQDIEGLAQLASQMFPIGKVYFFKKNQEVDQFDQ 451
           G+ SV LH+Q+I S+    +ID L+ YG  I+ LA+LASQ++P  +V++F K ++ DQF+
Sbjct: 361 GKDSVILHSQLIDSLTSGNIDQLVFYGDHIKELARLASQVYPAEQVHYFLKTEQEDQFEA 420

Query: 452 LLAKVKDTLKEKDQILLKGSNSMNLSKIVDIL 483
           +   V++ L   DQILLKGS+SM+L K+VD L
Sbjct: 421 MAQYVQNILNPFDQILLKGSHSMSLEKLVDRL 452
```

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

Example 2209

A DNA sequence (GBSx2328) was identified in *S.agalactiae* <SEQ ID 6821> which encodes the amino acid sequence <SEQ ID 6822>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1381(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:AAC95435 GB:AF068901 D-Ala-D-Ala ligase [Streptococcus pneumoniae]
 Identities = 243/346 (70%), Positives = 289/346 (83%)

Query: 3   KETLILLYGGRSAEREVSVLSAESVMRAINYDKFFVKTYFITQVGQFIKTQEFDEMPSSD 62
           K+T+ILLYGGRSAEREVSVLSAESVMRA+NYD+F VKT+FI+Q G FIKTQEF   P  +
Sbjct: 2   KQTIILLYGGRSAEREVSVLSAESVMRAVNYDRFTVKTFFISQSGDFIKTQEFSHAPGQE 61

Query: 63  EKLMTNQTVDLDKMVRPSDIYDDNAIVFPVLHGPMGEDGSIQGFLEVLRMPYVGTNILSS 122
           ++LMTN+T+D DK V PS IY++ A+VFPVLHGPMGEDGS+QGFLEVL+MPYVG NILSS
Sbjct: 62  DRLMTNETIDWDKKVAPSAIYEEGAVVFPVLHGPMGEDGSVQGFLEVLKMPYVGCNILSS 121

Query: 123 SVAMDKITTKQVLATVGVPQVAYQTYFEGDDLEHAIKLSLETLSFPIFVKPANMGSSVGI 182
           S+AMDKITTK+VL + G+ QV Y    EGDD+   I    E L++P+F KP+NMGSSVGI
Sbjct: 122 SLAMDKITTKRVLESAGIAQVPYVAIVEGDDVTAKIAEVEEKLAYPVFTKPSNMGSSVGI 181

Query: 183 SKATDESSLRSAIDLALKYDSRILIEQGVTAREIEVGILGNNDVKTTFPGEVVKDVDFYD 242
           SK+ ++  LR A+ LA +YDSR+L+EQGV AREIEVG+LGN DVK+T PGEVVKDV FYD
Sbjct: 182 SKSENQEELRQALKLAFRYDSRVLVEQGVNAREIEVGLLGNYDVKSTLPGEVVKDVAFYD 241

Query: 243 YDAKYIDNKITMDIPAKVDEATMEAMRQYASKAFKAIGACGLSRCDFFLTKDGQIFLNEL 302
           YDAKYIDNKITMDIPAK+ +  + MRQ A  AF+AIG  GLSRCDFF T  G+IFLNEL
Sbjct: 242 YDAKYIDNKITMDIPAKISDDVVAVMRQNAETAFRAIGGLGLSRCDFFYTDKGEIFLNEL 301

Query: 303 NTMPGFTQWSMYPLLWENMGLTYSDLIEKLVMLAKEMFEKRESHLI 348
           NTMPGFTQWSMYPLLW+NMG++Y +LIE+LV LAKE F+KRE+HLI
Sbjct: 302 NTMPGFTQWSMYPLLWDNMGISYPELIERLVDLAKESFDKREAHLI 347
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4559> which encodes the amino acid sequence <SEQ ID 4560>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1451(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 = 261/348 (75%), Positives = 306/348 (87%)

Query: 1   MSKETLILLYGGRSAEREVSVLSAESVMRAINYDKFFVKTYFITQVGQFIKTQEFDEMPS 60
           MSK+TL+LLYGGRSAEREVSVLSAESVMRA+NYDKF VKTYFITQ+GQFIKTQ+F E PS
Sbjct: 1   MSKQTLVLLYGGRSAEREVSVLSAESVMRAVNYDKFLVKTYFITQMGQFIKTQQFSEKPS 60

Query: 61  SDEKLMTNQTVDLDKMVRPSDIYDDNAIVFPVLHGPMGEDGSIQGFLEVLRMPYVGTNIL 120
             E+LMTN+T++L + ++PSDIY++ A+VFPVLHGPMGEDGSIQGFLEVLRMPY+GTN++
Sbjct: 61  ESERLMTNETIELTQKIKPSDIYEEGAVVFPVLHGPMGEDGSIQGFLEVLRMPYIGTNVM 120

Query: 121 SSSVAMDKITTKQVLATVGVPQVAYQTYFEGDDLEHAIKLSLETLSFPIFVKPANMGSSV 180
           SSS+AMDKITTK+VL ++G+PQVAY  Y +G DLE  +  +L  L+FPIFVKPANMGSSV
Sbjct: 121 SSSIAMDKITTKRVLESIGIPQVAYTVYIDGQDLEACLVETLARLTFPIFVKPANMGSSV 180

Query: 181 GISKATDESSLRSAIDLALKYDSRILIEQGVTAREIEVGILGNNDVKTTFPGEVVKDVDF 240
           GISKA  +  LR AI LAL YDSR+LIEQGV AREIEVG+LGN+ VK+T PGEV+KDVDF
Sbjct: 181 GISKAQTKVELRKAIQLALTYDSRVLIEQGVVAREIEVGLLGNDKVKSTLPGEVIKDVDF 240

Query: 241 YDYDAKYIDNKITMDIPAKVDEATMEAMRQYASKAFKAIGACGLSRCDFFLTKDGQIFLN 300
           YDY AKY+DNKITM IPA VD++ +  MR YA  AFKA+G CGLSRCDFFLT+DGQ++LN
Sbjct: 241 YDYQAKYVDNKITMAIPADVDQSIVTEMRSYAEVAFKALGGCGLSRCDFFLTQDGQVYLN 300

Query: 301 ELNTMPGFTQWSMYPLLWENMGLTYSDLIEKLVMLAKEMFEKRESHLI 348
```

```
            ELNTMPGFTQWSMYPLLWENMGL Y DLIE+LV LA+EMF++RESHLI
Sbjct: 301 ELNTMPGFTQWSMYPLLWENMGLAYPDLIEELVTLAQEMFDQRESHLI 348
```

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

Example 2210

A DNA sequence (GBSx2329) was identified in *S.agalactiae* <SEQ ID 6823> which encodes the amino acid sequence <SEQ ID 6824>. This protein is predicted to be recombination protein (recR). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2540(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:AAC44615 GB:U58210 RecM [Streptococcus thermophilus]
 Identities = 181/198 (91%), Positives = 189/198 (95%)

Query: 1   MLYPTPIAKLIDSFSKLPGIGTKTATRLAFYTIGMSDEDVNEFAKNLLAAKRELTYCSVC 60
           MLYPTPIAKLIDSFSKLPGIG KTATRLAFYTI MSDEDVN+FAKNLLAAKRELTYCSVC
Sbjct: 1   MLYPTPIAKLIDSFSKLPGIGAKTATRLAFYTISMSDEDVNDFAKNLLAAKRELTYCSVC 60

Query: 61  GNLTDDDPCLICTDKTRDQSVILVVEDSKDVSAMEKIQEYNGLYHVLHGLISPMNGISPD 120
           G LTDDDPC+ICTD+TRD++ ILVVEDSKDVSAMEKIQEY GLYHVL GLISPMNG+ PD
Sbjct: 61  GRLTDDDPCIICTDETRDRTKILVVEDSKDVSAMEKIQEYRGLYHVLQGLISPMNGVGPD 120

Query: 121 DINLKSLITRLMDGQVTEVIVATNATADGEATSMYISRVLKPAGIKVTRLARGLAVGSDI 180
           DINLKSLITRLMD +V EVI+ATNATADGEATSMYISRVLKPAGIKVTRLARGLAVGSDI
Sbjct: 121 DINLKSLITRLMDSEVDEVIIATNATADGEATSMYISRVLKPAGIKVTRLARGLAVGSDI 180

Query: 181 EYADEVTLLRAIENRTEL 198
           EYADEVTLLRAIENRTEL
Sbjct: 181 EYADEVTLLRAIENRTEL 198
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6825> which encodes the amino acid sequence <SEQ ID 6826>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2652(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 = 180/198 (90%), Positives = 192/198 (96%)

Query: 1   MLYPTPIAKLIDSFSKLPGIGTKTATRLAFYTIGMSDEDVNEFAKNLLAAKRELTYCSVC 60
           +LYPTPIAKLIDS+SKLPGIG KTATRLAFYTIGMS+EDVN+FAKNLLAAKRELTYCS+C
Sbjct: 1   VLYPTPIAKLIDSYSKLPGIGIKTATRLAFYTIGMSNEDVNDFAKNLLAAKRELTYCSIC 60

Query: 61  GNLTDDDPCLICTDKTRDQSVILVVEDSKDVSAMEKIQEYNGLYHVLHGLISPMNGISPD 120
           GNLTDDDPC ICTD +RDQ+ ILVVED+KDVSAMEKIQEY+G YHVLHGLISPMNG+ PD
Sbjct: 61  GNLTDDDPCHICTDTSRDQTTILVVEDAKDVSAMEKIQEYHGYYHVLHGLISPMNGVGPD 120
```

```
Query: 121 DINLKSLITRLMDGQVTEVIVATNATADGEATSMYISRVLKPAGIKVTRLARGLAVGSDI 180
           DINLKSLITRLMDG+V+EVIVATNATADGEATSMYISRVLKPAGIKVTRLARGLAVGSDI
Sbjct: 121 DINLKSLITRLMDGKVSEVIVATNATADGEATSMYISRVLKPAGIKVTRLARGLAVGSDI 180

Query: 181 EYADEVTLLRAIENRTEL 198
           EYADEVTLLRAIENRTEL
Sbjct: 181 EYADEVTLLRAIENRTEL 198
```

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

Example 2211

A DNA sequence (GBSx2330) was identified in *S.agalactiae* <SEQ ID 6827> which encodes the amino acid sequence <SEQ ID 6828>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3144(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 2212

A DNA sequence (GBSx2331) was identified in *S.agalactiae* <SEQ ID 6829> which encodes the amino acid sequence <SEQ ID 6830>. This protein is predicted to be penicillin-binding protein 2b. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.69   Transmembrane   23 -  39 (  17 -  46)

----- Final Results -----
               bacterial membrane --- Certainty=0.6477(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:AAC44614 GB:U58210 penicillin-binding protein 2b [Streptococcus thermophilus]
 Identities = 341/683 (49%), Positives = 477/683 (68%), Gaps = 12/683 (1%)

Query: 4   RKKRYRLTVKKQNASIPRRLNLLFFIIVLLFTVLILRLEQMQIGQQSFYMKKLTALTSYT 63
           ++K  R    ++  +I RR+ LLF ++ +LF +L  RL  MQ+  +SFY KKL   + YT
Sbjct: 18  KRKEKRANKPRKPVNISRRVYLLFGVVFVLFLLLFARLTYMQVYNKSFYTKKLEDNSKYT 77

Query: 64  VKESKARGQIFDAKGVVLVENDERPTVAFSRGNNISSQSIKELANKLSHYITLTEVASSD 123
           V+ +  RGQIFDAKG+ L  N  +  + F+R N +SS ++K +A +L+  +TLTE   +D
Sbjct: 78  VRIASERGQIFDAKGIALTTNQSKDVITFTRSNLVSSDTMKSVAERLATLVTLTETKVTD 137

Query: 124 RAKRDYYLADKANYKKVVESLPDSKRYDKFGNHLAESTVYANAVAAVPVSAINYSEDELK 183
           R KR++YLAD ANYK+VV  LP+ K+ DKFGN LAE+T+Y NA+ AVP  A++YSEDELK
Sbjct: 138 RQKREFYLADSANYKRVVNDLPNDKKTDKFGNKLAEATIYNNAINAVPDEAVDYSEDELK 197
```

```
Query: 184 VVALFNQMNATPTFGSVKLSTGELSDDQIKKLDADKKELLGISVTSNWHRRKKGTSLSDI 243
           +V +++ MNA   F +V L T +L+ DQI  + A +KEL GI V  +W R    +SLS +
Sbjct: 198 IVYIYSHMNAVSNFSTVILKTADLTPDQIAIVAAKQKELNGIRVAKDWERHTSDSSLSPL 257

Query: 244 LGTISTEKAGLPREEVKKYLKKGYSLNDRVGTSYLEKQYEDDLQGIRQIRKVVVNKKGKV 303
           +G +S+ +AGLP+E+ K YLKKGY+LNDRVGTSYLEK+YE++LQG   +R++ V+K+GKV
Sbjct: 258 IGRVSSSEAGLPQEDAKDYLKKGYALNDRVGTSYLEKEYEEELQGKHTVREITVDKEGKV 317

Query: 304 VSDNITQEGKSGRNLKLTIDLNYQNKVESILKQYYGSELSSGRASFSEGMYAVAIEPSTG 363
            SD I Q+G  G NLKLTIDL++Q  VE IL Q   SE+S  +A++SEGMYAV +   TG
Sbjct: 318 DSDKIIQKGSKGNNLKLTIDLDFQKGVEDILGQQLSSEISGNKATYSEGMYAVVMNADTG 377

Query: 364 KVLAMAGLKNDHG--NLVDDSLGTIAKNFTPGSVVKGATLSSGWENKVLRGNEVLYDQEI 421
            VLAMAG K++ G  +   D+LGTI   FTPGSVVKGATL++GW +  + G++VL DQ I
Sbjct: 378 AVLAMAGQKHEQGAQDFKADALGTITDVFTPGSVVKGATLTAGWRSGAIYGDQVLTDQPI 437

Query: 422 -----ANIRSWFT-RGLTPISAAQALEYSSNTYMVQVALRLMGQDYNTGDALTDRGYQEA 475
                I SWFT +G   I+A QALEYSSNTYMVQ+A++ +GQ Y  G +L+    ++A
Sbjct: 438 NIASSPPITSWFTDKGSRAITATQALEYSSNTYMVQIAIKRLGQQYVPGMSLSTDNMEKA 497

Query: 476 MAKLRKTYGEYGLGVSTGLDLP-ESEGYVPGKYSLGTTLMESFGQYDAYTPMQLGQYIST 534
           M  LR TY E+G+GVSTGLDLP ESEGY+P  Y++   L E+FGQYD+YT +QL QY+++
Sbjct: 498 MTTLRDTYAEFGMGVSTGLDLPGESEGYIPKNYNVANVLTEAFGQYDSYTTIQLAQYVAS 557

Query: 535 IANNGNRLAPHVVSDIYEGNDSNKFAQLVRSITPKTLNKIAISDQELAIIQEGFYNVVNS 594
           IAN G R+APH+V  IY+   +     L  ++  + LNK+++  ++L IIQ+GF++VVNS
Sbjct: 558 IANGGKRVAPHIVGGIYDAGKNGSLGTLSSTVDTRVLNKLSLDSKQLGIIQQGFHDVVNS 617

Query: 595 GSGYATGTSMRGNVTTISGKTGTAETFAKNVNGQTVSTYNLNAIAYDTNR---KIAVAVM 651
           GS  ATG +M  ++  ISGKTGTAET+A +G +V+T NLNA+AY T +   K+AV +M
Sbjct: 618 GSSLATGKAMASSIIPISGKTGTAETYATDGSGNSVTTVNLNAVAYATAKDGTKLAVGIM 677

Query: 652 YPHVTTDTTKSHQLVARDMIDQY 674
           YPH     +K+HQ   + +++ Y
Sbjct: 678 YPHALDWKSKAHQNAVKAIMELY 700
```

A related GBS gene <SEQ ID 8997> and protein <SEQ ID 8998> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 8
McG: Discrim Score:    -12.38
GvH: Signal Score (-7.5): -5.9
     Possible site: 35
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value: -12.42 threshold:  0.0
   INTEGRAL    Likelihood =-12.42   Transmembrane   23 -  39 (  18 -  46)
   PERIPHERAL  Likelihood =  4.56     355
 modified ALOM score:   2.98

*** Reasoning Step: 3

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

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

```
50.5/71.3% over 683aa

                                                                Streptococcus
thermophilus
  GP|1685112| penicillin-binding protein 2b Insert characterized

ORF02276(307 - 2322 of 2643)
GP|1685112|gb|AAC44614.1||U58210(17   -   700   of   704)   penicillin-binding   protein   2b
(Streptococcus thermophilus)
%Match = 38.5
```

```
%Identity = 50.4  %Similarity = 71.2
Matches = 342  Mismatches = 189  Conservative Sub.s = 141

108       138       168       198       228       258       288       318
NHGR*NS*LPTTCFRI**KIKPCFRILLR*II*SLYKKFRPSWLEFFIIYNILSVCKKPFL*YNSSQSFYSKELMLNRKK
                                                                :     ::    :    ::|
                                                               MTSFWEKNSQKWKKWRQKRK
                                                                         10        20

348       378       408       438       468       498       528       558
RYRLTVKKQNASIPRRLNLLFFIIXLLFTVLILRLEQMQIGQQSFYMKKLTALTSYTVKESKARGQIFDAKGVVLVENDE
  |    ::   :| ||: ||| :: :|| :|  ||  ||:  :||| |||   : |||: :  |||||||||: |  |
EKRANKPRKPVNISRRVYLLFGVVFVLFLLLFARLTYMQVYNKSFYTKKLEDNSKYTVRIASERGQIFDAKGIALTTNQS
        30        40        50        60        70        80        90       100

588       618       648       678       708       738       768       798
RPTVAFSRGNNISSQSIKELANKLSHYITLTEVASSDRAKRDYYLADKANYKKVVESLPDSKRYDKFGNHLAESTVYANA
:  : |:| | :|| ::| :| :|:  :|||| :|| ||::|||| ||||:|| ||: |: ||||| |||:|:| ||
KDVITFTRSNLVSSDTMKSVAERLATLVTLTETKVTDRQKREFYLADSANYKRVVNDLPNDKKTDKFGNKLAEATIYNNA
       110       120       130       140       150       160       170       180

828       858       888       918       948       978      1008      1038
VAAVPVSAINYSEDELKVVALFNQMNATPTFGSVKLSTGELSDDQIKKLDADKKELLGISVTSNWHRRKKGTSLSDILGT
: |||  |::|||||||:| ::::||| . | :| | | :|: |||  : | :||| || |  :| |    :||| ::|
INAVPDEAVDYSEDELKIVYIYSHMNAVSNFSTVILKTADLTPDQIAIVAAKQKELNGIRVAKDWERHTSDSSLSPLIGR
       190       200       210       220       230       240       250       260

1068      1098      1128      1158      1188      1218      1248      1278
ISTEKAGLPREEVKKYLKKGYSLNDRVGTSYLEKQYEDDLQGIRQIRKVVVNKKGKVVSDNITQEGKSGRNLKLTIDLNY
:|: :||||:|: | |||||||:||||||||||||:||::|||   :|:: |:|:||| || | |:|  | ||||||||::
VSSSEAGLPQEDAKDYLKKGYALNDRVGTSYLEKEYEEELQGKHTVREITVDKEGKVDSDKIIQKGSKGNNLKLTIDLDF
       270       280       290       300       310       320       330       340

1308      1338      1368      1398      1428      1452      1482      1512
QNKVESILKQYYGSELSSGRASFSEGMYAVAIEPSTGKVLAMAGLKNDHG--NLVDDSLGTIAKNFTPGSVVKGATLSSG
|  || || |   ||:|  :|::|||||||| :   || |||||| |:::|  ::  |:|||| ||||||||||||::|
QKGVEDILGQQLSSEISGNKATYSEGMYAVVMNADTGAVLAMAGQKHEQGAQDFKADALGTITDVFTPGSVVKGATLTAG
       350       360       370       380       390       400       410       420

1542      1566      1587      1614      1644      1674      1704      1734
WENKVLRGNEVLYDQ--EIAN---IRSWFT-RGLTPISAAQALEYSSNTYMVQVALRLMGQDYNTGDALTDRGYQEAMAK
| :  : |::|| ||   ||:   | |||| :|   |:| ||||||||||||||:|:: :|| |  | :|:    ::||
WRSGAIYGDQVLTDQPINIASSPPITSWFTDKGSRAITATQALEYSSNTYMVQIAIKRLGQQYVPGMSLSTDNMEKAMTT
       430       440       450       460       470       480       490       500

1764              1821      1851      1881      1911      1941      1971
LRKTYGEYGLGVSTGLDLP-ESEGYVPGKYSLGTTLMESFGQYDAYTPMQLGQYISTIANNGNRLAPHVVSDIYEGNDSN
|| || |:|:|||||||| |||||:|  |::   | |:|||||:|| :|| ||:::||| | |:|||:|  ||:  :
LRDTYAEFGMGVSTGLDLPGESEGYIPKNYNVANVLTEAFGQYDSYTTIQLAQYVASIANGGKRVAPHIVGGIYDAGKNG
       510       520       530       540       550       560       570       580

2001      2031      2061      2091      2121      2151      2181      2211
KFAQLVRSITPKTLNKIAISDQELAIIQEGFYNVVNSGSGYATGTSMRGNVTTISGKTGTAETFAKNVNGQTVSTYNLNA
 :  |  ::  : |||::: ::| |||:||::|||||| ||| :|  :: ||||||||||:| : :| :|:| ||||
SLGTLSSTVDTRVLNKLSLDSKQLGIIQQGFHDVVNSGSSLATGKAMASSIIPISGKTGTAETYATDGSGNSVTTVNLNA
       590       600       610       620       630       640       650       660

2262      2292      2322      2352      2382      2412      2442
IAYDTNR---KIAVAVMYPHVTTDTTKSHQLVARDMIDQYISQFTGQ*ERTFECFTQHQLLN*LTAFQNYRV*VLKQQHV
:|| | :   |:|| :|||| :|:||  : ::: | :
VAYATAKDGTKLAVGIMYPHALDWKSKAHQNAVKAIMELYQNTH
       670       680       690       700
```

SEQ ID 8998 (GBS292) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 9; MW 103kDa).

GBS292-GST was purified as shown in Figure 211, lane 7.

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

Example 2213

A DNA sequence (GBSx2332) was identified in *S.agalactiae* <SEQ ID 6831> which encodes the amino acid sequence <SEQ ID 6832>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2644(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:CAB51328 GB:AJ131985 phosphoglyceromutase [Streptococcus pneumoniae]
 Identities = 219/230 (95%), Positives = 226/230 (98%)

Query: 1   MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLIQAAGIEFDLAFTSVLKR 60
           MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLI+ AGI+FD A+TSVLKR
Sbjct: 1   MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLIKEAGIKFDQAYTSVLKR 60

Query: 61  AIKTTNLALEAADQLWVPVEKSWRLNERHYGGLTGKNKAEAAEQFGDEQVHIWRRSYDVL 120
           AIKTTNLALEA+DQLWVPVEKSWRLNERHYGGLTGKNKAEAAEQFGDEQVHIWRRSYDVL
Sbjct: 61  AIKTTNLALEASDQLWVPVEKSWRLNERHYGGLTGKNKAEAAEQFGDEQVHIWRRSYDVL 120

Query: 121 PPDMAKDDEHSAHTDRRYASLDDSVIPDAENLKVTLERALPFWEDKIAPALKDGKNVFVG 180
           PP+M +DDEHSAHTDRRYASLDDSVIPDAENLKVTLERALPFWEDKIAPALKDGKNVFVG
Sbjct: 121 PPNMDRDDEHSAHTDRRYASLDDSVIPDAENLKVTLERALPFWEDKIAPALKDGKNVFVG 180

Query: 181 AHGNSIRALVKHIKQLSDDEIMDVEIPNFPPLVFEFDEKLNLVSEYYLGK 230
           AHGNSIRALVKHIK LSDDEIMDVEIPNFPPLVFEFDEKLN+VSEYYLGK
Sbjct: 181 AHGNSIRALVKHIKGLSDDEIMDVEIPNFPPLVFEFDEKLNVVSEYYLGK 230
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6833> which encodes the amino acid sequence <SEQ ID 6834>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2646(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 = 206/229 (89%), Positives = 214/229 (92%)

Query: 1   MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLIQAAGIEFDLAFTSVLKR 60
           MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLI+ AGIEFDLAFTSVL R
Sbjct: 1   MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLIKEAGIEFDLAFTSVLTR 60

Query: 61  AIKTTNLALEAADQLWVPVEKSWRLNERHYGGLTGKNKAEAAEQFGDEQVHIWRRSYDVL 120
           AIKTTNLALE A QLWVP EKSWRLNERHYG LTGKNKAEAAEQF DEQVHIWRRSYDVL
Sbjct: 61  AIKTTNLALENAGQLWVPTEKSWRLNERHYGALTGKNKAEAAEQFCDEQVHIWRRSYDVL 120

Query: 121 PPDMAKDDEHSAHTDRRYASLDDSVIPDAENLKVTLERALPFWEDKIAPALKDGKNVFVG 180
           PP MAKDDE+SAH DRRYA LD ++IPDAENLKVTLERA+P+WE+KIAPAL DGKNVFVG
Sbjct: 121 PPAMAKDDEYSAHKDRRYADLDPALIPDAENLKVTLERAMPYWEEKIAPALLDGKNVFVG 180

Query: 181 AHGNSIRALVKHIKQLSDDEIMDVEIPNFPPLVFEFDEKLNLVSEYYLG 229
```

```
          AHGNSIRALVKHIK LSDDEIMDVEIPNFPPLVFE DEKLN+V EYYLG
Sbjct: 181 AHGNSIRALVKHIKGLSDDEIMDVEIPNFPPLVFELDEKLNIVKEYYLG 229
```

SEQ ID 6832 (GBS110) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 8; MW 28.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 10; MW 53.9kDa).

The GBS110-GST fusion product was purified (Figure 204, lane 5) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 252A), FACS (Figure 252B), 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 2214

A DNA sequence (GBSx2333) was identified in *S.agalactiae* <SEQ ID 6835> which encodes the amino acid sequence <SEQ ID 6836>. This protein is predicted to be triosephosphate isomerase (tpiA). Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane   36 -  52 (  36 -  52)

----- 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:AAC43268 GB:U07640 triosephosphate isomerase [Lactococcus
          lactis]
 Identities = 164/252 (65%), Positives = 202/252 (80%)

Query: 1   MSRKPFIAGNWKMNKNPEEAKAFIEAVASKLPSSELVEAGIAAPALTLSTVLEAAKGSEL 60
           MSRKP IAGNWKMNK   EA+AF+EAV + LPSS+ VE+ I APAL L+ +    +GSEL
Sbjct: 1   MSRKPIIAGNWKMNKTLSEAQAFVEAVKNNLPSSDNVESVIGAPALFLAPMAYLRQGSEL 60

Query: 61  KIAAQNSYFENSGAFTGENSPKVLAEMGTDYVVIGHSERRDYFHETDQDINKKAKAIFAN 120
           K+AA+NSYFEN+GAFTGENSP  + ++G +Y++IGHSERR+YFHETD+DINKKAKAIFA
Sbjct: 61  KLAAENSYFENAGAFTGENSPAAIVDLGIEYIIIGHSERREYFHETDEDINKKAKAIFAA 120

Query: 121 GLTPIICCGESLETYEAGKAVEFVGAQVSAALAGLSEEQVSSLVIAYEPIWAIGTGKSAT 180
           G TPI+CCGE+LET+EAGK  E+V  Q+ A LAGL+ EQVS+LVIAYEPIWAIGTGK+AT
Sbjct: 121 GATPILCCGETLETFEAGKTAEWVSGQIEAGLAGLTAEQVSNLVIAYEPIWAIGTGKTAT 180

Query: 181 QDDAQNMCKAVRDVVAADFGQAVADKVRVQYGGSVKPENVAEYMACPDVDGALVGGASLE 240
            + A   C  VR  V   +G+ V++ VR+QYGGSVKPE +   MA  ++DGALVGGASLE
Sbjct: 181 NEIADETCGVVRSTVEKLYGKEVSEAVRIQYGGSVKPETIEGLMAKENIDGALVGGASLE 240

Query: 241 AESFLALLDFVK 252
           A+SFLALL+  K
Sbjct: 241 ADSFLALLEMYK 252
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6837> which encodes the amino acid sequence <SEQ ID 6838>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
```

```
   INTEGRAL    Likelihood = -1.81   Transmembrane   36 -  52 (  36 -  52)

----- Final Results -----
               bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 220/251 (87%), Positives = 237/251 (93%)

Query: 1   MSRKPFIAGNWKMNKNPEEAKAFIEAVASKLPSSELVEAGIAAPALTLSTVLEAAKGSEL 60
           MSRKP IAGNWKMNKNP+EAKAF+EAVASKLPS++LV+  +AAPA+ L T +EAAK S L
Sbjct: 1   MSRKPIIAGNWKMNKNPQEAKAFVEAVASKLPSTDLVDVAVAAPAVDLVTTIEAAKDSVL 60

Query: 61  KIAAQNSYFENSGAFTGENSPKVLAEMGTDYVVIGHSERRDYFHETDQDINKKAKAIFAN 120
           K+AAQN YFEN+GAFTGE SPKVLAEMG DYVVIGHSERRDYFHETD+DINKKAKAIFAN
Sbjct: 61  KVAAQNCYFENTGAFTGETSPKVLAEMGADYVVIGHSERRDYFHETDEDINKKAKAIFAN 120

Query: 121 GLTPIICCGESLETYEAGKAVEFVGAQVSAALAGLSEEQVSSLVIAYEPIWAIGTGKSAT 180
           GLTPI+CCGESLETYEAGKAVEFVGAQVSAALAGLS EQV+SLV+AYEPIWAIGTGKSAT
Sbjct: 121 GLTPIVCCGESLETYEAGKAVEFVGAQVSAALAGLSAEQVASLVLAYEPIWAIGTGKSAT 180

Query: 181 QDDAQNMCKAVRDVVAADFGQAVADKVRVQYGGSVKPENVAEYMACPDVDGALVGGASLE 240
           QDDAQNMCKAVRDVVAADFGQ VADKVRVQYGGSVKPENV +YMACPDVDGALVGGASLE
Sbjct: 181 QDDAQNMCKAVRDVVAADFGQEVADKVRVQYGGSVKPENVKDYMACPDVDGALVGGASLE 240

Query: 241 AESFLALLDFV 251
           A+SFLALLDF+
Sbjct: 241 ADSFLALLDFL 251
```

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

Example 2215

A DNA sequence (GBSx2334) was identified in *S.agalactiae* <SEQ ID 6839> which encodes the amino acid sequence <SEQ ID 6840>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3050(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:AAB41198 GB:U75481 elongation factor-Tu [Streptococcus mutans]
 Identities = 44/45 (97%), Positives = 45/45 (99%)

Query: 1   MVMPGDNVTIEVELIHPIAVEQGTTFSIREGGRTVGSGIVSEIEA 45
           MVMPGDNVTI+VELIHPIAVEQGTTFSIREGGRTVGSGIVSEIEA
Sbjct: 117 MVMPGDNVTIDVELIHPIAVEQGTTFSIREGGRTVGSGIVSEIEA 161
```

There is also homology to SEQ ID 1022:

```
 Identities = 44/45 (97%), Positives = 44/45 (97%)

Query: 1   MVMPGDNVTIEVELIHPIAVEQGTTFSIREGGRTVGSGIVSEIEA 45
           MVMPGDNVTI VELIHPIAVEQGTTFSIREGGRTVGSGIVSEIEA
Sbjct: 371 MVMPGDNVTINVELIHPIAVEQGTTFSIREGGRTVGSGIVSEIEA 415
```

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

Example 2216

A DNA sequence (GBSx2335) was identified in *S.agalactiae* <SEQ ID 6841> which encodes the amino acid sequence <SEQ ID 6842>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.66   Transmembrane   81 -  97 (  80 -  97)
   INTEGRAL    Likelihood = -2.60   Transmembrane   18 -  34 (  17 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.2062(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 2217

A DNA sequence (GBSx2336) was identified in *S.agalactiae* <SEQ ID 6843> which encodes the amino acid sequence <SEQ ID 6844>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0596(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 2218

A DNA sequence (GBSx2337) was identified in *S.agalactiae* <SEQ ID 6845> which encodes the amino acid sequence <SEQ ID 6846>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3559(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 2219

A DNA sequence (GBSx2338) was identified in *S.agalactiae* <SEQ ID 6847> which encodes the amino acid sequence <SEQ ID 6848>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have an uncleavable N-term signal seq

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

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

```
>GP:AAF96286 GB:AE004374 hypothetical protein [Vibrio cholerae]
 Identities = 56/167 (33%), Positives = 89/167 (52%), Gaps = 12/167 (7%)

Query: 18  LAIIKSLPLNDCWLCAGTLRNFVWNKLS-GINETLTSDIDVVFFDKNI---SYEETVVLE 73
           L  +  L L  C++ AG +RN VW+ L    +  T  +DIDV+FFD +      YE++  LE
Sbjct: 41  LECVYQLELPQCYIAAGFVRNLVWDSLHHNVKLTPLNDIDVIFFDADCLDSDYEKS--LE 98

Query: 74  QQLKDNYPQYDWELKNEFYMNTHSPNTPKYTSSKDAISKFPEKCTAVGARLDDRNQLELY 133
            +L +  PQ +W++KN+  M+  + + P Y S+ DA+S +PEK TAV  R  + ++ E
Sbjct: 99  LKLSEQMPQLNWQVKNQAKMHLQNGDNP-YQSTLDAMSYWPEKETAVAVRKVEHDRYECI 157

Query: 134 LPYGEEEILNFIVSPTPYFEEDLLRYNVYLKRVDKKKWNNIWPRLTI 180
             +G E +    ++  P        Y ++  RV  K W  +WP L I
Sbjct: 158 SAFGFESLFQGFITHNP-----KRAYGIFENRVKSKGWLAMWPNLRI 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 2220

A DNA sequence (GBSx2339) was identified in *S.agalactiae* <SEQ ID 6849> which encodes the amino acid sequence <SEQ ID 6850>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2779(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:CAB13060 GB:Z99110 yjdF [Bacillus subtilis]
 Identities = 47/138 (34%), Positives = 93/138 (67%), Gaps = 2/138 (1%)

Query: 1   MKMTVYFDGNFWLGLIEYDDDGDYKVFRYFFGKEPKDDDVFNFINHKLNDLIKKYEFVKT 60
           MK+T+Y+DG FW+G++E  D+G  + FR+ FGKEP+D +V  F++++L +++ + E  +
Sbjct: 24  MKLTIYYDGQFWVGVVEVVDNGKLRAFRHLFGKEPRDSEVLEFVHNQLLNMMAQAE--QE 81

Query: 61  DISLKRTNEHKKSPKRMQREINREKRKPVVSTKAQLAMKTIHMSIKNERQLSQKCKKNEL 120
            + L+   + K +PKR+QR++++E +   V++KAQ A+K    + K +++   K ++  +
Sbjct: 82  GVRLQGRRQKKINPKRLQRQVSKELKNAGVTSKAQEAIKLELEARKQKKKQIMKEQREHV 141
```

```
Query: 121 RKHRYQLKQEKRYQKKKG 138
           ++ RY LK++K  +K +G
Sbjct: 142 KEQRYMLKKQKAKKKHRG 159
```

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 2221

A DNA sequence (GBSx2340) was identified in *S.agalactiae* <SEQ ID 6851> which encodes the amino acid sequence <SEQ ID 6852>. This protein is predicted to be ComX1. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3143(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 9469> which encodes amino acid sequence <SEQ ID 9470> was also identified.

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

```
>GP:AAD50429 GB:AF161701 ComX2 [Streptococcus pneumoniae]
 Identities = 61/152 (40%), Positives = 95/152 (62%)

Query: 5   EELFDKVKPIVMKLRRNYFVQLWEYDDWIQEGRIVLFRLLEEHPYLLDNESKLFIYFKTK 64
           +EL+++V+  V K R  Y++ LWE  DW QEG + L  L+     L+D+  +L  YFKTK
Sbjct: 3   KELYEEVQGTVYKCRNEYYLHLWELSDWDQEGMLCLHELISREEGLVDDIPRLRKYFKTK 62

Query: 65  FSNYLNDVLRHQDCQKRQFNKMPYEEISEVSHYVKSKGLVLDDYIAYRDTLTKVEETLSD 124
           F N + D +R Q+ QKR+++K PYEE+ E+SH +   GL LDDY  + +TL       S
Sbjct: 63  FRNRILDYIRKQESQKRRYDKEPYEEVGEISHRISEGGLWLDDYYLFHETLRDYRNKQSK 122

Query: 125 IDKEKFEKLISGERFAGKKQFIRDIQPFFNAF 156
             +E+ E+++S ERF G+++ +RD++  F  F
Sbjct: 123 EKQEELERVLSNERFRGRQRVLRDLRIVFKEF 154
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6853> which encodes the amino acid sequence <SEQ ID 6854>. Analysis of this protein sequence reveals the following:

```
      Possible site: 39

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.35   Transmembrane    9 -  25 (   7 -  28)

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

A related sequence was also identified in GAS <SEQ ID 9163> which encodes the amino acid sequence <SEQ ID 9164>. Analysis of this protein sequence reveals the following:

```
    Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.35   Transmembrane    2 -  18 (   1 -  18)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.160(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:AAD50429 GB:AF161701 ComX2 [Streptococcus pneumoniae]
 Identities = 60/149 (40%), Positives = 98/149 (65%)

Query: 41  FEKVKPIILKLKRHYYIQLWDRDDWLQEGHIILLQLLERYPELIEEEERLYRYFKTKFSS 100
           +E+V+  + K +  YY+ LW+  DW QEG + L +L+ R   L+++  RL +YFKTKF +
Sbjct: 6   YEEVQGTVYKCRNEYYLHLWELSDWDQEGMLCLHELISREEGLVDDIPRLRKYFKTKFRN 65

Query: 101 YLKDLLRRQESQKRQFHKLAYEEIGEVAHAIPSRGLWLDDYVAYQEVIASLENQLNSQER 160
            + D +R+QESQKR++ K  YEE+GE++H I   GLWLDDY  + E +    N+ + +++
Sbjct: 66  RILDYIRKQESQKRRYDKEPYEEVGEISHRISEGGLWLDDYYLFHETLRDYRNKQSKEKQ 125

Query: 161 MQFQALIRGERFKGRRALLRKISPYFKEF 189
            + + ++  ERF+GR+ +LR +   FKEF
Sbjct: 126 EELERVLSNERFRGRQRVLRDLRIVFKEF 154
```

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

```
 Identities = 78/149 (52%), Positives = 116/149 (77%)

Query: 8   FDKVKPIVMKLRRNYFVQLWEYDDWIQEGRIVLFRLLEEHPYLLDNESKLFIYFKTKFSN 67
           F+KVKPI++KL+R+Y++QLW+ DDW+QEG I+L +LLE +P L++ E +L+ YFKTKFS+
Sbjct: 41  FEKVKPIILKLKRHYYIQLWDRDDWLQEGHIILLQLLERYPELIEEEERLYRYFKTKFSS 100

Query: 68  YLNDVLRHQDCQKRQFNKMPYEEISEVSHYVKSKGLVLDDYIAYRDTLTKVEETLSDIDK 127
           YL D+LR Q+ QKRQF+K+ YEEI EV+H + S+GL LDDY+AY++ +  +E  L+  ++
Sbjct: 101 YLKDLLRRQESQKRQFHKLAYEEIGEVAHAIPSRGLWLDDYVAYQEVIASLENQLNSQER 160

Query: 128 EKFEKLISGERFAGKKQFIRDIQPFFNAF 156
            +F+ LI GERF G++  +R I P+F  F
Sbjct: 161 MQFQALIRGERFKGRRALLRKISPYFKEF 189
```

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

Example 2222

A DNA sequence (GBSx2341) was identified in *S.agalactiae* <SEQ ID 6855> which encodes the amino acid sequence <SEQ ID 6856>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.23   Transmembrane  166 - 182 ( 166 - 182)

----- Final Results -----
              bacterial membrane --- Certainty=0.1893(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:CAA99510 GB:Z75191 ORF YOR283w [Saccharomyces cerevisiae]
 Identities = 57/226 (25%), Positives = 97/226 (42%), Gaps = 22/226 (9%)

Query: 4   VRLYIARHGKTMFNTIGRAQGWSDTPLTTFGELGIKELGLGLKASNISFKEAFSSDSGRT 63
           +RL+I RHG+T  N     QG  DT +   GE    +LG  L++  I F +  SSD  R
Sbjct: 17  IRLFIIRHGQTEHNVKKILQGHKDTSINPTGEEQATKLGHYLRSRGIHFDKVVSSDLKRC 76
```

```
Query: 64  LQTMEIILREVQQENIPYTRDKRIREWCFGSLDGGYDGDLFNGVLPRVSNGDMSHLTHEE 123
           QT  ++L+  +QEN+P +    +RE   G ++G                  M     E+
Sbjct: 77  RQTTALVLKHSKQENVPTSYTSGLRERYMGVIEG-----------------MQITEAEK 118

Query: 124 IANLICQVDTAGWAEPWAILSNRILSGFTAIAKKIEDIGGGNAIVVSHGMTIATFL-WL- 181
            A+   +     + E       R+         +  + G  N  +VSHG  I   L WL
Sbjct: 119 YADKHGEGSFRNFGEKSDDFVARLTGCVEEEVAEASNEGVKNLALVSHGGAIRMILQWLK 178

Query: 182 IDHSTPRSLGLDNGSVSVVDF--EDGTFSIQSIGDMSYREKGREIL 225
           ++     + + N SV++VD+  +   F ++ +G+  +   G  ++
Sbjct: 179 YENHQAHKIIVFNTSVTIVDYVKDSKQFIVRRVGNTQHLGDGEFVV 224
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6857> which encodes the amino acid sequence <SEQ ID 6858>. Analysis of this protein sequence reveals the following:

```
      Possible site: 57

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.69   Transmembrane  170 - 186 ( 170 - 186)

----- Final Results -----
               bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAA99510 GB:Z75191 ORF YOR283w [Saccharomyces cerevisiae]
 Identities = 64/231 (27%), Positives = 98/231 (41%), Gaps = 27/231 (11%)

Query: 5   RLYIARHGKTMFNTIGRAQGWSDTPLTKKGEEGIRELGLGLKDATIPFKAAFSSDSGRTM 64
           RL+I RHG+T  N     QG  DT +   GEE   +LG  L+   I F    SSD  R
Sbjct: 18  RLFIIRHGQTEHNVKKILQGHKDTSINPTGEEQATKLGHYLRSRGIHFDKVVSSDLKRCR 77

Query: 65  QTIEIILRESENEFLPYTKDNRIREWCFGSLEGTYDSELFLGVLPRTKAFENRDNLRDVP 124
           QT  ++L+ S+ E +P +  + +RE   G +EG   +E
Sbjct: 78  QTTALVLKHSKQENVPTSYTSGLRERYMGVIEGMQITEA---------------------- 116

Query: 125 YSELAESIVEVDTANWAEPWEVLRKRIWEGFEAIALSIQNAGGGNALVVSHGMTIGTFL- 183
            + A+    E    N+ E  +    R+    E        N G  N  +VSHG  I   L
Sbjct: 117 -EKYADKHGEGSFRNFGEKSDDFVARLTGCVEEEVAEASNEGVKNLALVSHGGAIRMILQ 175

Query: 184 WL--IDPDRDKQYIDNGSVTVVEF--DDGQFTIKTIGDMSYRYRGREIIEE 230
           WL   +    K  + N SVT+V++  D  QF ++ +G+  +   G  ++ +
Sbjct: 176 WLKYENHQAHKIIVFNTSVTIVDYVKDSKQFIVRRVGNTQHLGDGEFVVSD 226
```

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

```
 Identities = 150/231 (64%), Positives = 182/231 (77%), Gaps = 5/231 (2%)

Query: 1   MSKVRLYIARHGKTMFNTIGRAQGWSDTPLTTFGELGIKELGLGLKASNISFKEAFSSDS 60
           M+K RLYIARHGKTMFNTIGRAQGWSDTPLT  GE GI+ELGLGLK + I FK AFSSDS
Sbjct: 1   MTKTRLYIARHGKTMFNTIGRAQGWSDTPLTKKGEEGIRELGLGLKDATIPFKAAFSSDS 60

Query: 61  GRTLQTMEIILREVQQENIPYTRDKRIREWCFGSLDGGYDGDLFNGVLPRV----SNGDM 116
           GRT+QT+EIILRE + E +PYT+D RIREWCFGSL+G YD +LF GVLPR     +  ++
Sbjct: 61  GRTMQTIEIILRESENEFLPYTKDNRIREWCFGSLEGTYDSELFLGVLPRTKAFENRDNL 120

Query: 117 SHLTHEEIANLICQVDTAGWAEPWAILSNRILSGFTAIAKKIEDIGGGNAIVVSHGMTIA 176
            + + E+A  I +VDTA WAEPW +L  RI  GF AIA  I++ GGGNA+VVSHGMTI
Sbjct: 121 RDVPYSELAESIVEVDTANWAEPWEVLRKRIWEGFEAIALSIQNAGGGNALVVSHGMTIG 180

Query: 177 TFLWLIDHSTPRSLGLDNGSVSVVDFEDGTFSIQSIGDMSYREKGREILEK 227
           TFLWLID     +DNGSV+VV+F+DG F+I++IGDMSYR +GREI+E+
Sbjct: 181 TFLWLIDPDRDKQY-IDNGSVTVVEFDDGQFTIKTIGDMSYRYRGREIIEE 230
```

A related GBS gene <SEQ ID 8999> and protein <SEQ ID 9000> were also identified. Analysis of this protein sequence reveals the following:

```
Cytoplasmic predicted but experimentally found on the surface of Streptococci

32.3/52.0% over 184aa
                                        Thermotoga maritima
  EGAD|165681| phosphoglycerate mutase Insert characterized
  GP|4981935|gb|AAD36444.1|AE001791_6|AE001791 phosphoglycerate mutase Insert characterized
  PIR|G72260|G72260 phosphoglycerate mutase - (strain MSB8) Insert characterized

ORF01265(268 - 870 of 1248)
EGAD|165681|TM1374(1 - 185 of 201) phosphoglycerate mutase (Thermotoga maritima)
GP|4981935|gb|AAD36444.1|AE001791_6|AE001791 phosphoglycerate mutase (Thermotoga maritima)
PIR|G72260|G72260 phosphoglycerate mutase - Thermotoga maritima (strain MSB8)
%Match = 6.3
%Identity = 32.2  %Similarity = 52.0
Matches = 57  Mismatches = 78  Conservative Sub.s = 35

105       135       165       195       225       255       285       315
RGRNNSYEIFNPFSMLLKRINRFYFCSR*LQNFFIGKVR*YIPVKAFVFCYNIIKCL*GVSMSKVRLYIARHGKTMFNTI
                                                            ::||: |||:|::|
                                                           MKLYLIRHGETIWNEK
                                                                  10

345       375       405       435       465       495       519       549
GRAQGWSDTPLTTFGELGIKELGLGLKASNISFKEAFSSDSXRTLQTMEIILREVQQENI--PYTRDKRIREWCFGSLDG
|  || :| ||   |    ::|   ||          :||   |:|:| | | |  ::| |      |:  |  |
GLWQGVTDVPLNERGREQARKLANSLK----RVDAIYSSPLKRSLETAEEIARRFEKEIIVEEDLRECEISLW-------
             30        40            50        60        70        80

579       609       639       669                   699       729       759
GYDGDLFNGVLPRVSNGDMSHLTHEEIANLICQVDTAGWA----------EPWAILSNRILSGFTAIAKKIEDIGGGNAI
                  : || || |     |:   |:          |    : ||::     |  : :  |  | :
------------------NGLTVEE-AIREYPVEFKKWSSDPNFGMEGLESMRNVQNRVVKAIMKIVSQEKLNGSENVV
                     90        100       110       120       130       140

789       816       840       870       900       930       960       990
VVSHGMTIATFL-WLIDHST--PRSLGLDNGSVSVVDFEDGTFSIQSIGDMSYREKGREILEKTLQ*KKIKLSDSV*LVF
:||| :::  |: |::       |:: ||| |:|||: |     :       :|
IVSHSLSLRAFICWILGLPLYLHRNFKLDNASLSVVEIESKPRLVLLNDTCHLKES
             160       170       180       190       200
```

SEQ ID 9000 (GBS44) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 6; MW 27kDa), in Figure 168 (lane 8-10; MW 42kDa – thioredoxin fusion) and in Figure 238 (lane 7; MW 42kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 12 (lane 8; MW 52.4kDa).

Purified Thio-GBS44-His is shown in Figure 244, lanes 7 & 8.

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

Example 2223

A DNA sequence (GBSx2342) was identified in *S.agalactiae* <SEQ ID 6859> which encodes the amino acid sequence <SEQ ID 6860>. This protein is predicted to be d-alanyl-d-alanine carboxypeptidase. 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 homology with the following sequences in the GENPEPT database.

```
>GP:AAD00280 GB:U78599 putative D,D-carboxypeptidase [Streptococcus mutans]
 Identities = 108/169 (63%), Positives = 139/169 (81%)

Query: 79  ELSPDVVPVENIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEM 138
           E++PDV  ++ + +D RI +   +FL AA+ IDS EHLISGYRSVAYQE+L+N+Y+ QE
Sbjct: 4   EMNPDVTDIDGVKVDSRIAENTRKFLAAAQEIDSSEHLISGYRSVAYQEELYNNYIAQEK 63

Query: 139 TSNPNLTRGQAEKLVKTYSQPAGASEHQTGLAMDMSTVDSLNESDPRVVSQLKKIAPQYG 198
            +NP+L++ +A+K V+TYSQP G+SEHQTGLA+DMSTVDSLN+SD  VV+++  IAP+YG
Sbjct: 64  ANNPSLSQEEAQKQVQTYSQPPGSSEHQTGLAIDMSTVDSLNQSDANVVAKVAAIAPKYG 123

Query: 199 FVLRFPDGKTAETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKE 247
           FVLRFP+GK   TG+ YEDWHYRYVGV+SAKYM KH LTLEEY+  LKE
Sbjct: 124 FVLRFPEGKKDATGIDYEDWHYRYVGVKSAKYMTKHDLTLEEYLKKLKE 172
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6861> which encodes the amino acid sequence <SEQ ID 6862>. Analysis of this protein sequence reveals the following:

```
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.66   Transmembrane   10 -  26 (   3 -  29)

----- Final Results -----
              bacterial membrane --- Certainty=0.4864(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:AAD00280 GB:U78599 putative D,D-carboxypeptidase [Streptococcus mutans]
 Identities = 118/173 (68%), Positives = 139/173 (80%)

Query: 74  ITKEMSPELADINGISVDKRIEQATSDFLAAAQAIDLQEHLISGYRSVDYQTELYQSYIK 133
           IT EM+P++ DI+G+ VD RI + T  FLAAAQ ID  EHLISGYRSV YQ ELY +YI
Sbjct: 1   ITAEMNPDVTDIDGVKVDSRIAENTRKFLAAAQEIDSSEHLISGYRSVAYQEELYNNYIA 60

Query: 134 KEMANDPTLTQEAAEALVQTYSQPPGASEHHTGLAIDMSTVDTLNASDPSVAKAVQKIAP 193
           +E AN+P+L+QE A+  VQTYSQPPG+SEH TGLAIDMSTVD+LN SD +V   V  IAP
Sbjct: 61  QEKANNPSLSQEEAQKQVQTYSQPPGSSEHQTGLAIDMSTVDSLNQSDANVVAKVAAIAP 120

Query: 194 DYGFVLRFPEGKKTSTGVDYEDWHYRYVGKASARYMAQHNLTLEEYIAALKEK 246
            YGFVLRFPEGKK +TG+DYEDWHYRYVG  SA+YM +H+LTLEEY+  LKEK
Sbjct: 121 KYGFVLRFPEGKKDATGIDYEDWHYRYVGVKSAKYMTKHDLTLEEYLKKLKEK 173
```

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

```
 Identities = 131/235 (55%), Positives = 172/235 (72%), Gaps = 3/235 (1%)

Query: 15  LLAILCF--SLFALLKPNSQQSSSQKLRNEDIKKISSQKRNKKLQLPAVSSKDWNLILVN 72
           LL ++ F   L+  +KP    + +Q L  ++I++   +K ++   LP VS +DW L+LVN
Sbjct: 12  LLIVIVFLGGLYLFIKPEESVTPTQ-LNKKEIQQKDIKKTDRLRALPKVSVEDWELVLVN 70

Query: 73  RDHKHEELSPDVVPVENIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNS 132
           RDH  +E+SP++  +  I +DKRI +  + FL AA+AID +EHLISGYRSV YQ +L+ S
Sbjct: 71  RDHITKEMSPELADINGISVDKRIEQATSDFLAAAQAIDLQEHLISGYRSVDYQTELYQS 130

Query: 133 YVTQEMTSNPNLTRGQAEKLVKTYSQPAGASEHQTGLAMDMSTVDSLNESDPRVVSQLKK 192
           Y+ +EM ++P LT+  AE LV+TYSQP G SEH TGLA+DMSTVD+LN SDP V   ++K
Sbjct: 131 YIKKEMANDPTLTQEAAEALVQTYSQPPGASEHHTGLAIDMSTVDTLNASDPSVAKAVQK 190

Query: 193 IAPQYGFVLRFPDGKTAETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKE 247
```

```
          IAP YGFVLRFP+GK   TGV YEDWHYRYVG  SA+YMA+H+LTLEEYI  LKE
Sbjct: 191 IAPDYGFVLRFPEGKKTSTGVDYEDWHYRYVGKASARYMAQHNLTLEEYIAALKE 245
```

A related GBS gene <SEQ ID 9001> and protein <SEQ ID 9002> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:      14.03
GvH: Signal Score (-7.5): -1.02
     Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:  10.08 threshold:  0.0
   PERIPHERAL  Likelihood = 10.08      56
 modified ALOM score:  -2.52

*** 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:

```
33.7/55.1% over 183aa

Enterococcus faecalis
  EGAD|41322| d-alanyl-d-alanine carboxypeptidase Insert characterized
GP|1209528|gb|AAB05624.1||U35369 D,D-carboxypeptidase Insert characterized

ORF01266(484 - 1038 of 1350)
EGAD|41322|43646(85  -  268  of  268)  d-alanyl-d-alanine  carboxypeptidase  {Enterococcus
faecalis}SP|Q47746|VANY_ENTFA D-ALANYL-D-ALANINE
 CARBOXYPEPTIDASE          (EC          3.4.16.4)          (DD-PEPTIDASE)          (DD-
CARBOXYPEPTIDASE).GP|1209528|gb|AAB05624.1||U35369 D,D-carboxypeptidase {Enteroco
ccus faecalis}
%Match = 10.1
%Identity = 33.7  %Similarity = 55.1
Matches = 63  Mismatches = 79  Conservative Sub.s = 40

234       264       294       324       354       384       414       444
SR*F*RWNIFYSIYWGYVLSRKRKRNFRKNIAMKKNKIIRFSLVGVLLAILCFSLFALLKPNSQQSSSQKLRNEDIKKIS

          MEKSNYHSNVNHHKRHMKQSGEKRAFLWAFIISFTVCTLFLGWRLVSVLEATQLPPIPATHTGSGTGVAEN
                   10        20        30        40        50        60        70

474       504       531       561       588       618       648       678
SQKRNKKLQLPAVSSKDWNLILVNRDHK-HEELSPDVVPVEN-IYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQ
 :             ::|:|||||| :    :   ::  : |   :| ||:     :::|||       : ||||:   |
PEENTLATAKEQGDEQEWSLILVNRQNPIPAQYDVELEQLSNGERIDIRISPYLQDLFDAARADGVYPIVASGYRTTEKQ
                   90        100       110       120       130       140       150

708       738       768       798       828       858       888       918
EKLFNSYVTQEMTSNPNLTRGQAEKLVKTYSQPAGASEHQTGLAMDMSTVDSLNESDPRVVSQLKKIAPQYGFVLRFPDG
::: :  |  |  :    |  ||:   :|:    | |||| |||:|::  | :: :   |   | : : ::||: |:|
QEIMDEKV-AEYKAK-GYTSAQAKAEAETWVAVPGTSEHQLGLAVDINA-DGIHSTGNEVYRWLDENSYRFGFIRRYPPD
       160       170       180       190       200       210       220

948       978       1008      1038      1068      1098      1128      1158
KTAETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQ*GNVFPC*ILLLLLLFSFSLFFFRF*TIREK*MLIL
||  |||  | |||||||:|:|  :  : | ||||:   |
KTEITGVSNEPWHYRYVGIEAATKIYHQGLCLEEYLNTEK
```

SEQ ID 6860 (GBS18) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 3; MW 31kDa).

The GBS18-His fusion product was purified (Figure 93A; see also Figure 189, lane 11) and used to immunise mice (lane 2 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 93B), FACS (Figure 93C), 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.

Example 2224

A DNA sequence (GBSx2343) was identified in *S.agalactiae* <SEQ ID 6863> which encodes the amino acid sequence <SEQ ID 6864>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.58   Transmembrane   10 -  26 (   3 -  29)

----- 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>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6865> which encodes the amino acid sequence <SEQ ID 6866>. Analysis of this protein sequence reveals the following:

```
     Possible site: 33
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.83   Transmembrane   10 -  26 (   4 -  33)

----- Final Results -----
               bacterial membrane --- Certainty=0.5734(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAD00279 GB:U78599 putative N-acetyl-muramidase [Streptococcus mutans]
 Identities = 66/150 (44%), Positives = 97/150 (64%), Gaps = 5/150 (3%)

Query: 18  LLLIVCPLLSSQRIASADKEVRVNYSQKQFITKMGKEVKPLAKYYGIRPSILIAQILLET 77
           LL+I+ P+L+S  +A A+K++    YS K+F+ ++     + L+K YG+R SI+I Q  L++
Sbjct: 3   LLVILLPILASGGLADANKKMPSPYSHKEFVKEIAPTAQKLSKIYGVRSSIIIGQAALDS 62

Query: 78  HDGKTLLASKYHNLFSKKATPGQVAITLKSPKQTN---QNV--RYAIYKDDASAIRDYLR 132
           H G TLLASKYHNLFS +A+PGQ A+ LKS +  N   Q V  RY +Y+     ++ DY+
Sbjct: 63  HFGSTLLASKYHNLFSIEASPGQGAVRLKSHEYKNGRWQEVTNRYLVYESWKESLYDYMA 122

Query: 133 MLRQGKEVDKRLYRNLATEKGYKAPAKSLQ 162
           +L   K  DK LY  + T  GYK  A++LQ
Sbjct: 123 ILHGNKIWDKALYTTMMTSSGYKTVARALQ 152
```

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

```
 Identities = 67/190 (35%), Positives = 102/190 (53%), Gaps = 1/190 (0%)

Query: 1   MRKRFSLLNFIVVTFIFFFFILFPLLNHKGKVDANSRQSVTYTKEEFIQKIVPDAQDLGK 60
           MRKR     F+ +   F   I+ PLL+ +    A+    V Y++++FI K+  + + L K
Sbjct: 1   MRKRLKFPYFLTLLACFLLLIVCPLLSSQRIASADKEVRVNYSQKQFITKMGKEVKPLAK 60

Query: 61  SYGIRPSFIIAQAALDSDFGEKILANKYHNLFGLLAEPGTPSITLNDSSTGKKQEKQFTH 120
            YGIRPS +IAQ  L++  G+ +LA+KYHNLF   A PG  +ITL  S     Q  ++
Sbjct: 61  YYGIRPSILIAQILLETHDGKTLLASKYHNLFSKKATPGQVAITLK-SPKQTNQNVRYAI 119

Query: 121 YKSWKYSMYDYLAHIKSGATGKKDSYTIMVSVKNPKTLVQKLQDSGFDNDKKYAKKMTEI 180
           YK    ++ DYL  ++ G    K  Y  + + K  K    + LQ      DK YA+++ ++
```

```
Sbjct: 120 YKDDASAIRDYLRMLRQGKEVDKRLYRNLATEKGYKAPAKSLQKYLHYTDKTYARRLIQV 179

Query: 181 IDLYDLTRYD 190
           I+  DLT YD
Sbjct: 180 IESNDLTNYD 189
```

SEQ ID 6864 (GBS246) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 7; MW 24.6kDa).

GBS246d was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 154 (lanes 14 & 15; MW 21kDa) and in Figure 183 (lane 4; MW 21kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 187 (lane 12; MW 46kDa). Purified GBS246d-GST is shown in Figure 243, lane 12.

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

Example 2225

A DNA sequence (GBSx2344) was identified in *S.agalactiae* <SEQ ID 6867> which encodes the amino acid sequence <SEQ ID 6868>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2541(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:AAC45610 GB:U78296 repressor of class I heat shock gene
          expression HrcA [Streptococcus mutans]
 Identities = 227/345 (65%), Positives = 287/345 (82%), Gaps = 1/345 (0%)

Query: 17  VITQRQNDILNLIVELFTQTHEPVGSKALQRTIDSSSATIRNDMAKLEKLGLLEKAHTSS 76
           +ITQRQ DILNLIVELFT+THEP+GSK LQ +I SS ATIRNDMA LEKLGLLEKA T
Sbjct: 1   MITQRQKDILNLIVELFTKTHEPIGSKTLQNSIASSRATIRNDMAALEKLGLLEKATTPP 60

Query: 77  GRM-PSPAGFKYFVEHSLRLDSIDEQDIYHVIKAFDFEAFKLEDMLQKASHILSEMTGYT 135
             + P      +YFVEHSL  DS+DEQD+Y VIKAFDFEAF+L D+LQ+AS +L+ +TGYT
Sbjct: 61  AVVCPVKKAIRYFVEHSLNPDSLDEQDVYQVIKAFDFEAFRLGDLLQRASDVLANLTGYT 120

Query: 136 SVILDVEPARQRLTGFDVVQLSNHDALAVMTLDESKPVTVQFAIPRNFLTRDLIAFKAIV 195
           ++ILDVEP +QRLT FD+V+LSNHDALAV+TLDE+ PVTVQFAIP+NFL  DL+    I
Sbjct: 121 ALILDVEPKKQRLTTFDIVKLSNHDALAVLTLDEASPVTVQFAIPKNFLDSDLMTVAKIA 180

Query: 196 EERLLDGSVMDIHYKLRTEIPQIVQKYFVTTDNVLQLFDYVFSELFLETVFVAGKVNSLT 255
            ER L+ +V+DIHY+LRTE PQI+QKYF  TDNVL LFD++F+ +F E VF++GK+ +L
Sbjct: 181 RERFLNQTVLDIHYRLRTEPPQIIQKYFPRTDNVLDLFDHIFNPIFQEEVFISGKIKTLE 240

Query: 256 YSDLSTYQFLDNEQQVAISLRQSLKEGEMASVQVADSQEAALADVSVLTHKFLIPYRGFG 315
           ++ L TYQFL+N Q VA+ +RQSL E E+  VQVADS+E +LAD++V++ KFLIPYRGFG
Sbjct: 241 FAGLDTYQFLENLQSVALEIRQSLPEDELHRVQVADSKEKSLADLTVISQKFLIPYRGFG 300

Query: 316 LLSLIGPIDMDYRRSVSLVNIIGKVLAAKLGDYYRYLNSNHYEVH 360
           +L++IGP+D+DY+R++SL+N+I +VLA KLGD+YRYLNSNHYEVH
Sbjct: 301 ILTVIGPVDLDYQRTISLINVISRVLAVKLGDFYRYLNSNHYEVH 345
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6869> which encodes the amino acid sequence <SEQ ID 6870>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0695(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 = 341/344 (99%), Positives = 343/344 (99%)

Query: 17  VITQRQNDILNLIVELFTQTHEPVGSKALQRTIDSSSATIRNDMAKLEKLGLLEKAHTSS 76
           VITQRQNDILNLIVELFTQTHEPVGSKALQRTIDSSSATIRNDMAKLEKLGLLEKAHTSS
Sbjct: 1   VITQRQNDILNLIVELFTQTHEPVGSKALQRTIDSSSATIRNDMAKLEKLGLLEKAHTSS 60

Query: 77  GRMPSPAGFKYFVEHSLRLDSIDEQDIYHVIKAFDFEAFKLEDMLQKASHILSEMTGYTS 136
           GRMPSPAGFKYFVEHSLRLDSIDEQDIYHVIKAFDFEAFKLEDMLQKASHIL+EMTGYTS
Sbjct: 61  GRMPSPAGFKYFVEHSLRLDSIDEQDIYHVIKAFDFEAFKLEDMLQKASHILAEMTGYTS 120

Query: 137 VILDVEPARQRLTGFDVVQLSNHDALAVMTLDESKPVTVQFAIPRNFLTRDLIAFKAIVE 196
           VILDVEPARQRLTGFDVVQLSNHDALAVMTLDESKPVTVQFAIPRNFLTRDLIAFKAIVE
Sbjct: 121 VILDVEPARQRLTGFDVVQLSNHDALAVMTLDESKPVTVQFAIPRNFLTRDLIAFKAIVE 180

Query: 197 ERLLDGSVMDIHYKLRTEIPQIVQKYFVTTDNVLQLFDYVFSELFLETVFVAGKVNSLTY 256
           ERLLD SV+DIHYKLRTEIPQIVQKYFVTTDNVLQLFDYVFSELFLETVFVAGKVNSLTY
Sbjct: 181 ERLLDNSVIDIHYKLRTEIPQIVQKYFVTTDNVLQLFDYVFSELFLETVFVAGKVNSLTY 240

Query: 257 SDLSTYQFLDNEQQVAISLRQSLKEGEMASVQVADSQEAALADVSVLTHKFLIPYRGFGL 316
           SDLSTYQFLDNEQQVAISLRQSLKEGEMASVQVADSQEAALADVSVLTHKFLIPYRGFGL
Sbjct: 241 SDLSTYQFLDNEQQVAISLRQSLKEGEMASVQVADSQEAALADVSVLTHKFLIPYRGFGL 300

Query: 317 LSLIGPIDMDYRRSVSLVNIIGKVLAAKLGDYYRYLNSNHYEVH 360
           LSLIGPIDMDYRRSVSLVNIIGKVLAAKLGDYYRYLNSNHYEVH
Sbjct: 301 LSLIGPIDMDYRRSVSLVNIIGKVLAAKLGDYYRYLNSNHYEVH 344
```

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

Example 2226

A DNA sequence (GBSx2345) was identified in *S.agalactiae* <SEQ ID 6871> which encodes the amino acid sequence <SEQ ID 6872>. This protein is predicted to be grpe protein (grpE). Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5138(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:AAC45611 GB:U78296 GrpE [Streptococcus mutans]
 Identities = 130/180 (72%), Positives = 151/180 (83%), Gaps = 3/180 (1%)

Query: 14  VSEEIKKDDLQEEVEATE--TEETVEEVIEEIPEKSELELANERADEFENKYLRAHAEM- 70
           +S++ KK++ +EEVEATE  TEE+VEEV EE  E  EL+ A ERA++FENKYLRAHAEM
Sbjct: 1   MSKKDKKEEYKEEVEATEPTTEESVEEVAEETSENKELQEALERAEDFENKYLRAHAEMP 60

Query: 71  QNIQRRSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGLEMTRDSLIQA 130
           +       +  +  QRYRSQDL KAILPSLDNLERALAVEGLTDDVKKGLEM ++SLIQA
Sbjct: 61  KTFSVALMKSDKVCQRYRSQDLRKAILPSLDNLERALAVEGLTDDVKKGLEMVQESLIQA 120
```

```
Query: 131 LKEEGVEEVEVDSFDHNFHMAVQTLPADDEHPADSIAEVFQKGYKLHERLLRPAMVVVYN 190
           LKEEGVEEVE+++FD N HMAVQTL ADD+HPADSIA+V QKGY+LHERLLRPAMVVVYN
Sbjct: 121 LKEEGVEEVELENFDANLHMAVQTLDADDDHPADSIAQVHQKGYQLHERLLRPAMVVVYN 180
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6873> which encodes the amino acid sequence <SEQ ID 6874>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5138(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 = 189/190 (99%), Positives = 189/190 (99%)

Query: 1   MAVFNKLFKRRHSVSEEIKKDDLQEEVEATETEETVEEVIEEIPEKSELELANERADEFE 60
           MAVFNKLFKRRHSVSEEIKKDDLQEEVEATETEETVEEVIEE PEKSELELANERADEFE
Sbjct: 1   MAVFNKLFKRRHSVSEEIKKDDLQEEVEATETEETVEEVIEETPEKSELELANERADEFE 60

Query: 61  NKYLRAHAEMQNIQRRSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGL 120
           NKYLRAHAEMQNIQRRSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGL
Sbjct: 61  NKYLRAHAEMQNIQRRSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGL 120

Query: 121 EMTRDSLIQALKEEGVEEVEVDSFDHNFHMAVQTLPADDEHPADSIAEVFQKGYKLHERL 180
           EMTRDSLIQALKEEGVEEVEVDSFDHNFHMAVQTLPADDEHPADSIAEVFQKGYKLHERL
Sbjct: 121 EMTRDSLIQALKEEGVEEVEVDSFDHNFHMAVQTLPADDEHPADSIAEVFQKGYKLHERL 180

Query: 181 LRPAMVVVYN 190
           LRPAMVVVYN
Sbjct: 181 LRPAMVVVYN 190
```

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

Example 2227

A DNA sequence (GBSx2346) was identified in *S.agalactiae* <SEQ ID 6875> which encodes the amino acid sequence <SEQ ID 6876>. This protein is predicted to be heat shock protein 70 (dnaK). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0996(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6877> which encodes the amino acid sequence <SEQ ID 6878>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0996(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 = 594/609 (97%), Positives = 603/609 (98%), Gaps = 1/609 (0%)

Query: 1   MSKIIGIDLGTTNSAVAVLEGTESKIIANPEGNRTTPSVVSFKNGEIIVGDAAKRQAVTN 60
            MSKIIGIDLGTTNSAVAVLEGTESKIIANPEGNRTTPSVVSFKNGEIIVGDAAKRQAVTN
Sbjct: 1   MSKIIGIDLGTTNSAVAVLEGTESKIIANPEGNRTTPSVVSFKNGEIIVGDAAKRQAVTN 60

Query: 61  PDTVISIKSKMGTSEKVSANGKEYTPQEISAMILQYLKGYAEDYLGEKVEKAVITVPAYF 120
            P+TVISIKSKMGTSEKVSANGKEYTPQEISAMILQYLKGYAEDYLGEKVEKAVITVPAYF
Sbjct: 61  PETVISIKSKMGTSEKVSANGKEYTPQEISAMILQYLKGYAEDYLGEKVEKAVITVPAYF 120

Query: 121 NDAQRQATKDAGKIAGLEVERIVNEPTAAALAYGMDKTDKDEKILVFDLGGGTFDVSILE 180
            NDAQRQATKDAGKIAGLEVERIVNEPTAAALAYGMDKTDKDEKILVFDLGGGTFDVSILE
Sbjct: 121 NDAQRQATKDAGKIAGLEVERIVNEPTAAALAYGMDKTDKDEKILVFDLGGGTFDVSILE 180

Query: 181 LGDGVFDVLATAGDNKLGGDDFDQKIIDFLVEEFKKENGIDLSQDKMALQRLKDAAEKAK 240
            LGDGVFDVLATAGDNKLGGDDFDQKIIDFLV EFKKENGIDLSQDKMALQRLKDAAEKAK
Sbjct: 181 LGDGVFDVLATAGDNKLGGDDFDQKIIDFLVAEFKKENGIDLSQDKMALQRLKDAAEKAK 240

Query: 241 KDLSGVTQTQISLPFITAGSAGPLHLEMSLSRAKFDDLTRDLVERTKTPVRQALSDAGLS 300
            KDLSGVTQTQISLPFITAGSAGPLHLEMSLSRAKFDDLTRDLVERTKTPVRQALSDAGLS
Sbjct: 241 KDLSGVTQTQISLPFITAGSAGPLHLEMSLSRAKFDDLTRDLVERTKTPVRQALSDAGLS 300

Query: 301 LSEIDEVILVGGSTRIPAVVEAVKAETGKEPNKSVNPDEVVAMGAAIQGGVITGDVKDVV 360
            LSEIDEVILVGGSTRIPAVVEAVKAETGKEPNKSVNPDEVVAMGAAIQGGVITGDVKDVV
Sbjct: 301 LSEIDEVILVGGSTRIPAVVEAVKAETGKEPNKSVNPDEVVAMGAAIQGGVITGDVKDVV 360

Query: 361 LLDVTPLSLGIETMGGVFTKLIDRNTTIPTSKSQVFSTAADNQPAVDIHVLQGERPMAAD 420
            LLDVTPLSLGIETMGGVFTKLIDRNTTIPTSKSQVFSTAADNQPAVDIHVLQGERPMAAD
Sbjct: 361 LLDVTPLSLGIETMGGVFTKLIDRNTTIPTSKSQVFSTAADNQPAVDIHVLQGERPMAAD 420

Query: 421 NKTLGRFQLTDIPAAPRGIPQIEVTFDIDKNGIVSVKAKDLGTQKEQHIVIQSNSGLTDE 480
            NKTLGRFQLTDIPAAPRGIPQIEVTFDIDKNGIVSVKAKDLGTQKEQHIVI+SN GL++E
Sbjct: 421 NKTLGRFQLTDIPAAPRGIPQIEVTFDIDKNGIVSVKAKDLGTQKEQHIVIKSNDGLSEE 480

Query: 481 EIDKMMKDAEANAEADAKRKEEVDLKNEVDQAIFATEKTIKETEGKGFDTERDAAQSALD 540
            EID+MMKDAEANAEADAKRKEEVDLKNEVDQAIFATEKTIKETEGKGFDTERDAAQSALD
Sbjct: 481 EIDRMMKDAEANAEADAKRKEEVDLKNEVDQAIFATEKTIKETEGKGFDTERDAAQSALD 540

Query: 541 ELKKAQESGNLDDMKAKLEALNEKAQALAVKLYEQAAAAQQAAQGAEGAQSADSSSKGDD 600
            ELK AQESGNLDDMKAKLEALNEKAQALAVK+YEQAAAAQQAAQGAEGAQ+ DS++  DD
Sbjct: 541 ELKAAQESGNLDDMKAKLEALNEKAQALAVKMYEQAAAAQQAAQGAEGAQANDSAN-NDD 599

Query: 601 VVDGEFTEK 609
            VVDGEFTEK
Sbjct: 600 VVDGEFTEK 608
```

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

Example 2228

A DNA sequence (GBSx2347) was identified in *S.agalactiae* <SEQ ID 6879> which encodes the amino acid sequence <SEQ ID 6880>. This protein is predicted to be Streptococcus pneumoniae DnaJ protein homologue (dnaJ). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4180(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6881> which encodes the amino acid sequence <SEQ ID 6882>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1322(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 = 330/377 (87%), Positives = 357/377 (94%), Gaps = 1/377 (0%)

Query: 1   MNNTEFYDRLGVSKDASQDEIKKAYRRMSKKYHPDINKETGAEEKYKEVQEAYETLSDTQ 60
           MNNTE+YDRLGVSKDASQD+IKKAYR+MSKKYHPDINKE GAE+KYK+VQEAYETLSD+Q
Sbjct: 19  MNNTEYYDRLGVSKDASQDDIKKAYRKMSKKYHPDINKEAGAEQKYKDVQEAYETLSDSQ 78

Query: 61  KRAAYDQYGAAGANGGFGGFDGGGFGGFDGGGFGGFEDIFSSFFGGGGMRNPNAPRQGDD 120
           KRAAYDQYGAAGA GGFGG   GGFGGFDGGGFGGFEDIFSSFFGGGG RNPNAPRQGDD
Sbjct: 79  KRAAYDQYGAAGAQGGFGG-GAGGFGGFDGGGFGGFEDIFSSFFGGGGSRNPNAPRQGDD 137

Query: 121 LQYRVNLSFEEAIFGAEKEVSYNRESSCHTCSGSGAKPGTSPVTCQKCHGSGVINVDTQT 180
           LQYRVNLSFEEA+FG EKEVSYNRE++C TC GSGAKPGT+PVTC+KCHGSGV+ +DTQT
Sbjct: 138 LQYRVNLSFEEAVFGVEKEVSYNREATCGTCLGSGAKPGTAPVTCRKCHGSGVMTIDTQT 197

Query: 181 PLGTMRRQVTCDVCQGSGQEIKEKCPTCHGTGHEKKTHKVSVKIPAGVETGQQIRLTGQG 240
           PLG MRRQVTCD+C GSG+EIKE C TCHGTGHEK+ HKVSVKIPAGVETGQQIRL GQG
Sbjct: 198 PLGMMRRQVTCDICHGSGKEIKEPCQTCHGTGHEKQAHKVSVKIPAGVETGQQIRLQGQG 257

Query: 241 EAGFNGGPYGDLFVIINVLPSQQFERNGSTIYYTLNISFVQAALGDTIDIPTVHGAVEMS 300
           EAGFNGGPYGDLFVI+NVLPS+QFERNGSTIYY L+ISF QAALGDT++IPTVHG VEM+
Sbjct: 258 EAGFNGGPYGDLFVILNVLPSKQFERNGSTIYYNLDISFTQAALGDTVEIPTVHGDVEMA 317

Query: 301 IPAGTQTGKTFRLRGKGAPKLRGGGQGDQHVTVNIVTPTKLNDAQKEALHAFAEASGDKM 360
           IPAGTQTGKTFRL+GKGAPKLRGGGQGDQHVTVNIVTPTKLNDAQ+EAL AFAEASG+KM
Sbjct: 318 IPAGTQTGKTFRLKGKGAPKLRGGGQGDQHVTVNIVTPTKLNDAQREALQAFAEASGEKM 377

Query: 361 VHPKKKGFFDKVKDALD 377
           +HPKKKGFFDKVKDAL+
Sbjct: 378 LHPKKKGFFDKVKDALE 394
```

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

Example 2229

A DNA sequence (GBSx2348) was identified in *S.agalactiae* <SEQ ID 6883> which encodes the amino acid sequence <SEQ ID 6884>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22   Transmembrane  281 - 297 ( 281 - 297)

----- 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:AAD24445 GB:AF118389 unknown [Streptococcus suis]
 Identities = 182/373 (48%), Positives = 257/373 (68%), Gaps = 5/373 (1%)
```

```
Query: 4   KVEEIRSYLIASIQNGKLAPGDRLPSIRQLANQFSCNKDTVQRVLMELRFDNYIYAKPRS 63
           K + I   ++  I+  +   G++LPSIRQL  Q+ C+KDTVQ+ ++EL++ N IYA  +S
Sbjct: 3   KYQVIIQDILTGIEEHRFKRGEKLPSIRQLREQYHCSKDTVQKAMLELKYQNKIYAVEKS 62

Query: 64  GYYVFDSHQEEVEEGVSLPNSEIANIAYDDFRLCLNETLIGREDYLFNYYYRQEGLLDLS 123
           GYY+ +   + +      +  ++ I Y+DFR+CL E+LIGRE+YLFNYY++QEGL +L
Sbjct: 63  GYYILEDRDFQ-DHTCRAQSYRLSRITYEDFRICLKESLIGRENYLFNYYHQQEGLAELI 121

Query: 124 KAVAKLMEETGVYVPLDDIVITAGTQQALFILTQVTFPNRKSRVLIEEPTYPRMIELIKT 183
            +V  L+ +  VY   D +VITAG+QQAL+ILTQ+      K+ +LIE PTY RMIELI+
Sbjct: 122 SSVQSLLMDYHVYTKKDQLVITAGSQQALYILTQMETLAGKTEILIENPTYSRMIELIRH 181

Query: 184 QNLPYETISRGTHGIDFQRLEEIFQTQSIKFFYVIPRMHNPLGTSYNPVEMKRLIEMAEK 243
           Q +PY+TI R   GID + LE IFQT  IKFFY IPR+HNPLG++Y+       ++++A++
Sbjct: 182 QGIPYQTIERNLDGIDLEELESIFQTGKIKFFYTIPRLHNPLGSTYDIATKTAIVKLAKQ 241

Query: 244 YDVYIVEDDYMSDFASQS--PLHYYDTHGRVIYLKSFSKAIFPALRLAAICLPQALKSTF 301
           YDVYI+EDDY++DF S    PLHY DT  RVIY+KSF+  +FPALR+ AI LP  L+  F
Sbjct: 242 YDVYIIEDDYLADFDSSHSLPLHYLDTDNRVIYIKSFTPTLFPALRIGAISLPNQLRDIF 301

Query: 302 MAYKKLMDYDTNLILQKALALYIENGLYAKNSQYLKYRYQKDLANSKSILADHP-NLPSY 360
           + +K L+DYDTNLI+QKAL+LYI+NG++A+N+Q+L + Y       K  L  +  N+P Y
Sbjct: 302 IKHKSLIDYDTNLIMQKALSLYIDNGMFARNTQHLHHIYHAQWNKIKDCLEKYALNIP-Y 360

Query: 361 SLHHDSVLFDCSK 373
            +   SV F  SK
Sbjct: 361 RIPKGSVTFQLSK 373
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6885> which encodes the amino acid sequence <SEQ ID 6886>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3043(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 = 176/382 (46%), Positives = 255/382 (66%), Gaps = 7/382 (1%)

Query: 1   MVTKVEEIRSYLIASIQNGKLAPGDRLPSIRQLANQFSCNKDTVQRVLMELRFDNYIYAK 60
           M TK + I S +   IQ  +L  GD+LPSIR L+  + C+KDTVQR L+EL++ + IYA
Sbjct: 1   MTTKYQTIISNIEQDIQKQRLKKGDKLPSIRVLSKVYYCSKDTVQRALLELKYRHLIYAV 60

Query: 61  PRSGYYVFDSHQEEVEEGVSLPNSEIANIAYDDFRLCLNETLIGREDYLFNYYYRQEGLL 120
           P+SGYYV   +   + ++L   +  N+AY+DFRLCLNE L  ++ YLF+YY++ EGL
Sbjct: 61  PKSGYYVL-GNVSMPDNVLNLSLEDYNNMAYEDFRLCLNEALSAKDKYLFHYYHKTEGLE 119

Query: 121 DLSKAVAKLMEETGVYVPLDDIVITAGTQQALFILTQVTFPNRKSRVLIEEPTYPRMIEL 180
           +L +A+   + E  VY   D ++IT+GTQQAL+IL+Q+ FPN    +L+E+PTY RM  +
Sbjct: 120 ELREALLLYLAENSVYSNKDQLLITSGTQQALYILSQMPFPNTGKTILLEKPTYHRMEAI 179

Query: 181 IKTQNLPYETISRGTHGIDFQRLEEIFQTQSIKFFYVIPRMHNPLGTSYNPVEMKRLIEM 240
           +   LPY+TISR  +G+D + LE +FQT  IKFFY I R  +PLG SY+  E + ++ +
Sbjct: 180 VAQLGLPYQTISRHFNGLDLELLESLFQTGDIKFFYTISRFSHPLGLSYSTKEKEAIVRL 239

Query: 241 AEKYDVYIVEDDYMSDFA--SQSPLHYYDTHGRVIYLKSFSKAIFPALRLAAICLPQALK 298
           A++Y VYI+EDDY+ DF     P+HYYDTH R+IYLKSFS ++FPALR+ A+ LP  LK
Sbjct: 240 AQRYQVYILEDDYLGDFVKLKEPPIHYYDTHHRIIYLKSFSMSVFPALRIGALVLPSGLK 299

Query: 299 STFMAYKKLMDYDTNLILQKALALYIENGLYAKNSQYLKYRYQKDLANSKSILADHPNLP 358
             F+  K L+D DTNL++QKALALY+ENG++ KN +++K RY K      ++      N P
Sbjct: 300 PHFLTQKSLIDLDTNLLMQKALALYLENGMFQKNLRFIK-RYLKQRERQLALFLKQ-NCP 357
```

```
Query: 359 S--YSLHHDSVLFDCSKLDNFK 378
             Y L    ++ D +  D+++
Sbjct: 358 DIHYQLTPTHLVIDYTTSDSYR 379
```

SEQ ID 6884 (GBS423) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 7; MW 49.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 2; MW 74kDa).

GBS423-GST was purified as shown in Figure 219, lane 2-3.

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

Example 2230

A DNA sequence (GBSx2349) was identified in *S.agalactiae* <SEQ ID 6887> which encodes the amino acid sequence <SEQ ID 6888>. This protein is predicted to be pseudouridylate synthase I (truA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3265(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:BAB03886 GB:AP001507 tRNA pseudouridine synthase A
          (pseudouridylate synthase I) [Bacillus halodurans]
 Identities = 105/240 (43%), Positives = 147/240 (60%), Gaps = 2/240 (0%)

Query: 1   MTRYKAQISYDGSAFSGFQRQPNCRTVQEEIERTLKRLNSGNDVIIHGAGRTDVGVHAYG 60
           M R    +++YDG+ F+G+Q QPN RTVQ E+E  LK ++ G  + +  +GRTD GVHA G
Sbjct: 1   MKRIGLKVAYDGTDFAGYQIQPNERTVQGELESVLKNIHKGMSIRVTASGRTDTGVHARG 60

Query: 61  QVIHFDLPQARDVEKLRFGLDTQCPDDIDIVKVEQVSDDFHCRYDKHIKTYEFLVDIGRP 120
           Q++HFD    +  V++     L++Q P DI +++    V  DFH RY     K Y + V
Sbjct: 61  QIVHFDTSLSFPVDRWPIALNSQLPADICVLEAADVPADFHARYSAKTKEYRYRVLTSAQ 120

Query: 121 KNPMMRNYATHYPYPVIIELMQEAIKDLVGTHDFTGFTASGTSVENKVRTIFDAKIQFEA 180
            +   RNY  H  YP+ +E MQ A   L+GTHDF+ F A+   VE+KVRTI D  +  E
Sbjct: 121 ADVFRRNYTYHVRYPLDVEAMQRAAVQLLGTHDFSSFCAAKAEVEDKVRTIEDVALWREG 180

Query: 181 SKNLLIFTFTGNGFLYKQVRNMVGTLLKIGNGRMPISQIKTILQAKNRDLAGPTAAGNGL 240
           + LIF+  GNGFLY  VR +VGTLL+IG G+    ++  IL A++R+ AG TA G+GL
Sbjct: 181 DE--LIFSIRGNGFLYNMVRIIVGTLLEIGAGKRSAEEVAKILAARSREAAGKTAPGHGL 238
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6889> which encodes the amino acid sequence <SEQ ID 6890>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2558(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 = 184/249 (73%), Positives = 214/249 (85%)

Query: 1   MTRYKAQISYDGSAFSGFQRQPNCRTVQEEIERTLKRLNSGNDVIIHGAGRTDVGVHAYG 60
           M RYKA ISYDG+ FSGFQRQ + RTVQEEIE+TL +LN+G  +IIHGAGRTD GVHAYG
Sbjct: 1   MVRYKATISYDGTLFSGFQRQRHLRTVQEEIEKTLYKLNNGTKIIIHGAGRTDAGVHAYG 60

Query: 61  QVIHFDLPQARDVEKLRFGLDTQCPDDIDIVKVEQVSDDFHCRYDKHIKTYEFLVDIGRP 120
           QVIHFDLPQ ++VEKLRF LDTQ P+DID+V +E+V+DDFHCRY KH+KTYEFLVD GRP
Sbjct: 61  QVIHFDLPQEQEVEKLRFALDTQTPEDIDVVNIEKVADDFHCRYQKHLKTYEFLVDNGRP 120

Query: 121 KNPMMRNYATHYPYPVIIELMQEAIKDLVGTHDFTGFTASGTSVENKVRTIFDAKIQFEA 180
           KNPMMR+Y THYPY + I+LMQEAI  LVGTHDFTGFTA+GTSV+NKVRTI  A +  +
Sbjct: 121 KNPMMRHYTTHYPYTLNIKLMQEAINGLVGTHDFTGFTAAGTSVQNKVRTITKATVSRDE 180

Query: 181 SKNLLIFTFTGNGFLYKQVRNMVGTLLKIGNGRMPISQIKTILQAKNRDLAGPTAAGNGL 240
             + L+FTF+GNGFLYKQVRNMVGTLLKIGNG+MP+ Q+K IL +KNR LAGPT +GNGL
Sbjct: 181 KTDFLVFTFSGNGFLYKQVRNMVGTLLKIGNGQMPVEQVKVILSSKNRQLAGPTISGNGL 240

Query: 241 YLKEIIYED 249
           YLKEI YE+
Sbjct: 241 YLKEICYEN 249
```

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

Example 2231

A DNA sequence (GBSx2350) was identified in *S.agalactiae* <SEQ ID 6891> which encodes the amino acid sequence <SEQ ID 6892>. This protein is predicted to be phosphomethypyrimidine kinase (thiD). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2051(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:CAB15828 GB:Z99123 phosphomethylpyrimidine kinase [Bacillus subtilis]
 Identities = 95/253 (37%), Positives = 150/253 (58%), Gaps = 13/253 (5%)

Query: 1   MKTRNVLAISGNDIFSGGGLHADLATYVVNKLHGFVAVTCLTAMSDKG---FEVIPIEAS 57
           M     L I+G+D   G G+ ADL T+    ++G  A+T + AM        +V PI+
Sbjct: 1   MSMHKALTIAGSDSSGGAGIQADLKTFQEKNVYGMTALTVIVAMDPNNSWNHQVFPIDTD 60

Query: 58  ILKQQLESLKD-VEFGSIKLGLLPNVETAQVVLEFVKSKQECPVVLDPVLVCKENHDL-- 114
            ++ QL ++ D +   ++K G+LP V+  ++  + +K KQ   VV+DPV+VCK  +++
Sbjct: 61  TIRAQLATITDGIGVDAMKTGMLPTVDIIELAAKTIKEKQLKNVVIDPVMVCKGANEVLY 120

Query: 115 --EVSQLREQLIAFFPYADVITPNLVEAQLLTGLS-IENLDQMKIAAEKLYDMGAKHVVI 171
                LREQL    P A VITPNL EA  L+G+  ++ +D M  AA+K++ +GA++VVI
Sbjct: 121 PEHAQALREQLA---PLATVITPNLFEASQLSGMDELKTVDDMIEAAKKIHALGAQYVVI 177

Query: 172 KGGNRLNAEEATDLYYDGERFETYVFPVVDANNT-GAGCTFASSIASQLAMGKNVEDAVK 230
            GG +L  E+A D+ YDGE  E      ++D   T GAGCTF+++  ++LA G  V++A+
Sbjct: 178 TGGGKLKHEKAVDVLYDGETAEVLESEMIDTPYTHGAGCTFSAAVTAELAKGAEVKEAIY 237

Query: 231 MSKGFVYQAIKAS 243
           +K F+  AIK S
Sbjct: 238 AAKEFITAAIKES 250
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4407> which encodes the amino acid sequence <SEQ ID 4408>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2029(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 = 135/252 (53%), Positives = 174/252 (68%)

Query: 1   MKTRNVLAISGNDIFSGGGLHADLATYVVNKLHGFVAVTCLTAMSDKGFEVIPIEASILK 60
           MKT  ++ ISGNDI SGGGL+ADLATY+   L  FVAVTCLT  S++GF + P+   I +
Sbjct: 1   MKTDYIVTISGNDILSGGGLYADLATYIRYDLQAFVAVTCLTTRSEEGFSLFPVAKEIFR 60

Query: 61  QQLESLKDVEFGSIKLGLLPNVETAQVVLEFVKSKQECPVVLDPVLVCKENHDLEVSQLR 120
            QL S  +    +IK+GLLPN E  ++VL+F+K     PVVLDPVL CKE  D+++  LR
Sbjct: 61  DQLNSFTNAPISAIKIGLLPNAEMCEIVLDFIKGHLGIPVVLDPVLACKEIDDVKIVPLR 120

Query: 121 EQLIAFFPYADVITPNLVEAQLLTGLSIENLDQMKIAAEKLYDMGAKHVVIKGGNRLNAE 180
           ++++   PY  V+TPNLVEAQLL+   I +L  M+ AA+  Y +GAK VVIKGGNR + +
Sbjct: 121 QEILQLLPYVTVVTPNLVEAQLLSQKEIVSLKDMQEAAKYFYQLGAKQVVIKGGNRFSQK 180

Query: 181 EATDLYYDGERFETYVFPVVDANNTGAGCTFASSIASQLAMGKNVEDAVKMSKGFVYQAI 240
           +A DL+YDG+   T   PV++ NN GAGCTFASSIASQL   K   +AVK SK  VYQAI
Sbjct: 181 KAIDLFYDGKEIVTLECPVLEKNNIGAGCTFASSIASQLVKKKTPLEAVKNSKELVYQAI 240

Query: 241 KASDKYGVVQHF 252
             SD+YGV Q +
Sbjct: 241 LQSDRYGVKQSY 252
```

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

Example 2232

A DNA sequence (GBSx2351) was identified in *S.agalactiae* <SEQ ID 6893> which encodes the amino acid sequence <SEQ ID 6894>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.05   Transmembrane   97 - 113 (  96 - 119)
   INTEGRAL    Likelihood = -0.22   Transmembrane   54 -  70 (  54 -  70)

----- Final Results -----
            bacterial membrane --- Certainty=0.3421(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:BAA30952 GB:AP000007 202aa long hypothetical protein [Pyrococcus
          horikoshii]
 Identities = 48/148 (32%), Positives = 78/148 (52%), Gaps = 9/148 (6%)

Query: 10  VQLAIVTAISIVLGMFISIPTPTGFLTLLDAGIFFAAFYFGKKEGAVVGALAGFLIDLLK 69
           V  A+VTA+++V+   I IP   G+L   D I   +  FG   G   G +     DLL
Sbjct: 49  VMAALVTAMTMVIR--IPIPASQGYLNFGDIMIMLTSVLFGPLVGGFAGGVGSAFADLL- 105

Query: 70  GYPNWMFFSLLIHGTQGYLAGLPGR------RRLLGLISATLVMVLGYAIASGLMYGWGA 123
           GYP+W  F+L+I GT+G + G   +       + LLG +    VMV+GY   + ++YG
```

```
Sbjct: 106 GYPSWALFTLVIKGTEGIIVGYFSKGEANYGKILLGTVLGGSVMVIGYVSVAYVLYGPAG 165

Query: 124 VLPDIPGNIMQNMVGMVVGFALNKSLER 151
           + ++  +I+Q + G+V+G  L   L++
Sbjct: 166 AIGELYNDIVQAVSGIVIGGGLGYILKK 193
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6895> which encodes the amino acid sequence <SEQ ID 6896>. Analysis of this protein sequence reveals the following:

```
     Possible site: 54

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -4.62   Transmembrane   98 - 114 (  97 - 119)
   INTEGRAL    Likelihood = -0.00   Transmembrane  135 - 151 ( 135 - 151)

----- 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 databases:

```
>GP:CAB49310 GB:AJ248284 hypothetical protein [Pyrococcus abyssi]
 Identities = 42/145 (28%), Positives = 73/145 (49%), Gaps = 10/145 (6%)

Query: 7   RQMSLTGILTALVVVLGRFVMLPTPT--GFLTLLDAGIYAVSFSFGSAQGAIVGGLSGFL 64
           R ++++ +  ALV  +   + +P P   G+L   D  I  V+  FG   G   GG+   +
Sbjct: 39  RTVAISAVAAALVTAMTMVIRIPIPASQGYLNFGDIMIMLVAVLFGPLVGGFAGGVGSAI 98

Query: 65  IDLVAGYPQWMFHSLIAHSVQGYFAGWRGR------KRWLGVVIGSFIMIFWYFLGSLML 118
            DL+ GYP W   +LI    +G   G+  +      K  +G V+G FIM+  Y   S +L
Sbjct: 99  ADLI-GYPSWALFTLIIKGSEGLVVGYFSKGEPNYSKILIGTVLGGFIMVLGYVSVSYVL 157

Query: 119 GYGLSGSLAGIWGNVMQNTLGLFVG 143
           YG +G+++  ++ + +Q   G+ +G
Sbjct: 158 -YGPAGAISELYNDTVQAVSGIVIG 181
```

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

```
 Identities = 77/155 (49%), Positives = 106/155 (67%), Gaps = 1/155 (0%)

Query: 1   MRKEKTSQLVQLAIVTAISIVLGMFISIPTPTGFLTLLDAGIFFAAFYFGKKEGAVVGAL 60
           M+  K  Q+    I+TA+ +VLG F+ +PTPTGFLTLLDAGI+  +F FG  +GA+VG L
Sbjct: 1   MQNSKIRQMSLTGILTALVVVLGRFVMLPTPTGFLTLLDAGIYAVSFSFGSAQGAIVGGL 60

Query: 61  AGFLIDLLKGYPNWMFFSLLIHGTQGYLAGLPGRRRLLGLISATLVMVLGYAIASGLM-Y 119
           +GFLIDL+ GYP WMF SL+ H  QGY AG  GR+R LG++   + +M+  Y + S ++ Y
Sbjct: 61  SGFLIDLVAGYPQWMFHSLIAHSVQGYFAGWRGRKRWLGVVIGSFIMIFWYFLGSLMLGY 120

Query: 120 GWGAVLPDIPGNIMQNMVGMVVGFALNKSLERVKK 154
           G    L  I GN+MQN +G+ VGF + K++ R KK
Sbjct: 121 GLSGSLAGIWGNVMQNTLGLFVGFIIFKAILRQKK 155
```

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

Example 2233

A DNA sequence (GBSx2352) was identified in *S.agalactiae* <SEQ ID 6897> which encodes the amino acid sequence <SEQ ID 6898>. Analysis of this protein sequence reveals the following:

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

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

```
bacterial cytoplasm --- Certainty=0.0881(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:CAB15708 GB:Z99122 alternate gene name: ipc-33d [Bacillus subtilis]
 Identities = 91/176 (51%), Positives = 115/176 (64%)

Query: 6   NKLKQETKAIVVDIIERSALKKGQIFVLGLSSSEVSGGLIGKNSSSEIGEIIVEVILKEL 65
           N+LKQ  K ++ +  +++ LK+ Q+FVLG S+SEV+G  IG + S +I E I   + +
Sbjct: 2   NELKQTWKTMLSEFQDQAELKQDQLFVLGCSTSEVAGSRIGTSGSVDIAESIYSGLAELR 61

Query: 66  HSRGIYLAVQGCEHVNRALVVEAELAERQQLEVVNVVPNLHAGGSGQVAAFKLMTSPVEV 125
              GI+LA Q CEH+NRALVVEAE A+  +L  V+ VP   AGG+    AFK M SPV V
Sbjct: 62  EKTGIHLAFQCCEHLNRALVVEAETAKLFRLPTVSAVPVPKAGGAMASYAFKQMKSPVLV 121

Query: 126 EEIVAHAGIDIGDTSIGMHIKRVQVPLIPISRELGGAHVTALASRPKLIGGARAGY 181
           E I A AGIDIGDT IGMH+K V VP+      LG AHVT   +RPKLIGG RA Y
Sbjct: 122 ETIQADAGIDIGDTFIGMHLKPVAVPVRVSQNSLGSAHVTLARTRPKLIGGVRAVY 177
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6899> which encodes the amino acid sequence <SEQ ID 6900>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2166(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 = 132/183 (72%), Positives = 161/183 (87%)

Query: 6   NKLKQETKAIVVDIIERSALKKGQIFVLGLSSSEVSGGLIGKNSSSEIGEIIVEVILKEL 65
           N L+++T+ IV+D++ERSA++ G +FVLGLSSSE+ G  IGK SS E+G+I+VEV+L EL
Sbjct: 3   NNLEKQTREIVIDVVERSAIQPGNLFVLGLSSSEILGSRIGKQSSLEVGQIVVEVVLDEL 62

Query: 66  HSRGIYLAVQGCEHVNRALVVEAELAERQQLEVVNVVPNLHAGGSGQVAAFKLMTSPVEV 125
           + RG++LAVQGCEHVNRALVVE  +AE +QLE+VNVVPNLHAGGS Q+AAF+LM+ PVEV
Sbjct: 63  NKRGVHLAVQGCEHVNRALVVERHVAESKQLEIVNVVPNLHAGGSAQMAAFQLMSDPVEV 122

Query: 126 EEIVAHAGIDIGDTSIGMHIKRVQVPLIPISRELGGAHVTALASRPKLIGGARAGYTSDP 185
           EE++AHAG+DIGDT+IGMHIKRVQ+PLIP  RELGGAHVTALASRPKLIGGARA Y  D
Sbjct: 123 EEVIAHAGLDIGDTAIGMHIKRVQIPLIPCQRELGGAHVTALASRPKLIGGARADYNMDI 182

Query: 186 IRK 188
           IRK
Sbjct: 183 IRK 185
```

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

Example 2234

A DNA sequence (GBSx2353) was identified in *S.agalactiae* <SEQ ID 6901> which encodes the amino acid sequence <SEQ ID 6902>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.25   Transmembrane   21 -  37 (  13 -  46)
   INTEGRAL    Likelihood = -4.30   Transmembrane   78 -  94 (  76 - 113)
   INTEGRAL    Likelihood = -2.07   Transmembrane   96 - 112 (  95 - 113)
```

```
----- 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:BAB06385 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 105/261 (40%), Positives = 150/261 (57%), Gaps = 2/261 (0%)

Query: 12  NVEEVLFTFFTKLIS--ILLLIIAFVIVRQVINYLFEKTVNRSLAFSRQKVARQKTLAKL 69
           N+    F   T +I+  +L+ +IAF+IVR +   +   + R    ++    R  TL KL
Sbjct: 7   NITSGAFLASTFIIAGKVLVAVIAFLIVRAIGKRIISNSFARMAKNNQLSSGRVVTLEKL 66

Query: 70  SHNVLNYTLYFFLFYWILSILGVPISSLLAGAGIAGVAIGLGAQGFLSDVVNGFFILLEN 129
           S N  +YTL F     +L+I G+  S+L+AGAGI G+AIG GAQG +SD+V GFFILLE
Sbjct: 67  SLNAFSYTLMFIFATTLLTIFGLNPSALIAGAGIVGLAIGFGAQGLVSDIVTGFFILLEK 126

Query: 130 QFDVGDIINVGTVSGTVTNVGIRTTQIHDFDGTLHFIPNRNITIVSNKSRSNMRAQIDIP 189
           Q DVGD +  G V G V  VG+RT  I  FDGTLH+IPNRNI  VSN SR NMRA +DI
Sbjct: 127 QIDVGDYVTAGGVDGIVEEVGLRTALIRGFDGTLHYIPNRNIANVSNHSRGNMRALVDIS 186

Query: 190 LFVHTNLDQISDIVTKINEEYVSKHPAIVGEPTVFGPTTNANGQFVYRINIFTQNGAQFD 249
           +  + N+D+   ++ K+ ++   +  I+  P V G     +  V RI   T+N  Q+
Sbjct: 187 ISYNDNIDEAISVMQKVCDQLAEQDERIIEGPDVIGVQNLGDSDVVIRIIAKTENMEQWS 246

Query: 250 IYAEFYKLYQKAILEEGIDLP 270
           +     K  ++A+    I++P
Sbjct: 247 VERLLRKQLKEALEAHNIEIP 267
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6903> which encodes the amino acid sequence <SEQ ID 6904>. Analysis of this protein sequence reveals the following:

```
     Possible site: 54
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.49   Transmembrane   24 -  40 (  15 -  45)
   INTEGRAL    Likelihood = -4.83   Transmembrane   78 -  94 (  73 -  99)
   INTEGRAL    Likelihood = -2.07   Transmembrane   96 - 112 (  95 - 113)

----- 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:BAB06385 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 104/249 (41%), Positives = 151/249 (59%), Gaps = 4/249 (1%)

Query: 22  KKLVSLIILLLFFAILKRVTNYLFEKTINKSFAYSRQSEARKKTLSKLTHNILNYLLYFL 81
           K LV++I  L+  AI KR+ +  F +    +   + S  R  TL KL+ N  +Y L F+
Sbjct: 23  KVLVAVIAFLIVRAIGKRIISNSFARMAKNN----QLSSGRVVTLEKLSLNAFSYTLMFI 78

Query: 82  LIYWILSLFGIPVSSLLAGAGIAGVAIGLGAQGFLSDVVNGFFILFENQFEVGDNVTISD 141
               +L++FG+  S+L+AGAGI G+AIG GAQG +SD+V GFFIL E Q +VGD VT
Sbjct: 79  FATTLLTIFGLNPSALIAGAGIVGLAIGFGAQGLVSDIVTGFFILLEKQIDVGDYVTAGG 138

Query: 142 IEGSVFGVGIRTTQIRGFDGTLHFIPNRSITVVSNKSRGNMRALIEIPLYSTVNLSQVTR 201
           ++G V  VG+RT  IRGFDGTLH+IPNR+I  VSN SRGNMRAL++I +    N+ +
Sbjct: 139 VDGIVEEVGLRTALIRGFDGTLHYIPNRNIANVSNHSRGNMRALVDISISYNDNIDEAIS 198

Query: 202 IIDEVNQKELPNHPQIVGKPNILGPQNNSNGQFTFRIAIFTENGEQFKIYHTFYRLYQEA 261
           ++ +V  +      +I+  P+++G QN  +     RI   TEN EQ+ +     +  +EA
Sbjct: 199 VMQKVCDQLAEQDERIIEGPDVIGVQNLGDSDVVIRIIAKTENMEQWSVERLLRKQLKEA 258

Query: 262 LLKEGIQLP 270
```

```
            L    I++P
Sbjct: 259 LEAHNIEIP 267
```

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

```
 Identities = 164/265 (61%), Positives = 215/265 (80%)

Query: 7   FIDHLNVEEVLFTFFTKLISILLLIIAFVIVRQVINYLFEKTVNRSLAFSRQKVARQKTL 66
           +++  ++E +  T F KL+S+++L++ F I+++V NYLFEKT+N+S A+SRQ  AR+KTL
Sbjct: 7   YLEQSHIENIGLTIFKKLVSLIILLLFFAILKRVTNYLFEKTINKSFAYSRQSEARKKTL 66

Query: 67  AKLSHNVLNYTLYFFLFYWILSILGVPISSLLAGAGIAGVAIGLGAQGFLSDVVNGFFIL 126
           +KL+HN+LNY LYF L YWILS+ G+P+SSLLAGAGIAGVAIGLGAQGFLSDVVNGFFIL
Sbjct: 67  SKLTHNILNYLLYFLLIYWILSLFGIPVSSLLAGAGIAGVAIGLGAQGFLSDVVNGFFIL 126

Query: 127 LENQFDVGDIINVGTVSGTVTNVGIRTTQIHDFDGTLHFIPNRNITIVSNKSRSNMRAQI 186
            ENQF+VGD + +  + G+V  VGIRTTQI  FDGTLHFIPNR+IT+VSNKSR NMRA I
Sbjct: 127 FENQFEVGDNVTISDIEGSVFGVGIRTTQIRGFDGTLHFIPNRSITVVSNKSRGNMRALI 186

Query: 187 DIPLFVHTNLDQISDIVTKINEEYVSKHPAIVGEPTVFGPTTNANGQFVYRINIFTQNGA 246
           +IPL+   NL Q++ I+ ++N++ +  HP IVG+P + GP  N+NGQF +RI IFT+NG
Sbjct: 187 EIPLYSTVNLSQVTRIIDEVNQKELPNHPQIVGKPNILGPQNNSNGQFTFRIAIFTENGE 246

Query: 247 QFDIYAEFYKLYQKAILEEGIDLPT 271
           QF IY  FY+LYQ+A+L+EGI LPT
Sbjct: 247 QFKIYHTFYRLYQEALLKEGIQLPT 271
```

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

Example 2235

A DNA sequence (GBSx2354) was identified in *S.agalactiae* <SEQ ID 6905> which encodes the amino acid sequence <SEQ ID 6906>. This protein is predicted to be RopA (tig). Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1785(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 9283> which encodes amino acid sequence <SEQ ID 9284> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6907> which encodes the amino acid sequence <SEQ ID 6908>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0776(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 = 303/354 (85%), Positives = 337/354 (94%)

Query: 1   MSTSFENKATNRGIITFTISQDEIKPALDQAFNKVKKDLNVPGFRKGHMPRTVFNQKFGE 60
```

```
            MSTSFENKATNRG+ITFTISQD+IKPALD+AFNK+KKDLN PGFRKGHMPR VFNQKFGE
Sbjct: 30   MSTSFENKATNRGVITFTISQDKIKPALDKAFNKIKKDLNAPGFRKGHMPRPVFNQKFGE 89

Query: 61   EALYENALNLVLPKAYEAAVAELGLDVVAQPKIDVVSMEKGQDWKLTAEVVTKPEVKLGD 120
             E LYE+ALN+VLP+AYEAAV ELGLDVVAQPKIDVVSMEKG++W L+AEVVTKPEVKLGD
Sbjct: 90   EVLYEDALNIVLPEAYEAAVTELGLDVVAQPKIDVVSMEKGKEWTLSAEVVTKPEVKLGD 149

Query: 121  YKDLSVEVDASKEVSDEEVDAKVERERNNLAELTVKDGEAAQGDTVVIDFVGSVDGVEFD 180
             YK+L VEVDASKEVSDE+VDAK+ERER NLAEL +KDGEAAQGDTVVIDFVGSVDGVEFD
Sbjct: 150  YKNLVVEVDASKEVSDEDVDAKIERERQNLAELIIKDGEAAQGDTVVIDFVGSVDGVEFD 209

Query: 181  GGKGDNFSLELGSGQFIPGFEEQLVGSKAGQTVDVNVTFPEDYQAEDLAGKDAKFVTTIH 240
             GGKGDNFSLELGSGQFIPGFE+QLVG+KAG  V+VNVTFPE YQAEDLAGK AKF+TTIH
Sbjct: 210  GGKGDNFSLELGSGQFIPGFEDQLVGAKAGDEVEVNVTFPESYQAEDLAGKAAKFMTTIH 269

Query: 241  EVKTKEVPALDDELAKDIDDEVETLDELKAKYRKELESAKEIAFDDAVEGAAIELAVANA 300
             EVKTKEVP LDDELAKDID++V+TL++LK KYRKELE+A+E A+DDAVEGAAIELAVANA
Sbjct: 270  EVKTKEVPELDDELAKDIDEDVDTLEDLKVKYRKELEAAQETAYDDAVEGAAIELAVANA 329

Query: 301  EIVELPEEMVHDEVHRAMNEFMGNMQRQGISPEMYFQLTGTTEEDLHKQYQADA 354
             EIV+LPEEM+H+EV+R++NEFMGNMQRQGISPEMYFQLTGTT+EDLH QY A+A
Sbjct: 330  EIVDLPEEMIHEEVNRSVNEFMGNMQRQGISPEMYFQLTGTTQEDLHNQYSAEA 383
```

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

Example 2236

A DNA sequence (GBSx2355) was identified in *S.agalactiae* <SEQ ID 6909> which encodes the amino acid sequence <SEQ ID 6910>. This protein is predicted to be galactose-6-phosphate isomerase laca subunit (rpiB). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3491(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:AAA25177 GB:M60447 galactose 6-P isomerase [Lactococcus lactis]
 Identities = 92/141 (65%), Positives = 115/141 (81%)

Query: 1    MTIIIGADAHGVELKEVIRQHLTSIGKEIIDLTDTSKDFVDNTLAIVAKVNQKEDNLGIM 60
             M I++GAD  G  LK+V++  L   G E+ID+T   +DFVD TLA+ ++VN+ E NLGI+
Sbjct: 1    MAIVVGADLKGTRLKDVVKNFLVEEGFEVIDVTKDGQDFVDVTLAVASEVNKDEQNLGIV 60

Query: 61   VDAYGVGPFMVATKVKGMIAAEVSDERSAYMTRAHNNARMITLGSEIVGPGVAKHIVEGF 120
             +DAYG GPFMVATK+KGM+AAEVSDERSAYMTR HNNARMIT+G+EIVG  +AK+I + F
Sbjct: 61   IDAYGAGPFMVATKIKGMVAAEVSDERSAYMTRGHNNARMITVGAEIVGDELAKNIAKAF 120

Query: 121  VDGTYDAGRHQIRVDMLNKMC 141
             V+G YD GRHQ+RVDMLNKMC
Sbjct: 121  VNGKYDGGRHQVRVDMLNKMC 141
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6911> which encodes the amino acid sequence <SEQ ID 6912>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

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

```
bacterial cytoplasm --- Certainty=0.3224(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 = 101/140 (72%), Positives = 117/140 (83%)

Query: 1   MTIIIGADAHGVELKEVIRQHLTSLGKEIIDLTDTSKDFVDNTLAIVAKVNQKEDNLGIM 60
            M II+GADAHG  LKE+I+  L   G +IID+TD + DF+DNTLA+   VN+ E  LGIM
Sbjct: 1   MAIILGADAHGNALKELIKSFLQEEGYDIIDVTDINSDFIDNTLAVAKAVNEAEGRLGIM 60

Query: 61  VDAYGVGPFMVATKVKGMIAAEVSDERSAYMTRAHNNARMITLGSEIVGPGVAKHIVEGF 120
            VDAYG GPFMVATK+KGM+AAEVSDERSAYMTR HNNARMIT+G+EIVGP +AK+IV+GF
Sbjct: 61  VDAYGAGPFMVATKLKGMVAAEVSDERSAYMTRGHNNARMITIGAEIVGPELAKNIVKGF 120

Query: 121 VDGTYDAGRHQIRVDMLNKM 140
            V G YD GRHQIRVDMLNKM
Sbjct: 121 VTGPYDGGRHQIRVDMLNKM 140
```

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

Example 2237

A DNA sequence (GBSx2356) was identified in *S.agalactiae* <SEQ ID 6913> which encodes the amino acid sequence <SEQ ID 6914>. This protein is predicted to be galactose-6-phosphate isomerase lacb subunit (rpiB). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2511(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 10189> which encodes amino acid sequence <SEQ ID 10190> was also identified.

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

```
>GP:AAA25178 GB:M60447 galactose 6-P isomerase [Lactococcus lactis]
 Identities = 138/171 (80%), Positives = 157/171 (91%)

Query: 10  MKIAVGCDHIVTYDKIAVVDYLKTKGYEVIDCGTYDNIRTHYPIYGKKVGEAVASGKADL 69
            M+IA+GCDHIVT  K+AV ++LK+KGYEV+D GTYD++RTHYPIYGKKVGEAV SG+ADL
Sbjct: 1   MRIAIGCDHIVTDVKMAVSEFLKSKGYEVLDFGTYDHVRTHYPIYGKKVGEAVVSGQADL 60

Query: 70  GVCICGTGVGINNAVNKVPGIRSALVRDLTSAIYAKEELNANVIGFGGKITGGLLMTDII 129
            GVCICGTGVGINNAVNKVPG+RSALVRD+TSA+YAKEELNANVIGFGG ITGGLLM DII
Sbjct: 61  GVCICGTGVGINNAVNKVPGVRSALVRDMTSALYAKEELNANVIGFGGMITGGLLMNDII 120

Query: 130 EAFIRAKYKPTKENKVLIEKIAEVETHNAHQEENDFFTEFLDKWNRGEYHD 180
            EAFI A+YKPT+ENK LI KI  VETHNAHQ + +FFTEFL+KW+RGEYHD
Sbjct: 121 EAFIEAEYKPTEENKKLIAKIEHVETHNAHQADEEFFTEFLEKWDRGEYHD 171
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6915> which encodes the amino acid sequence <SEQ ID 6916>. Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3048(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 = 136/171 (79%), Positives = 160/171 (93%)

Query: 10  MKIAVGCDHIVTYDKIAVVDYLKTKGYEVIDCGTYDNIRTHYPIYGKKVGEAVASGKADL 69
           MKIA+GCDHIVT +K+AV D+LK+KGY+VIDCGTYD+ RTHYPI+GKKVGEAV +G+ADL
Sbjct: 2   MKIAIGCDHIVTNEKMAVSDFLKSKGYDVIDCGTYDHTRTHYPIFGKKVGEAVVNGQADL 61

Query: 70  GVCICGTGVGINNAVNKVPGIRSALVRDLTSAIYAKEELNANVIGFGGKITGGLLMTDII 129
           GVCICGTGVGINNAVNKVPGIRSALVRD+T+A+YAKEELNANVIGFGGKITG LLM DII
Sbjct: 62  GVCICGTGVGINNAVNKVPGIRSALVRDMTTALYAKEELNANVIGFGGKITGELLMCDII 121

Query: 130 EAFIRAKYKPTKENKVLIEKIAEVETHNAHQEENDFFTEFLDKWNRGEYHD 180
           +AFI+A+YK T+ENK LI KIA +E+H+A+QE+ DFFTEFL+KW+RGEYHD
Sbjct: 122 DAFIKAEYKETEENKKLIAKIAHLESHHANQEDPDFFTEFLEKWDRGEYHD 172
```

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

Example 2238

A DNA sequence (GBSx2357) was identified in *S.agalactiae* <SEQ ID 6917> which encodes the amino acid sequence <SEQ ID 6918>. 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>
```

A related GBS nucleic acid sequence <SEQ ID 10187> which encodes amino acid sequence <SEQ ID 10188> was also identified.

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

```
>GP:AAA25179 GB:M60447 tagatose 6-P kinase [Lactococcus lactis]
 Identities = 192/310 (61%), Positives = 236/310 (75%)

Query: 11  MILTVTLNPSIDISYCLENFNMDTVNRVTDVSKTPGGKGLNVTRVLSQLGDNVVATGLLG 70
           MILTVTLNPS+DISY LE   +DTVNRV DVSKT GGKGLNVTRVL + GD V ATG LG
Sbjct: 1   MILTVTLNPSVDISYPLETLKIDTVNRVKDVSKTAGGKGLNVTRVLYESGDKVTATGFLG 60

Query: 71  GDFGDFIRSGLDALEIRHQFLSIGGETRHCIAVLHEGQQTEILEKGPHITKDEADAFLNH 130
           G  G+FI S L+   +   F  I G TR+CIA+LHEG QTEILE+GP I+ +EA+ FL+H
Sbjct: 61  GKIGEFIESELEQSPVSPAFYKISGNTRNCIAILHEGNQTEILEQGPTISHEEAEGFLDH 120

Query: 131 LKLIFDAATIITVSGSLPKGLPSDYYARLISLANHFNKKVVLDCSGEALRSVLKSSAKPT 190
              +   + ++T+SGSLP GLP+DYY +LI LA+      VVLDCSG  L +VLKSSAKPT
Sbjct: 121 YSNLIKQSEVVTISGSLPSGLPNDYYEKLIQLASDEGVAVVLDCSGAPLETVLKSSAKPT 180

Query: 191 VIKPNLEELTQLIGKPISYSLDELKSTLQQDLFRGIDWVIVSLGARGAFAKHGNHYYQVT 250
            IKPN EEL+QL+GK ++  ++ELK  L++ LF GI+W++VSLG  GAFAKHG+ +Y+V
Sbjct: 181 AIKPNNEELSQLLGKEVTKDIEELKDVLKESLFSGIEWIVVSLGRNGAFAKHGDVFYKVD 240

Query: 251 IPKIEVINPVGSGDATVAGIASALEHQLDDTNLLKRANVLGMLNAQETLTGHINLTYYQE 310
           IP I V+NPVGSGD+TVAGIASAL  +  D +LLK A  LGMLNAQET+TGH+N+T Y+
Sbjct: 241 IPDIPVVNPVGSGDSTVAGIASALNSKKSDADLLKHAMTLGMLNAQETMTGHVNMTNYET 300
```

```
Query: 311 LISQIQVKEV 320
           L SQI VKEV
Sbjct: 301 LNSQIGVKEV 310
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6919> which encodes the amino acid sequence <SEQ ID 6920>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1178(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 = 184/310 (59%), Positives = 232/310 (74%), Gaps = 1/310 (0%)

Query: 11  MILTVTLNPSIDISYCLENFNMDTVNRVTDVSKTPGGKGLNVTRVLSQLGDNVVATGLLG 70
           +ILTVTLNP+ID+SY L+     DTVNRV DV+KTPGGKGLNV+RVL++ G+ V ATG +G
Sbjct: 1   VILTVTLNPAIDVSYPLDELKCDTVNRVVDVTKTPGGKGLNVSRVLNEFGETVKATGCVG 60

Query: 71  GDFGDFIRSGLDALEIRHQFLSIGGETRHCIAVLHEGQQTEILEKGPHITKDEADAFLNH 130
           G+ GDFI + L     I  +F  I G+TR CIA+LHEG QTEILEKGP ++ DE D F +H
Sbjct: 61  GESGDFIINHLPD-SILSRFYKISGDTRTCIAILHEGNQTEILEKGPMLSVDEIDGFTHH 119

Query: 131 LKLIFDAATIITVSGSLPKGLPSDYYARLISLANHFNKKVVLDCSGEALRSVLKSSAKPT 190
            K + +   ++T+SGSLP G+P DYY +LI +AN   KK VLDCSG AL +VLK  +KPT
Sbjct: 120 FKYLLNDVDVVTLSGSLPAGMPDDYYQKLIKIANLNGKKTVLDCSGNALEAVLKGDSKPT 179

Query: 191 VIKPNLEELTQLIGKPISYSLDELKSTLQQDLFRGIDWVIVSLGARGAFAKHGNHYYQVT 250
           VIKPNLEEL+QL+GK ++   D LK  LQ +LF GI+W+IVSLGA G FAKH + +Y V
Sbjct: 180 VIKPNLEELSQLLGKEMTKDFDALKEVLQDELFDGIEWIIVSLGADGVFAKHKDTFYNVD 239

Query: 251 IPKIEVINPVGSGDATVAGIASALEHQLDDTNLLKRANVLGMLNAQETLTGHINLTYYQE 310
           IPKI++++ VGSGD+TVAGIAS L +  DD  LL +ANVLGMLNAQE  TGH+N+  Y +
Sbjct: 240 IPKIKIVSAVGSGDSTVAGIASGLANDEDDRALLTKANVLGMLNAQEKTTGHVNMANYDK 299

Query: 311 LISQIQVKEV 320
           L   I+VKEV
Sbjct: 300 LYQSIKVKEV 309
```

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

Example 2239

A DNA sequence (GBSx2358) was identified in *S.agalactiae* <SEQ ID 6921> which encodes the amino acid sequence <SEQ ID 6922>. This protein is predicted to be tagatose 1,6-diphosphate aldolase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0369(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:AAA25180 GB:M60447 tagatose 1,6-diP aldolase [Lactococcus
          lactis]
```

```
 Identities = 253/325 (77%), Positives = 295/325 (89%)

Query: 1   MGLTEQKQKHMEQLSDKNGIISALAFDQRGALKRLMAKYQSEEPTVSQIEALKVLVAEEL 60
            M LTEQK+K +E+LSDKNG ISALAFDQRGALKRLMA+YQ  EPTV+Q+E LKVLVA+EL
Sbjct: 1   MVLTEQKRKSLEKLSDKNGFISALAFDQRGALKRLMAQYQDTEPTVAQMEELKVLVADEL 60

Query: 61  TPYASSMLLDPEYGLPATKVLDDNAGLLLAYEKTGYDTSSTKRLPDCLDIWSAKRIKEEG 120
            T YASSMLLDPEYGLPATK LD  AGLLLA+EKTGYDTSSTKRLPDCLD+WSAKRIKE+G
Sbjct: 61  TKYASSMLLDPEYGLPATKALDKEAGLLLAFEKTGYDTSSTKRLPDCLDVWSAKRIKEQG 120

Query: 121 ADAVKFLLYYDVDSSDEVNEEKEAYIERIGSECVAEDIPFFLEILSYDEKITDSSGIEYA 180
            ADAVKFLLYYDVDSSDE+N++K+AYIER+GSECVAEDIPFFLEIL+YDE+I+D+  +EYA
Sbjct: 121 ADAVKFLLYYDVDSSDELNQQKQAYIERVGSECVAEDIPFFLEILAYDEEISDAGSVEYA 180

Query: 181 KIKPRKVIEAMKVFSNPRFNIDVLKVEVPVNMDYVEGFAQGETAYNKATAAAYFREQDQA 240
            K+KPRKVIEAMKVFS+PRFNIDVLKVEVPVN+ YVEGFA GE  Y+KA AA +F+ Q++A
Sbjct: 181 KVKPRKVIEAMKVFSDPRFNIDVLKVEVPVNVKYVEGFADGEVVYSKAEAADFFKAQEEA 240

Query: 241 TLLPYIFLSAGVPAQLFQETLVFAKEAGAKFNGVLCGRATWAGSVKEYVEKGEAGARQWL 300
            T LPYI+LSAGV A+LFQETL FA ++GAKFNGVLCGRATWAGSV+ Y+++GE  AR+WL
Sbjct: 241 TNLPYIYLSAGVSAKLFQETLQFAHDSGAKFNGVLCGRATWAGSVEPYIKEGEKAAREWL 300

Query: 301 RTIGFQNIDELNKILQKTATSWKER 325
            RT GF+NIDELNK+L KTA+ W ++
Sbjct: 301 RTTGFENIDELNKVLVKTASPWTDK 325
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6923> which encodes the amino acid sequence <SEQ ID 6924>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0600(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 = 230/323 (71%), Positives = 276/323 (85%), Gaps = 1/323 (0%)

Query: 3   LTEQKQKHMEQLSDKNGIISALAFDQRGALKRLMAKYQSEEPTVSQIEALKVLVAEELTP 62
            LTE K+K ME+LS  +G+ISALAFDQRGALKR+MA++Q++EPTV QIE LK LV+EELTP
Sbjct: 5   LTENKRKSMEKLS-VDGVISALAFDQRGALKRMMAQHQTKEPTVEQIEELKSLVSEELTP 63

Query: 63  YASSMLLDPEYGLPATKVLDDNAGLLLAYEKTGYDTSSTKRLPDCLDIWSAKRIKEEGAD 122
            +ASS+LLDPEYGLPA++V  + AGLLLAYEKTGYD ++T RLPDCLD+WSAKRIKE GA+
Sbjct: 64  FASSILLDPEYGLPASRVRSEEAGLLLAYEKTGYDATTTSRLPDCLDVWSAKRIKEAGAE 123

Query: 123 AVKFLLYYDVDSSDEVNEEKEAYIERIGSECVAEDIPFFLEILSYDEKITDSSGIEYAKI 182
            AVKFLLYYD+D   +VNE+K+AYIERIGSEC AEDIPF+LEIL+YDEKI D++  E+AK+
Sbjct: 124 AVKFLLYYDIDGDQDVNEQKKAYIERIGSECRAEDIPFYLEILTYDEKIADNASPEFAKV 183

Query: 183 KPRKVIEAMKVFSNPRFNIDVLKVEVPVNMDYVEGFAQGETAYNKATAAAYFREQDQATL 242
            K  KV EAMKVFS  RF +DVLKVEVPVNM +VEGFA GE  + K  AA  FR+Q+ +T
Sbjct: 184 KAHKVNEAMKVFSKERFGVDVLKVEVPVNMKFVEGFADGEVLFTKEEAAQAFRDQEASTD 243

Query: 243 LPYIFLSAGVPAQLFQETLVFAKEAGAKFNGVLCGRATWAGSVKEYVEKGEAGARQWLRT 302
            LPYI+LSAGV A+LFQ+TLVFA E+GAKFNGVLCGRATWAGSVK Y+E+G   AR+WLRT
Sbjct: 244 LPYIYLSAGVSAKLFQDTLVFAAESGAKFNGVLCGRATWAGSVKVYIEEGPQAAREWLRT 303

Query: 303 IGFQNIDELNKILQKTATSWKER 325
            GF+NIDELNK+L KTA+ W E+
Sbjct: 304 EGFKNIDELNKVLDKTASPWTEK 326
```

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

Example 2240

A DNA sequence (GBSx2359) was identified in *S.agalactiae* <SEQ ID 6925> which encodes the amino acid sequence <SEQ ID 6926>. This protein is predicted to be lacx protein, chromosomal. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0643(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 10185> which encodes amino acid sequence <SEQ ID 10186> was also identified.

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

```
>GP:AAA25184 GB:M60447 ORF [Lactococcus lactis]
 Identities = 173/298 (58%), Positives = 219/298 (73%)

Query: 24  MAITIQNHELQVTLKALGATMTSITDSQGVEYLWQGDATYWGGQAPILFPICGSVRNDCV 83
           M I ++N  L V  K LG  +TSI D  G+EYLWQ D  YW GQAPILFPICGS+RND
Sbjct: 1   MTIELKNEYLTVQFKTLGGQLTSIKDKDGLEYLWQADPEYWNGQAPILFPICGSLRNDWA 60

Query: 84  IYRPAQAPHFTGIIPRHGFVRHKTFDYDYISDSSVRFTIKSSKEMLINYPYRFSLEITYT 143
           IYRP + P FTG+I RHGFVR + F  + ++++SV F+IK + EML NY Y+F L + YT
Sbjct: 61  IYRPQERPFFTGLIRRHGFVRKEEFTLEEVNENSVTFSIKPNAEMLDNYLYQFELRVVYT 120

Query: 144 LRNKSIAITYIVKNLESEKNMPYAIGAHPGFNCPLFEKEVFSDYYLEFEQFETCTIPESF 203
           L  KSI   + V NLE+EK MPY IGAHP FNCPL E E + DY LEF + E+C+IP+SF
Sbjct: 121 LNGKSIRTEFQVTNLETEKTMPYFIGAHPAFNCPLVEGEKYEDYSLEFSEVESCSIPKSF 180

Query: 204 PDTGLLDLQARHPFLENQKQLSLNHALFEKDAITLDQLRSKTVYLKSRNHAKGIQLDFDD 263
           P+TGLLDLQ R PFLENQK L L+++LF  DAITLD+L+S++V L+SR   KG+++DFDD
Sbjct: 181 PETGLLDLQDRTPFLENQKSLDLDYSLFSHDAITLDRLKSRSVTLRSRKSGKGLRVDFDD 240

Query: 264 FENLILWTSNNGGPFLALEPWSSLSTSIEESDILEDKQNIVRLNPKQSKQHSIRITIL 321
           F NLILW++ N  PF+ALEPWS LST+EE +ILEDK  + ++ P  + + S  ITIL
Sbjct: 241 FPNLILWSTTNKSPFIALEPWSGLSTSLEEGNILEDKPQVTKVLPLDTSKKSYDITIL 298
```

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 2241

A DNA sequence (GBSx2361) was identified in *S.agalactiae* <SEQ ID 6927> which encodes the amino acid sequence <SEQ ID 6928>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3272(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 10183> which encodes amino acid sequence <SEQ ID 10184> was also identified.

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

```
>GP:CAA51350 GB:X72832 leucine rich protein [Streptococcus
          equisimilis]
 Identities = 101/278 (36%), Positives = 160/278 (57%), Gaps = 1/278 (0%)

Query: 10  MDFKELFPEVITKQEVKQSEDYIIVEQDGHVLHFPKSSLTKRELYLLQMTPSLEDASSVD 69
           M+ K+ FPE+          ++++ V++     +HFPKS L+++E  LL++      +
Sbjct: 1   MELKDYFPEMQVGPHPLGDKEWVSVKEGDQYVHFPKSCLSEKERLLLEVGLGQYEVLQ-P 59

Query: 70  SQNPWYRYLVEGRGRLPQSHSAVQFIFIEHQFTLSEELKDFLSPLVINVETIMTINQTQS 129
             +PW RYL++ +G  PQ      QFI++ HQ  L  +L + L  ++   +E I+ I+ TQ+
Sbjct: 60  LGSPWQRYLLDHQGNPPQLFETSQFIYLNHQQVLPADLVELLQQMIAGLEVILPISTTQT 119

Query: 130 VMILNQDNFFNATELLTDILPTIENDFNTRLRCYFGNSWTHLQAVDWKELYEEEYKLFTL 189
             +  Q          L  +LPT+E+DF    L  + GN+W  + A    +E +EEE +L T
Sbjct: 120 AFLCRQATSIKVLRSLEGLLPTLESDFGLALTMFVGNAWYQVAAGTLRECFEEECQLLTA 179

Query: 190 FLSHKAEQHYCRFPKMALWALANQSPMPSIKAKCLQHILDTSDTSAIIKALWQEQGNLAK 249
           +L  K+       F ++ LW++ +     P++  +  Q +    SD + ++ ALW E GNL +
Sbjct: 180 YLKQKSGGKLLTFAEVMLWSILSHQSFPALTRQFHQFLNPQSDMADVVHALWSEHGNLVQ 239

Query: 250 TAKALFIHRNSLQYKLDKFTQSSGLNLKILDDLAYAYL 287
           TA+ L+IHRNSLQYKLDKF Q SGL+LK LDDLA+AYL
Sbjct: 240 TAQRLYIHRNSLQYKLDKFAQQSGLHLKQLDDLAFAYL 277
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6929> which encodes the amino acid sequence <SEQ ID 6930>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4332(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 = 106/287 (36%), Positives = 169/287 (57%), Gaps = 4/287 (1%)

Query: 3   KTVVED-AMDFKELFPEVITKQEVKQSEDYIIVEQDGHVLHFPKSSLTKRELYLLQM-TP 60
           KTV++  AM+ K+ FPE+          +D++ +++     +HFPKS L+++E  LL++
Sbjct: 7   KTVMKGMAMELKDYFPEMQVGPHPLGDKDWMSIKEGDQYVHFPKSCLSEKERLLLEVGLG 66

Query: 61  SLEDASSVDSQNPWYRYLVEGRGRLPQSHSAVQFIFIEHQFTLSEELKDFLSPLVINVET 120
             E    + S  PW RYL++ +G  PQ +   QFI++ HQ  L ++L + L  ++   +E
Sbjct: 67  QCEVLQPLGS--PWQRYLLDHQGNPPQLYETSQFIYLNHQQALPDDLVELLQQMIAGLEV 124

Query: 121 IMTINQTQSVMILNQDNFFNATELLTDILPTIENDFNTRLRCYFGNSWTHLQAVDWKELY 180
           I+ I+ TQ+  + Q          L D+LPT+E+DF    L  + GN+W  + A    +E +
Sbjct: 125 ILPISATQTAFLCRQAISIKVLRWLEDLLPTLESDFGLALTMFVGNAWYQVAAGTLRECF 184

Query: 181 EEEYKLFTLFLSHKAEQHYCRFPKMALWALANQSPMPSIKAKCLQHILDTSDTSAIIKAL 240
           EEE +L T +L  ++ +     F  + LW+L +       ++  +  Q +    SD + ++ AL
Sbjct: 185 EEECQLLTAYLRQQSGRKLLTFSGLMLWSLLSHHTFLALTRQFHQFLSPQSDMADVVHAL 244

Query: 241 WQEQGNLAKTAKALFIHRNSLQYKLDKFTQSSGLNLKILDDLAYAYL 287
           W E GNL +TA+ L+IHRNSLQYKLDKF Q SGL+LK LDDLA+A+L
Sbjct: 245 WSEHGNLVQTAQRLYIHRNSLQYKLDKFAQQSGLHLKQLDDLAFAHL 291
```

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

Example 2242

A DNA sequence (GBSx2362) was identified in *S.agalactiae* <SEQ ID 6931> which encodes the amino acid sequence <SEQ ID 6932>. This protein is predicted to be multiple sugar-binding transport ATP-binding protein msmk (malK). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4392(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:AAA26938 GB:M77351 ATP-binding protein [Streptococcus mutans]
 Identities = 320/377 (84%), Positives = 359/377 (94%)

Query: 1   MVELNLNHIYKKYPSASHYSVEDFDLDIKDKEFIVFVGPSGCGKSTTLRMIAGLEDISEG 60
            MVELNLNHIYKKYP++SHYSVEDFDLDIK+KEFIVFVGPSGCGKSTTLRM+AGLEDI++G
Sbjct: 1   MVELNLNHIYKKYPNSSHYSVEDFDLDIKNKEFIVFVGPSGCGKSTTLRMVAGLEDITKG 60

Query: 61  ELKIDGEVVNDKSPKDRDIAMVFQNYALYPHMTVYDNMAFGLKLRKFSKQEIDKRVREAA 120
            ELKIDGEVVNDK+PKDRDIAMVFQNYALYPHM+VYDNMAFGLKLR +SK+ IDKRV+EAA
Sbjct: 61  ELKIDGEVVNDKAPKDRDIAMVFQNYALYPHMSVYDNMAFGLKLRHYSKEAIDKRVKEAA 120

Query: 121 ANIGLTEFLERKPADLSGGQRQRVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAEIAK 180
              +GLTEFLERKPADLSGGQRQRVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAEIAK
Sbjct: 121 QILGLTEFLERKPADLSGGQRQRVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAEIAK 180

Query: 181 IHQRIGSTTIYVTHDQTEAMTLADRIVIMSATKNPDGDGTIGKIEQVGSPQELYNLPANK 240
            IH+RIG+TTIYVTHDQTEAMTLADRIVIMS+TKN DG GTIG++EQVG+PQELYN PANK
Sbjct: 181 IHRRIGATTIYVTHDQTEAMTLADRIVIMSSTKNEDGSGTIGRVEQVGTPQELYNRPANK 240

Query: 241 FVAGFIGSPSMNFFKVKVENGMIISEDGLRIAIPEGQEKLLESRGYKGKELIFGIRPEDI 300
            FVAGFIGSP+MNFF V +++G ++S+DGL IA+ EGQ K+LES+G+K K LIFGIRPEDI
Sbjct: 241 FVAGFIGSPAMNFFDVTIKDGHLVSKDGLTIAVTEGQLKMLESKGFKNKNLIFGIRPEDI 300

Query: 301 SSNLLVQDTYPNANVEAEVLVSELLGSETMLYVKLGQTEFASRVEARDFHNPGEKVNLTF 360
            SS+LLVQ+TYP+A V+AEV+VSELLGSETMLY+KLGQTEFA+RV+ARDFH PGEKV+LTF
Sbjct: 301 SSSLLVQETYPDATVDAEVVVSELLGSETMLYLKLGQTEFAARVDARDFHEPGEKVSLTF 360

Query: 361 NVAKGHFFDADTEQAIR 377
            NVAKGHFFDA+TE AIR
Sbjct: 361 NVAKGHFFDAETEAAIR 377
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6933> which encodes the amino acid sequence <SEQ ID 6934>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4642(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 = 332/377 (88%), Positives = 359/377 (95%)
```

```
Query: 1   MVELNLNHIYKKYPSASHYSVEDFDLDIKDKEFIVFVGPSGCGKSTTLRMIAGLEDISEG 60
           MVELNLNHIYKKYP+ +HY+VEDFDLDIKDKEFIVFVGPSGCGKSTTLRMIAGLEDISEG
Sbjct: 1   MVELNLNHIYKKYPNTTHYAVEDFDLDIKDKEFIVFVGPSGCGKSTTLRMIAGLEDISEG 60

Query: 61  ELKIDGEVVNDKSPKDRDIAMVFQNYALYPHMTVYDNMAFGLKLRKFSKQEIDKRVREAA 120
           ELKI GEVVNDKSPKDRDIAMVFQNYALYPHMTVYDNMAFGLKLRK+ K +ID+RV+EAA
Sbjct: 61  ELKIGGEVVNDKSPKDRDIAMVFQNYALYPHMTVYDNMAFGLKLRKYKKDDIDRRVKEAA 120

Query: 121 ANIGLTEFLERKPADLSGGQRQRVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAEIAK 180
             +GLTEFLERKPADLSGGQRQRVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAEIAK
Sbjct: 121 QILGLTEFLERKPADLSGGQRQRVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAEIAK 180

Query: 181 IHQRIGSTTIYVTHDQTEAMTLADRIVIMSATKNPDGDGTIGKIEQVGSPQELYNLPANK 240
           IH+RIGSTTIYVTHDQTEAMTLADRIVIMSATKNP G+GTIGKIEQVGSPQELYNLPANK
Sbjct: 181 IHRRIGSTTIYVTHDQTEAMTLADRIVIMSATKNPQGNGTIGKIEQVGSPQELYNLPANK 240

Query: 241 FVAGFIGSPSMNFFKVKVENGMIISEDGLRIAIPEGQEKLLESRGYKGKELIFGIRPEDI 300
           FVAGFIGSP+MNFF+V+V++G I+SEDGL IAIPEGQ K+LE+ GYKG+++ FGIRPEDI
Sbjct: 241 FVAGFIGSPAMNFFEVEVKDGRIVSEDGLDIAIPEGQAKMLEAAGYKGEKVTFGIRPEDI 300

Query: 301 SSNLLVQDTYPNANVEAEVLVSELLGSETMLYVKLGQTEFASRVEARDFHNPGEKVNLTF 360
           SS  +V DTYP+A V AEVLVSELLGSETMLYVKLGQTEFASRV+ARDFH+PGE+V+LTF
Sbjct: 301 SSRQIVHDTYPSATVTAEVLVSELLGSETMLYVKLGQTEFASRVDARDFHSPGEQVSLTF 360

Query: 361 NVAKGHFFDADTEQAIR 377
           NVAKGHFFD DTEQAIR
Sbjct: 361 NVAKGHFFDRDTEQAIR 377
```

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

Example 2243

A DNA sequence (GBSx2363) was identified in *S.agalactiae* <SEQ ID 6935> which encodes the amino acid sequence <SEQ ID 6936>. This protein is predicted to be glucan 1,6-alpha-glucosidase (dexB) (treC). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2525(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:CAA51348 GB:X72832 glucan 1,6-alpha-glucosidase [Streptococcus
          equisimilis]
 Identities = 413/535 (77%), Positives = 476/535 (88%), Gaps = 1/535 (0%)

Query: 1   MKKHWWHKATIYQIYPRSFMDSDGDGVGDIKGITSKLDYLEKLGITAIWLSPVYQSPMDD 60
           M+K WWHKATIYQIYPRSF D+ G+G+GD+KGITS+LDYL+KLGITAIWLSPVYQSPMDD
Sbjct: 1   MQKQWWHKATIYQIYPRSFKDTSGNGIGDLKGITSQLDYLQKLGITAIWLSPVYQSPMDD 60

Query: 61  NGYDISDYQAIADIFGDMNDMDQLLQEANQRGIKIIMDLVVNHTSDEHAWFVEARENPNS 120
           NGYDISDY+AIA++FG+M+DMD LL  AN+RGIKIIMDLVVNHTSDEHAWFVEARENPNS
Sbjct: 61  NGYDISDYEAIAEVFGNMDDMDDLLAAANERGIKIIMDLVVNHTSDEHAWFVEARENPNS 120

Query: 121 PERDFYIWRDEPNDLTSIFSGSAWEYDKVSGQYYLHLFSKRQPDLNWENEALRHKIYDMM 180
           PERD+YIWRDEPN+L SIFSGSAWE D+ SGQYYLHLFSK+QPDLNWEN  +R KIYDMM
Sbjct: 121 PERDYYIWRDEPNNLMSIFSGSAWELDEASGQYYLHLFSKKQPDLNWENAHVRQKIYDMM 180

Query: 181 NFWIDKGIGGFRMDVIDLIGKIPDKGITGNGPKLHDYLKEMNRASFGKHDLLTVGETWGA 240
```

```
            NFWI KGIGGFRMDVIDLIGKIPD  ITGNGP+LHDYLKEMN+A+FG HD++TVGETWGA
Sbjct: 181 NFWIAKGIGGFRMDVIDLIGKIPDSEITGNGPRLHDYLKEMNQATFGNHDVMTVGETWGA 240

Query: 241 TPDIAKQYSNPDNEELSMVFQFEHVGLQHKPDAPKWDYSDGLDVPALKDIFTKWQTQLEL 300
            TP+IA+QYS P+N+ELSMVFQFEHVGLQHKP+APKWDY++ LDVPALK IF+KWQT+L+L
Sbjct: 241 TPEIARQYSRPENKELSMVFQFEHVGLQHKPNAPKWDYAEELDVPALKTIFSKWQTELKL 300

Query: 301 GQGWNSLFWNNHDLPRVLSIWGNDSDNRKQSAKALAILLHLMRGTPYIYQGEEIGMTNYP 360
            G+GWNSLFWNNHDLPRVLSIWGNDS  R++SAKALAILLHLMRGTPYIYQGEEIGMTNYP
Sbjct: 301 GEGWNSLFWNNHDLPRVLSIWGNDSIYREKSAKALAILLHLMRGTPYIYQGEEIGMTNYP 360

Query: 361 FECLADVDDIESLNYAKEAMDNGVSEATILDSIRKVGRDNARTPMQWSQEHQAGFTKG-T 419
            F+ L +VDDIESLNYAKEAM+NGV  A ++ SIRKVGRDNARTPMQWS++  AGF++
Sbjct: 361 FKDLTEVDDIESLNYAKEAMENGVPAARVMSSIRKVGRDNARTPMQWSKDTHAGFSEAQE 420

Query: 420 PWLAVNPNYQEINVEAALNDTESIFYTYQKLVALRKEHDWLVDADFKLLETADKVFAYVR 479
             WL VNPNYQEINV  AL + +SIFYTYQ+L+ALRK+ DWLV+AD+ LL TADKVFAY R
Sbjct: 421 TWLPVNPNYQEINVADALANQDSIFYTYQQLIALRKDQDWLVEADYHLLPTADKVFAYQR 480

Query: 480 QTDKERYLIVANLSDQNQSFEFPEAVKETIISNTEVQEVLSSNTLKPWDAFCIEL 534
            Q  +E Y+IV N+SDQ Q F    A  E +I+NT+V +VL +  L+PWDAFC++L
Sbjct: 481 QFGEETYVIVVNVSDQEQVFAKDLAGAEVVITNTDVDKVLETKHLQPWDAFCVKL 535
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6937> which encodes the amino acid sequence <SEQ ID 6938>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2793(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 = 418/535 (78%), Positives = 474/535 (88%), Gaps = 1/535 (0%)

Query: 1   MKKHWWHKATIYQIYPRSFMDSDGDGVGDIKGITSKLDYLEKLGITAIWLSPVYQSPMDD 60
            M  HWWHKATIYQIYPRSF D+ G+G+GD+KGITS+LDYL+KLGITAIWLSPVYQSPMDD
Sbjct: 1   MNNHWWHKATIYQIYPRSFKDTSGNGIGDLKGITSQLDYLQKLGITAIWLSPVYQSPMDD 60

Query: 61  NGYDISDYQAIADIFGDMNDMDQLLQEANQRGIKIIMDLVVNHTSDEHAWFVEARENPNS 120
            NGYDISDY+AIAD+FGDM DMD+LL  AN+RGIKIIMDLVVNHTSDEHAWFVEARENPNS
Sbjct: 61  NGYDISDYEAIADVFGDMADMDELLAAANERGIKIIMDLVVNHTSDEHAWFVEARENPNS 120

Query: 121 PERDFYIWRDEPNDLTSIFSGSAWEYDKVSGQYYLHLFSKRQPDLNWENEALRHKIYDMM 180
            PERD+YIWRDEPN+L SIFSGSAWE D+ SGQYYLHLFSK+QPDLNWEN  LR KIYDMM
Sbjct: 121 PERDYYIWRDEPNNLMSIFSGSAWELDEASGQYYLHLFSKKQPDLNWENAQLRQKIYDMM 180

Query: 181 NFWIDKGIGGFRMDVIDLIGKIPDKGITGNGPKLHDYLKEMNRASFGKHDLLTVGETWGA 240
            NFWI KGIGGFRMDVIDLIGK+PD  ITGNGP+LHDYLKEMN+A+FG HD++TVGETWGA
Sbjct: 181 NFWIAKGIGGFRMDVIDLIGKVPDLEITGNGPRLHDYLKEMNQATFGNHDVMTVGETWGA 240

Query: 241 TPDIAKQYSNPDNEELSMVFQFEHVGLQHKPDAPKWDYSDGLDVPALKDIFTKWQTQLEL 300
            TP+IA+QYS P+N+ELSMVFQFEHVGLQHKPDAPKWDY+  LDVPALK IF+KWQT+L+L
Sbjct: 241 TPEIARQYSRPENKELSMVFQFEHVGLQHKPDAPKWDYAKELDVPALKAIFSKWQTELKL 300

Query: 301 GQGWNSLFWNNHDLPRVLSIWGNDSDNRKQSAKALAILLHLMRGTPYIYQGEEIGMTNYP 360
            G+GWNSLFWNNHDLPRVLSIWGNDS  R++SAKALAILLHLMRGTPYIYQGEEIGMTNYP
Sbjct: 301 GEGWNSLFWNNHDLPRVLSIWGNDSTYREKSAKALAILLHLMRGTPYIYQGEEIGMTNYP 360

Query: 361 FECLADVDDIESLNYAKEAMDNGVSEATILDSIRKVGRDNARTPMQWSQEHQAGFTKG-T 419
            F+ L +V+DIESLNYAKEAM NGVS A ++DSIRKVGRDNARTPMQWS++  AGF++
Sbjct: 361 FKDLTEVNDIESLNYAKEAMGNGVSAARVMDSIRKVGRDNARTPMQWSKDTHAGFSEAKE 420

Query: 420 PWLAVNPNYQEINVEAALNDTESIFYTYQKLVALRKEHDWLVDADFKLLETADKVFAYVR 479
```

```
            WL VNPNYQ+INV  AL D +SIFYTYQKL+ALRKE DWLV+AD+ LL TADKVFAY R
Sbjct: 421 TWLPVNPNYQDINVADALADPDSIFYTYQKLIALRKEQDWLVEADYHLLPTADKVFAYQR 480

Query: 480 QTDKERYLIVANLSDQNQSFEFPEAVKETIISNTEVQEVLSSNTLKPWDAFCIEL 534
           Q  +E Y+IV N+SD+ Q F    A  + II+NT+V  VL +  L+PWDAFC++L
Sbjct: 481 QLGEETYVIVVNVSDEEQVFATDLAGAQVIIANTDVDTVLETKHLQPWDAFCLKL 535
```

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

Example 2244

A DNA sequence (GBSx2364) was identified in *S.agalactiae* <SEQ ID 6939> which encodes the amino acid sequence <SEQ ID 6940>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have an uncleavable N-term signal seq

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

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

```
>GP:AAB49738 GB:U21942 UDP-galactose 4-epimerase [Streptococcus mutans]
 Identities = 267/331 (80%), Positives = 306/331 (91%)

Query: 1   MAVLILGGAGYIGSHMVDQLITQGKEKVIVVDNLVTGHRQAVHSDAIFYEGDLSDKTFMR 60
           MA+L+LGGAGYIGSHMVD+LI +G+E+V+VVD+LVTGHR AVH  A FY+GDL+D+ FM
Sbjct: 1   MAILVLGGAGYIGSHMVDRLIEKGEEEVVVVDSLVTGHRAAVHPAAKFYQGDLADREFMS 60

Query: 61  QVFRENPDVDAVIHFAAFSLVAESMENPLKYFDNNTAGMIKLLEVMNECDIKNIVFSSTA 120
            VFRENPDVDAVIHFAA+SLVAESM+ PLKYFDNNTAGMIKLLEVM+E  +K IVFSSTA
Sbjct: 61  MVFRENPDVDAVIHFAAYSLVAESMKKPLKYFDNNTAGMIKLLEVMSEFGVKYIVFSSTA 120

Query: 121 ATYGIPEQVPILETAPQNPINPYGESKLMMETIMKWADQAYGIKFVALRYFNVAGDKPDG 180
           ATYGIP ++PI ET PQ PINPYGESKLMMETIMKW+D+AYGIKFV +RYFNVAG KPDG
Sbjct: 121 ATYGIPNEIPIKETTPQRPINPYGESKLMMETIMKWSDRAYGIKFVPVRYFNVAGAKPDG 180

Query: 181 SIGEDHKPETHLLPIILQVAQGVRDKIMIFGDDYNTPDGTNVRDYVHPFDLADAHILAVD 240
           SIGEDH PETHLLPIILQVAQGVR+KIMIFGDDYNTPDGTNVRDYVHPFDLAD H+LA++
Sbjct: 181 SIGEDHSPETHLLPIILQVAQGVREKIMIFGDDYNTPDGTNVRDYVHPFDLADRHLLALN 240

Query: 241 YLRQGNESNVFNLGSSTGFSNLQMLEAARRITGKEIPAQKAARRPGDPDTLIASSEKARQ 300
           YLRQGN S  FNLGSSTGFSNLQ+LEAAR++TG++IPA+KAARR GDPDTLIASSEKAR+
Sbjct: 241 YLRQGNPSTAFNLGSSTGFSNLQILEAARKVTGQKIPAEKAARRSGDPDTLIASSEKARE 300

Query: 301 ILGWEPKFDNIDKIISSAWAWHSSHPNGYED 331
           ++GW+P+FD+I+KII+SAWAWHSSHP GY+D
Sbjct: 301 VVGWKPQFDDIEKIIASAWAWHSSHPKGYDD 331
```

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

Example 2245

A DNA sequence (GBSx2366) was identified in *S.agalactiae* <SEQ ID 6941> which encodes the amino acid sequence <SEQ ID 6942>. This protein is predicted to be two-component response regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3945(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:BAB06470 GB:AP001516 two-component response regulator [Bacillus halodurans]
 Identities = 71/223 (31%), Positives = 139/223 (61%), Gaps = 7/223 (3%)

Query: 3   VLIIEDDPMVEFIHRNYLEKLNYFQNIYSTASQTQAIAYLNDIKIQLVLLDIHIKEGNGL 62
           VL+IEDDPMV+ ++R ++EKL+ F  + +TA+  + +    +++  L+LLDI + + +GL
Sbjct: 9   VLLIEDDPMVQEVNRMFVEKLSGFTIVGTTATGEEGMVKTRELQPDLILLDIFMPKQDGL 68

Query: 63  ELLKLLRNQHQNTEVIVISAANEAHTVKEAFHLGIVDYLIKPFTFERFESSIEKFLNHYH 122
             +K +R Q+ + ++I ++AAN+  T+K     G++DYL+KPFTFER ++++ ++   +
Sbjct: 69  SFIKQIREQYIDVDIIAVTAANDTKTIKTLLRYGVMDYLVKPFTFERLKAALTQYEEMFR 128

Query: 123 TFEAD-KIYQDNIDHFQKIDSGWLEGEVKLDE--KGLSEITYQHILDAIQELEQPFTIQE 179
             + + ++ QD++D   K      + +  +D+  KGL   T Q +++ ++EL++P + +E
Sbjct: 129 KMQKEAAELSQDSLDEMIK----QKQAQANMDDLPKGLHAHTLQQVIERLEELDEPKSAEE 184

Query: 180 LAKCSQFSHVSVRKYIAYMEEKGLLTSQQIYTKVGRPYKVYKL 222
           + +    + V+VR+Y+ Y+E  G +     Y  +GRP + YKL
Sbjct: 185 IGRDVGLARVTVRRYLNYLESVGQVEMDLTYGSIGRPIQTYKL 227
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6943> which encodes the amino acid sequence <SEQ ID 6944>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4053(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 = 123/220 (55%), Positives = 156/220 (70%)

Query: 1   MDVLIIEDDPMVEFIHRNYLEKLNYFQNIYSTASQTQAIAYLNDIKIQLVLLDIHIKEGN 60
           M+VLIIEDDPMV+FIHRNYLEKLN F  I S+ S     + L D  I L+LLDIHI +GN
Sbjct: 1   MNVLIIEDDPMVDFIHRNYLEKLNLFDRIISSDSMKAVQSILTDYAIDLILLDIHITDGN 60

Query: 61  GLELLKLLRNQHQNTEVIVISAANEAHTVKEAFHLGIVDYLIKPFTFERFESSIEKFLNH 120
           G++ L+  R QH   EVI+ISAAN+ + +++ FHLGI+DYLIKPFTFERF+ SI++F+ H
Sbjct: 61  GIQFLEKWRTQHIPCEVIIISAANDGNIIRDGFHLGIIDYLIKPFTFERFQESIQQFVTH 120

Query: 121 YHTFEADKIYQDNIDHFQKIDSGWLEGEVKLDEKGLSEITYQHILDAIQELEQPFTIQEL 180
                ++ Q  ID  + + S       +L EKGLSE T+Q I++ I+  +QPFTIQEL
Sbjct: 121 REHLANQQLEQAQIDQLKCLTSKKDTKNKQLLEKGLSESTFQWIMENIKVFDQPFTIQEL 180

Query: 181 AKCSQFSHVSVRKYIAYMEEKGLLTSQQIYTKVGRPYKVY 220
           A     SHVSVRKYIAY+EE   L SQQI+TKVGRPY+VY
Sbjct: 181 ASACHLSHVSVRKYIAYLEENKQLNSQQIFTKVGRPYRVY 220
```

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

Example 2246

A DNA sequence (GBSx2367) was identified in *S.agalactiae* <SEQ ID 6945> which encodes the amino acid sequence <SEQ ID 6946>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.76   Transmembrane   12 -  28 (   6 -  34)
   INTEGRAL    Likelihood = -7.43   Transmembrane  178 - 194 ( 173 - 197)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9003> which encodes amino acid sequence <SEQ ID 9004> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length of UR:   27
     Peak Value of UR:   2.99
     Net Charge of CR: 3
McG: Discrim Score:     12.92
GvH: Signal Score (-7.5): -2.57
     Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 2 value:  -8.76 threshold:  0.0
   INTEGRAL    Likelihood = -8.76   Transmembrane   10 -  26 (   4 -  32)
   INTEGRAL    Likelihood = -7.43   Transmembrane  176 - 192 ( 171 - 195)
   PERIPHERAL  Likelihood =  3.18     149
 modified ALOM score:   2.25
icm1 HYPID: 7  CFP: 0.450

*** Reasoning Step: 3

----- 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:CAB15141 GB:Z99120 similar to two-component sensor histidine
          kinase [YufM] [Bacillus subtilis]
 Identities = 132/461 (28%), Positives = 245/461 (52%), Gaps = 7/461 (1%)

Query: 3   MKKKLSLWAFLSLILVTMTICIFSIFYYVTIHQSYRMVRVQEEKILKNTGYALSRNPQVI 62
           MKK L L   L++ +  + +     I ++    Q+ + +R QE+     T   ++  P
Sbjct: 1   MKKTLKLQTRLTIFVCIVVLIALLITFFTVGAQTTKRIRDQEKATALQTAEMVAEAPMTA 60

Query: 63  QTLKDNHYDQSLQKQMLFLSKKSNLDYIVLINLKGIRFTHPDSTKIGKPFQGGDEQAVFK 122
             L+     + LQ     + K +  +++V++++ GIR THPD +KIGK F+GGDE  V K
Sbjct: 61  AALESGKKQKELQSYTKRVQKITGTEFVVVMDMNGIRKTHPDPSKIGKKFRGGDESEVLK 120

Query: 123 GKAIMSTAEGSLGKSLRYLIPVY-DHQKQVGAIAVGLKLTTLGDLSQSSIKEFSKPLLIS 181
           G   +STA G+LGKS R  +PVY ++ KQVGA+AVG+ +  + ++   S++     + +S
Sbjct: 121 GHVHISTASGTLGKSQRAFVPVYAENGKQVGAVAVGITVNEIDEVISHSLRPLYFIICVS 180

Query: 182 ILISLVVTSIISYGLKKQLHNLHPSDIFQHLEERNATLDQIQAAVFVIDQRHIIKRNNPA 241
           I + ++   I++  +K  ++L P +I   LEER+A L+  +  +  +D+   IK  N
Sbjct: 181 IFVGVIGAVIVARTVKNIMYGLEPYEIATLLEERSAMLESTKEGILAVDEHGKIKLANAE 240

Query: 242 ASLLFKKEGQRDLFSGKLLESLIP--QLKQDHFSKK--TEQVLHFQGQDYLLSISPITVK 297
           A  LF K G       + ++ ++P  +LK+    +KK   ++ +   G + + +  PI +K
Sbjct: 241 AKRLFVKMGINTNPIDQDVDDILPKSRLKKVIETKKPLQDRDVRINGLELVFNEVPIQLK 300

Query: 298 TQNRGYVVFLRNVTETLFTLDQLAHTTAYASALQAQTHQFMNQLHVIYGLADIEYYDELK 357
            Q  G +   R+ TE      +QL+     YA+AL+AQ+H+FMN+LHVI GL   ++ YD+L
Sbjct: 301 GQTVGAIATFRDKTEVKHLAEQLSGVKMYANALRAQSHEFMNKLHVILGLVQLKEYDDLG 360
```

```
Query: 358 IYLKELLEPQNEFLARLSMLVREPRLASFIIGEREKFAEKHINLSTEILVEIPTKSTVED 417
           Y+K++   Q    + +   V+   LA F++G++    E+  NL  E    IP  +
Sbjct: 361 DYIKDIAIQQKSETSEIINDVKSSVLAGFLLGKQSFIREQGANLDIECNGVIPNAADPSV 420

Query: 418 VNNYL-LLHRYINTKILTLLN-STTLVSLRLNYQNNLIETD 456
           ++  + ++   IN  +  + +     +++ + + N++++ +
Sbjct: 421 IHELITIIGNLINNGLDAVADMPKKQITMSMRFHNSILDIE 461
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6947> which encodes the amino acid sequence <SEQ ID 6948>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL     Likelihood =-10.03   Transmembrane  174 - 190 ( 170 - 195)

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

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

```
 Identities = 236/488 (48%), Positives = 337/488 (68%), Gaps = 3/488 (0%)

Query: 3   MKKKLSLWAFLSLILVTMTICIFSIFYYVTIHQSYRMVRVQEEKILKNTGYALSRNPQVI 62
           MKK L LWA LSLILV+M +   S+FY + +H +++ ++ QE  +L +TG  L+ +  +
Sbjct: 1   MKKPLRLWASLSLILVSMIVVTTSLFYGIMLHDTHQSIKNQETHLLTSTGKMLASHQAIK 60

Query: 63  QTLKDNHYDQSLQKQMLFLSKKSNLDYIVLINLKGIRFTHPDSTKIGKPFQGGDEQAVFK 122
           + L +N  +          ++   NLDY+V++N+KGIR THP+   IGKPFQGGDE+AV
Sbjct: 61  ELLLNNQPNAKTTAYTNSIASIYNLDYVVVMNMKGIRLTHPNPKNIGKPFQGGDEEAVLA 120

Query: 123 GKAIMSTAEGSLGKSLRYLIPVYDHQKQVGAIAVGLKLTTLGDLSQSSIKEFSKPLLISI 182
           GK ++STA+G+LGKSLRYL+PV+D  KQ+GAIAVG+KLTTL D++ +S + ++  LL+ +
Sbjct: 121 GKKVISTAKGTLGKSLRYLVPVFDGDKQIGAIAVGIKLTTLNDVALTSKRNYTLSLLLCL 180

Query: 183 LISLVVTSIISYGLKKQLHNLHPSDIFQHLEERNATLDQIQAAVFVIDQRHIIKRNNPAA 242
           LISL+VTS IS+ LK+QLH L PS+I+Q  EERNA LDQI+AAVFV+D+  I++  N A
Sbjct: 181 LISLLVTSFISFRLKRQLHQLEPSEIYQLFEERNAMLDQIEAAVFVVDKAGILQLCNQAG 240

Query: 243 SLLFKKEGQRDLFSGKLLESLIPQLKQDHFSKKTEQVLHFQGQDYLLSISPITVKTQNRG 302
             L  ++ Q    +G     L P   +    EQ+  +  +DYLL+ISPI VK  +RG
Sbjct: 241 QKLIARKCQLGKPTGNSFNYLFPDFPKLSLQEGHEQLFRYGEEDYLLAISPICVKNDHRG 300

Query: 303 YVVFLRNVTETLFTLDQLAHTTAYASALQAQTHQFMNQLHVIYGLADIEYYDELKIYLKE 362
           +++F+R   + + TLDQLA+TTAYASALQAQTH+FMNQLHVIYGL DI YYD+LKIYL
Sbjct: 301 HIIFMREAVKAIDTLDQLAYTTAYASALQAQTHKFMNQLHVIYGLVDIAYYDQLKIYLDS 360

Query: 363 LLEPQNEFLARLSMLVREPRLASFIIGEREKFAEKHINLSTEILVEIPTKSTVEDVNNYL 422
           +LEP+NE L  LS+LV+EP LASF+IGE+EK+ E +++L  ++L EIP  +T   +NN L
Sbjct: 361 ILEPENEILTSLSVLVKEPLLASFLIGEQEKYQELNVHLKIDVLSEIPHSATKNQLNNGL 420

Query: 423 LLHRYINTKILTLLNSTTLVSLRLNYQNNLIETDYQWENEKWL-LNDYHQYFNDAYFQQL 481
           +++R+I+T +LT L   +LV    + QN+LI   +    + W+ L      F+  YFQQL
Sbjct: 421 MIYRFIHTNLLTTLRPKSLVLSIQHDQNHLI--SHYTLTDNWIDLERVQPIFDLPYFQQL 478

Query: 482 LVDSRATY 489
           L D+ + +
Sbjct: 479 LTDTNSQF 486
```

SEQ ID 9004 (GBS130d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 123 (lane 8-10; MW 63kDa) and in Figure 184 (lane 4; MW 63kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 123 (lane 11; MW 38kDa) and in Figure 181 (lane 7; MW 38kDa).

GBS130d-GST was purified as shown in Figure 237, lane 11. GBS130d-His was purified as shown in Figure 233, 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 2247

A DNA sequence (GBSx2368) was identified in *S.agalactiae* <SEQ ID 6949> which encodes the amino acid sequence <SEQ ID 6950>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.52   Transmembrane  364 - 380 ( 353 - 386)
   INTEGRAL    Likelihood = -9.66   Transmembrane   33 -  49 (  26 -  57)
   INTEGRAL    Likelihood = -7.80   Transmembrane   87 - 103 (  82 - 105)
   INTEGRAL    Likelihood = -6.85   Transmembrane  153 - 169 ( 144 - 174)
   INTEGRAL    Likelihood = -4.41   Transmembrane  301 - 317 ( 300 - 318)
   INTEGRAL    Likelihood = -2.81   Transmembrane  216 - 232 ( 212 - 235)
   INTEGRAL    Likelihood = -2.39   Transmembrane  120 - 136 ( 120 - 136)
   INTEGRAL    Likelihood = -1.65   Transmembrane   57 -  73 (  56 -  73)
   INTEGRAL    Likelihood = -1.17   Transmembrane  428 - 444 ( 428 - 444)
   INTEGRAL    Likelihood = -0.32   Transmembrane  276 - 292 ( 276 - 292)

----- Final Results -----
               bacterial membrane --- Certainty=0.5607(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:AAB18291 GB:U35658 L-malate permease [Streptococcus bovis]
 Identities = 329/428 (76%), Positives = 375/428 (86%)

Query: 18  DLKAKLFHIKIGSVPLPVYVCLALLILLAGFLQKLPVNMLGGFAVILTMGWFLGTIGASI 77
           D + KL   +IGSV LPVY+  A +IL+   L++LPVNMLGGFAVILTMGW LGTIG +I
Sbjct: 14  DWRNKLTKTRIGSVTLPVYLVTASIILVTALLEQLPVNMLGGFAVILTMGWLLGTIGGNI 73

Query: 78  PGFKNFGGPAILSLLVPSILVFFNLINKNVLESTNMLMKQANFLYFYIACLVSGSILGMN 137
           P  K+FGGPAILSLLVPSI+VFFNL+N+NVL+ST++LMKQANFLYFYIACLV GSILGMN
Sbjct: 74  PILKHFGGPAILSLLVPSIMVFFNLLNQNVLDSTDILMKQANFLYFYIACLVCGSILGMN 133

Query: 138 RKMLIQGLLRMIFPMLLGMVCAMMVGTFVGVILGLEWRHTLFYIVTPVLAGGIGEGILPL 197
           RK+L+QGL+RMI PM LGM+ AM VGT VG +LGL W+H+LFYIVTPVLAGGIGEGILPL
Sbjct: 134 RKILVQGLMRMIVPMALGMILAMGVGTLVGTLLGLGWKHSLFYIVTPVLAGGIGEGILPL 193

Query: 198 SLGYSSITGVASEQLVAQLIPATIIGNFFAILCTALLNRLGEKKPHLSGQGQLVRLNKGE 257
           SLGYS+ITG+ SEQLV QLIPATIIGNFFAI+C+ LL+RLGEK+P LSGQGQL+++   +
Sbjct: 194 SLGYSAITGLPSEQLVGQLIPATIIGNFFAIMCSGLLSRLGEKRPELSGQGQLIKITNSD 253

Query: 258 DMSDIIADHSGPIDVKKMGGGVLTACSLFIFGHLLQQLTGFPGPVLMIVAAAILKYINVI 317
           D+SD + +   PIDVK MG GVL AC+LFI G LLQ LTGFPGPVLMIV AA LKY+NV+
Sbjct: 254 DLSDALEEDKAPIDVKLMGAGVLIACTLFITGGLLQHLTGFPGPVLMIVVAAFLKYLNVV 313

Query: 318 PRETQNGAKQLYKFISGNFTFPLMAGLGLLYIPLKDVVATLSIQYFIVVISVVFTVISVG 377
           P+ETQ G+KQLYKFISGNFTFPLM GLG+LYIPLKDVV  LS QYF+VVISVVFTVI+ G
Sbjct: 314 PKETQRGSKQLYKFISGNFTFPLMVGLGMLYIPLKDVVGMLSWQYFVVVISVVFTVIATG 373

Query: 378 FFVSRFLNMNPVEAGIISACQSGMGGTGDVAILSTADRMNLMPFAQVATRLGGAITVITM 437
           FFVSRF+NMNPVEA I+SACQSGMGGTGDVAILSTA+RM LMPFAQVATRLGGAITVITM
Sbjct: 374 FFVSRFMNMNPVEAAIVSACQSGMGGTGDVAILSTANRMTLMPFAQVATRLGGAITVITM 433

Query: 438 TAILRMLF 445
           TAI RMLF
Sbjct: 434 TAIFRMLF 441
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6951> which encodes the amino acid sequence <SEQ ID 6952>. Analysis of this protein sequence reveals the following:

```
      Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.89   Transmembrane  361 - 377 ( 350 - 383)
   INTEGRAL    Likelihood = -7.43   Transmembrane   84 - 100 (  79 - 102)
   INTEGRAL    Likelihood = -6.16   Transmembrane  150 - 166 ( 137 - 171)
   INTEGRAL    Likelihood = -4.88   Transmembrane   30 -  46 (  24 -  48)
   INTEGRAL    Likelihood = -4.35   Transmembrane  299 - 315 ( 297 - 316)
   INTEGRAL    Likelihood = -4.14   Transmembrane  117 - 133 ( 115 - 134)
   INTEGRAL    Likelihood = -3.19   Transmembrane   54 -  70 (  51 -  75)
   INTEGRAL    Likelihood = -2.92   Transmembrane  425 - 441 ( 425 - 442)
   INTEGRAL    Likelihood = -2.81   Transmembrane  213 - 229 ( 209 - 232)
   INTEGRAL    Likelihood = -2.44   Transmembrane  273 - 289 ( 271 - 290)

----- 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>
```

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

```
>GP:AAB18291 GB:U35658 L-malate permease [Streptococcus bovis]
 Identities = 344/443 (77%), Positives = 394/443 (88%), Gaps = 6/443 (1%)

Query: 4   ISKKMPQKDLSEHSKAWQNR----RIGSVPLPVYLVLATLILVTGWLQQLPVNMLGGFAV 59
           + KK+P    +E    W+N+    RIGSV LPVYLV A++ILVT  L+QLPVNMLGGFAV
Sbjct: 1   MEKKLPATAANETD--WRNKLTKTRIGSVTLPVYLVTASIILVTALLEQLPVNMLGGFAV 58

Query: 60  ILTLGWLLGTIGATIPGLKHFGGPAILSLLVPSILVFFNLLNPNVLEATNVLMKQANFLY 119
           ILT+GWLLGTIG  IP LKHFGGPAILSLLVPSI+VFFNLLN NVL++T++LMKQANFLY
Sbjct: 59  ILTMGWLLGTIGGNIPILKHFGGPAILSLLVPSIMVFFNLLNQNVLDSTDILMKQANFLY 118

Query: 120 FYIACLVCGSILGMNRKILIQGLFRMIIPMLLGMVCAMGVGTLVGVILGLDWQHTLFYVV 179
           FYIACLVCGSILGMNRKIL+QGL RMI+PM LGM+ AMGVGTLVG +LGL W+H+LFY+V
Sbjct: 119 FYIACLVCGSILGMNRKILVQGLMRMIVPMALGMILAMGVGTLVGTLLGLGWKHSLFYIV 178

Query: 180 TPVLAGGIGEGILPLSLGYSAITGVGSEQLVAQLIPATIIGNFFAILCTALLNRFGEKHP 239
           TPVLAGGIGEGILPLSLGYSAITG+ SEQLV QLIPATIIGNFFAI+C+ LL+R GEK P
Sbjct: 179 TPVLAGGIGEGILPLSLGYSAITGLPSEQLVGQLIPATIIGNFFAIMCSGLLSRLGEKRP 238

Query: 240 SYSGQGQLVKIGHSEDMSDALKDNSGALDVKLMGAGVLTACSLFIAGGLLQHLTDFPGPV 299
             SGQGQL+KI +S+D+SDAL+++    +DVKLMGAGVL AC+LFI GGLLQHLT FPGPV
Sbjct: 239 ELSGQGQLIKITNSDDLSDALEEDKAPIDVKLMGAGVLIACTLFITGGLLQHLTGFPGPV 298

Query: 300 LMIILAAFLKYLNVIPQETQNGAKQLYKFISSNFTFPLMAGLGLLYIPLKEVVATLSWQY 359
           LMI++AAFLKYLNV+P+ETQ G+KQLYKFIS NFTFPLM GLG+LYIPLK+VV  LSWQY
Sbjct: 299 LMIVVAAFLKYLNVVPKETQRGSKQLYKFISGNFTFPLMVGLGMLYIPLKDVVGMLSWQY 358

Query: 360 FIVVISVVLTVVSVGFFVSRFLNMSPVEAAIISACQSGMGGTGDVAILSTADRMNLMPFA 419
           F+VVISVV TV++ GFFVSRF+NM+PVEAAI+SACQSGMGGTGDVAILSTA+RM LMPFA
Sbjct: 359 FVVVISVVFTVIATGFFVSRFMNMNPVEAAIVSACQSGMGGTGDVAILSTANRMTLMPFA 418

Query: 420 QVATRLGGAITVITMTAILRIIF 442
           QVATRLGGAITVITMTAI R++F
Sbjct: 419 QVATRLGGAITVITMTAIFRMLF 441
```

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

```
 Identities = 356/419 (84%), Positives = 385/419 (90%)

Query: 27  KIGSVPLPVYVCLALLILLAGFLQKLPVNMLGGFAVILTMGWFLGTIGASIPGFKNFGGP 86
           +IGSVPLPVY+ LA LIL+ G+LQ+LPVNMLGGFAVILT+GW LGTIGA+IPG K+FGGP
Sbjct: 24  RIGSVPLPVYLVLATLILVTGWLQQLPVNMLGGFAVILTLGWLLGTIGATIPGLKHFGGP 83
```

```
Query: 87   AILSLLVPSILVFFNLINKNVLESTNMLMKQANFLYFYIACLVSGSILGMNRKMLIQGLL 146
            AILSLLVPSILVFFNL+N NVLE+TN+LMKQANFLYFYIACLV GSILGMNRK+LIQGL
Sbjct: 84   AILSLLVPSILVFFNLLNPNVLEATNVLMKQANFLYFYIACLVCGSILGMNRKILIQGLF 143

Query: 147  RMIFPMLLGMVCAMMVGTFVGVILGLEWRHTLFYIVTPVLAGGIGEGILPLSLGYSSITG 206
            RMI PMLLGMVCAM VGT VGVILGL+W+HTLFY+VTPVLAGGIGEGILPLSLGYS+ITG
Sbjct: 144  RMIIPMLLGMVCAMGVGTLVGVILGLDWQHTLFYVVTPVLAGGIGEGILPLSLGYSAITG 203

Query: 207  VASEQLVAQLIPATIIGNFFAILCTALLNRLGEKKPHLSGQGQLVRLNKGEDMSDIIADH 266
            V SEQLVAQLIPATIIGNFFAILCTALLNR GEK P  SGQGQLV++   EDMSD + D+
Sbjct: 204  VGSEQLVAQLIPATIIGNFFAILCTALLNRFGEKHPSYSGQGQLVKIGHSEDMSDALKDN 263

Query: 267  SGPIDVKKMGGGVLTACSLFIFGHLLQQLTGFPGPVLMIVAAAILKYINVIPRETQNGAK 326
            SG +DVK MG GVLTACSLFI G LLQ LT FPGPVLMI+ AA LKY+NVIP+ETQNGAK
Sbjct: 264  SGALDVKLMGAGVLTACSLFIAGGLLQHLTDFPGPVLMIILAAFLKYLNVIPQETQNGAK 323

Query: 327  QLYKFISGNFTFPLMAGLGLLYIPLKDVVATLSIQYFIVVISVVFTVISVGFFVSRFLNM 386
            QLYKFIS NFTFPLMAGLGLLYIPLK+VVATLS QYFIVVISVV TV+SVGFFVSRFLNM
Sbjct: 324  QLYKFISSNFTFPLMAGLGLLYIPLKEVVATLSWQYFIVVISVVLTVVSVGFFVSRFLNM 383

Query: 387  NPVEAGIISACQSGMGGTGDVAILSTADRMNLMPFAQVATRLGGAITVITMTAILRMLF 445
            +PVEA IISACQSGMGGTGDVAILSTADRMNLMPFAQVATRLGGAITVITMTAILR++F
Sbjct: 384  SPVEAAIISACQSGMGGTGDVAILSTADRMNLMPFAQVATRLGGAITVITMTAILRIIF 442
```

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

Example 2248

A DNA sequence (GBSx2369) was identified in *S.agalactiae* <SEQ ID 6953> which encodes the amino acid sequence <SEQ ID 6954>. This protein is predicted to be malic enzyme (mae). Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.28   Transmembrane  164 - 180 ( 164 - 181)

----- 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:AAB07709 GB:U35659 malic enzyme [Streptococcus bovis]
 Identities = 285/386 (73%), Positives = 332/386 (85%), Gaps = 1/386 (0%)

Query: 2    SENLGQLAINQARENGGKLEVISKVKVEDKRDLSIAYTPGVASVSSAIAEDVELAYELTT 61
             ++++ +LAI QA++ GGKLEV  KV +E K DL IAYTPGVA+VSSAI E  E AYELTT
Sbjct: 3    TKDVKELAIEQAKKFGGKLEVCPKVPIETKADLGIAYTPGVAAVSSAIYEKKERAYELTT 62

Query: 62   KKNTVAVVSDGSAVLGLGDIGPEAAMPVMEGKAALFKRFANVDAVPIVLKTNDTEEIISI 121
             KKNTVAV+SDGSAVLGLG+IGPEAAMPVMEGKAALFKRFA VD++P+VL T DTEEII
Sbjct: 63   KKNTVAVISDGSAVLGLGNIGPEAAMPVMEGKAALFKRFAGVDSIPLVLDTQDTEEIIQT 122

Query: 122  VKAISPTFGGINLEDISAPRCFEIEQRLIEECDIPVFHDDQHGTAIVVLAALFNSLKLVK 181
             VK ++PTFGGINLEDISAPRCFEIEQRLI+E DIPVFHDDQHGTAIVVLAAL+NSLKL+
Sbjct: 123  VKFLAPTFGGINLEDISAPRCFEIEQRLIDELDIPVFHDDQHGTAIVVLAALYNSLKLIN 182

Query: 182  KDIEDIRVVVNGGGSAGLSITRKLLSAGAKHVTVVDRFGIINDKDRESLAPHHKAIAKLT 241
             K IEDI VV+NGGGSAGLSITRK L+AG KH+ +VDR GI+++ D  +L PHH  IAKLT
Sbjct: 183  KKIEDIHVVINGGGSAGLSITRKFLAAGVKHIIIVDRTGILSETD-TALPPHHAEIAKLT 241

Query: 242  NREFQSGSLEDALENADVFIGVSAPEALHAEWISKMADKPIVFAMANPIPEIYPDQALKA 301
             NRE ++G L  ALE ADVF+GVSAP  L  EWI +M ++P++FAMANP+PEI+PD+AL A
```

```
Sbjct: 242 NREHRTGDLATALEGADVFVGVSAPGVLKPEWIQQMNEQPVIFAMANPVPEIFPDEALAA 301

Query: 302 GAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAARGIASLIPEEELST 361
           GAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAK IT+EMQIAAA+GIA LIP+ EL+
Sbjct: 302 GAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKKITIEMQIAAAKGIAKLIPDNELTP 361

Query: 362 THIIPNAFQNDVADVVAKSVSNAVQK 387
           T+IIP+ FQ  VA VVA+SV NAV++
Sbjct: 362 TNIIPDPFQEGVAKVVAESVRNAVKE 387
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6955> which encodes the amino acid sequence <SEQ ID 6956>. Analysis of this protein sequence reveals the following:

```
     Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.44   Transmembrane  164 - 180 ( 164 - 181)
   INTEGRAL    Likelihood = -1.75   Transmembrane   94 - 110 (  94 - 110)

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

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

```
>GP:AAB07709 GB:U35659 malic enzyme [Streptococcus bovis]
 Identities = 289/379 (76%), Positives = 334/379 (87%), Gaps = 1/379 (0%)

Query: 7   QLALEQAKTFGGKLEVQPKVDIKTKHDLSIAYTPGVASVSSAIAKDKTLAYDLTTKKNTV 66
           +LA+EQAK FGGKLEV PKV I+TK DL IAYTPGVA+VSSAI + K  AY+LTTKKNTV
Sbjct: 8   ELAIEQAKKFGGKLEVCPKVPIETKADLGIAYTPGVAAVSSAIYEKKERAYELTTKKNTV 67

Query: 67  AVISDGTAVLGLGDIGPEAAMPVMEGKAALFKAFAGVDAIPIVLDTKDTEEIISIVKALA 126
           AVISDG+AVLGLG+IGPEAAMPVMEGKAALFK FAGVD+IP+VLDT+DTEEII  VK LA
Sbjct: 68  AVISDGSAVLGLGNIGPEAAMPVMEGKAALFKRFAGVDSIPLVLDTQDTEEIIQTVKFLA 127

Query: 127 PTFGGINLEDISAPRCFEIEQRLIKECHIPVFHDDQHGTAIVVLAAIFNSLKLLKKSLDE 186
           PTFGGINLEDISAPRCFEIEQRLI E  IPVFHDDQHGTAIVVLAA++NSLKL+ K +++
Sbjct: 128 PTFGGINLEDISAPRCFEIEQRLIDELDIPVFHDDQHGTAIVVLAALYNSLKLINKKIED 187

Query: 187 VSIVVNGGGSAGLSITRKLLAAGATKVTVVDKFGIINEQEAAQLAPHHLDIAKVTNREFK 246
           + +V+NGGGSAGLSITRK LAAG   + +VD+ GI++E + A L PHH +IAK+TNRE +
Sbjct: 188 IHVVINGGGSAGLSITRKFLAAGVKHIIIVDRTGILSETDTA-LPPHHAEIAKLTNREHR 246

Query: 247 SGTLEDALEGADIFIGVSAPGVLKAEWISKMAARPVIFAMANPIPEIYPDEALEAGAYIV 306
           +G L  ALEGAD+F+GVSAPGVLK EWI +M  +PVIFAMANP+PEI+PDEAL AGAYIV
Sbjct: 247 TGDLATALEGADVFVGVSAPGVLKPEWIQQMNEQPVIFAMANPVPEIFPDEALAAGAYIV 306

Query: 307 GTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAAKGIASLVPDDALSTTNIIP 366
           GTGRSDFPNQINNVLAFPGIFRGALDARAK IT+EMQIAAAKGIA L+PD+ L+ TNIIP
Sbjct: 307 GTGRSDFPNQINNVLAFPGIFRGALDARAKKITIEMQIAAAKGIAKLIPDNELTPTNIIP 366

Query: 367 DAFKEGVAEIVAKSVRSVV 385
           D F+EGVA++VA+SVR+ V
Sbjct: 367 DPFQEGVAKVVAESVRNAV 385
```

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

```
 Identities = 306/387 (79%), Positives = 349/387 (90%)

Query: 1   MSENLGQLAINQARENGGKLEVISKVKVEDKRDLSIAYTPGVASVSSAIAEDVELAYELT 60
           M   LGQLA+ QA+  GGKLEV  KV ++ K DLSIAYTPGVASVSSAIA+D  LAY+LT
Sbjct: 1   MKNQLGQLALEQAKTFGGKLEVQPKVDIKTKHDLSIAYTPGVASVSSAIAKDKTLAYDLT 60

Query: 61  TKKNTVAVVSDGSAVLGLGDIGPEAAMPVMEGKAALFKRFANVDAVPIVLKTNDTEEIIS 120
           TKKNTVAV+SDG+AVLGLGDIGPEAAMPVMEGKAALFK FA VDA+PIVL T DTEEIIS
Sbjct: 61  TKKNTVAVISDGTAVLGLGDIGPEAAMPVMEGKAALFKAFAGVDAIPIVLDTKDTEEIIS 120
```

```
Query: 121 IVKAISPTFGGINLEDISAPRCFEIEQRLIEECDIPVFHDDQHGTAIVVLAALFNSLKLV 180
           IVKA++PTFGGINLEDISAPRCFEIEQRLI+EC IPVFHDDQHGTAIVVLAA+FNSLKL+
Sbjct: 121 IVKALAPTFGGINLEDISAPRCFEIEQRLIKECHIPVFHDDQHGTAIVVLAAIFNSLKLL 180

Query: 181 KKDIEDIRVVVNGGGSAGLSITRKLLSAGAKHVTVVDRFGIINDKDRESLAPHHKAIAKL 240
           KK ++++ +VVNGGGSAGLSITRKLL+AGA  VTVVD+FGIIN+++   LAPHH  IAK+
Sbjct: 181 KKSLDEVSIVVNGGGSAGLSITRKLLAAGATKVTVVDKFGIINEQEAAQLAPHHLDIAKV 240

Query: 241 TNREFQSGSLEDALENADVFIGVSAPEALHAEWISKMADKPIVFAMANPIPEIYPDQALK 300
           TNREF+SG+LEDALE AD+FIGVSAP  L AEWISKMA +P++FAMANPIPEIYPD+AL+
Sbjct: 241 TNREFKSGTLEDALEGADIFIGVSAPGVLKAEWISKMAARPVIFAMANPIPEIYPDEALE 300

Query: 301 AGAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAARGIASLIPEEELS 360
           AGAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAA+GIASL+P++ LS
Sbjct: 301 AGAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAAKGIASLVPDDALS 360

Query: 361 TTHIIPNAFQNDVADVVAKSVSNAVQK 387
           TT+IIP+AF+  VA++VAKSV + V K
Sbjct: 361 TTNIIPDAFKEGVAEIVAKSVRSVVLK 387
```

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

Example 2249

A DNA sequence (GBSx2370) was identified in *S.agalactiae* <SEQ ID 6957> which encodes the amino acid sequence <SEQ ID 6958>. This protein is predicted to be Bta. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.02   Transmembrane   29 -  45 (  29 -  45)

----- Final Results -----
               bacterial membrane --- Certainty=0.1808(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:AAD56628 GB:AF165218 Bta [Streptococcus pneumoniae]
 Identities = 35/112 (31%), Positives = 63/112 (56%)

Query: 1   MYSFEELLATMTLITAAEIEDKIDSNQDFVLFIGRISCPFCHLFVPKIVEVADEDEFELF 60
           M  F + +  + + T    ++ +D  +    FIGR +CP+C  F   + V  E +  ++
Sbjct: 1   MEQFLDNIKDLEVTTVVRAQEALDKKETATFFIGRKTCPYCRKFAGTLSGVVAETKAHIY 60

Query: 61  HLDSEDFDHWTANKEFRNKYDIPTVPGLMVVKNGTIKVKCDSKMTKEEIREF 112
            ++SE+       + FR++Y IPTVPG + + +G I V+CDS M+ +EI++F
Sbjct: 61  FINSEEASQLNDLQAFRSRYGIPTVPGFVHITDGQINVRCDSSMSAQEIKDF 112
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6959> which encodes the amino acid sequence <SEQ ID 6960>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0900(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 = 39/111 (35%), Positives = 66/111 (59%)

Query: 3   SFEELLATMTLITAAEIEDKIDSNQDFVLFIGRISCPFCHLFVPKIVEVADEDEFELFHL 62
           +FEE++A      + AE+   I S +D ++F+GR SCP+C   F PK+ +VA +++ E++ +
Sbjct: 11  TFEEIVANFIPSSVAEVTSAIASGKDMIVFLGRSSCPYCRRFAPKLAQVATDNQKEVYFV 70

Query: 63  DSEDFDHWTANKEFRNKYDIPTVPGLMVVKNGTIKVKCDSKMTKEEIREFI 113
           DSE+          FR  Y + TVP L+V  +    +  CDS +T ++I  F+
Sbjct: 71  DSENAADAAELAAFRENYQLVTVPALLVSYDQHQRAVCDSSLTPDDILAFL 121
```

SEQ ID 6958 (GBS427) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 5; MW 16.2kDa).

GBS427-His was purified as shown in Figure 214, lane 8.

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

Example 2250

A DNA sequence (GBSx2371) was identified in *S.agalactiae* <SEQ ID 6961> which encodes the amino acid sequence <SEQ ID 6962>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.75   Transmembrane    2 -  18 (   1 -  21)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9437> which encodes amino acid sequence <SEQ ID 9438> was also identified.

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

```
>GP:BAA11328 GB:D78257 ORF11 [Enterococcus faecalis]
 Identities = 36/80 (45%), Positives = 58/80 (72%)

Query: 1  MSLPIIMLVVMVGMMFFMQRQQKKQAQERQKQLNAVQKGDEIVTIGGLFGVVDEVNTEAQ 60
          M L +IML+V+V M F++ R QKKQ +ERQ  LN +Q GD +VTIGGL GV+ E++++ +
Sbjct: 1  MKLMLIMLLVIVAMYFYLFRTQKKQQKERQDFLNNLQPGDAVVTIGGLHGVISEISSDKK 60

Query: 61 RMVLDVDGVYLTFELAAIKS 80
          ++ LD +G +  F+  +I++
Sbjct: 61 KVTLDCEGAFFDFDQQSIRT 80
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6963> which encodes the amino acid sequence <SEQ ID 6964>. Analysis of this protein sequence reveals the following:

```
     Possible site: 60
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.10   Transmembrane    3 -  19 (   1 -  22)
   INTEGRAL    Likelihood = -3.03   Transmembrane   63 -  79 (  63 -  79)

----- Final Results -----
               bacterial membrane --- Certainty=0.3442(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:BAA11328 GB:D78257 ORF11 [Enterococcus faecalis]
 Identities = 29/75 (38%), Positives = 52/75 (68%)

Query: 6   ILMFVVMLGLIWFMQRQQKKQAQERQNQLNAIEKGDEVVTIGGMFAIVDEVDTTAKKIVL 65
           ++M +V++ + +++ R QKKQ +ERQ+ LN ++ GD VVTIGG+  ++ E+ +  KK+ L
Sbjct: 5   LIMLLVIVAMYFYLFRTQKKQQKERQDFLNNLQPGDAVVTIGGLHGVISEISSDKKKVTL 64

Query: 66 DVDGVFLTFELLAIK 80
           D +G F  F+  +I+
Sbjct: 65 DCEGAFFDFDQQSIR 79
```

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

```
 Identities = 63/90 (70%), Positives = 80/90 (88%)

Query: 4   PIIMLVVMVGMMFFMQRQQKKQAQERQKQLNAVQKGDEIVTIGGLFGVVDEVNTEAQRMV 63
           PI+M VVM+G+++FMQRQQKKQAQERQ QLNA++KGDE+VTIGG+F +VDEV+T A+++V
Sbjct: 5   PILMFVVMLGLIWFMQRQQKKQAQERQNQLNAIEKGDEVVTIGGMFAIVDEVDTTAKKIV 64

Query: 64 LDVDGVYLTFELAAIKSVVSKAATPTEPVE 93
           LDVDGV+LTFEL AIK +V+KA T T  VE
Sbjct: 65 LDVDGVFLTFELLAIKRIVTKATTETTLVE 94
```

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

Example 2251

A DNA sequence (GBSx2372) was identified in *S.agalactiae* <SEQ ID 6965> which encodes the amino acid sequence <SEQ ID 6966>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq

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

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 2252

A DNA sequence (GBSx2373) was identified in *S.agalactiae* <SEQ ID 6967> which encodes the amino acid sequence <SEQ ID 6968>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38   Transmembrane  164 - 180 ( 164 - 180)

----- 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 = 72/216 (33%), Positives = 120/216 (55%), Gaps = 1/216 (0%)

Query: 14  AQALEARGQKLYSVANRTYDKGLEFATKYGIQKVYDHIDQVFEDPEVDIIYISTPHNTHI 73
           A  ++    ++ +VA+RT       FA ++GI + Y   + +  D +VD++Y++TPH+ H
Sbjct: 25  ADLVDLPDAEVVAVASRTEASAKTFAERFGIPRAYGGWETLARDEDVDVVYVATPHSAHR 84

Query: 74  SFLRKALANGKHVLCEKSITLNSTELKEAIDLAETNHVVLAEAMTIFHMPIYRQLKTLVD 133
           +     L  G++VLCEK  TLN+ E  E + LA  N V L EAM ++  P+ R+LK LV
Sbjct: 85  TAAGLCLEAGRNVLCEKPFTLNAREAAELVALARENGVFLMEAMWMYCNPLVRRLKELVA 144

Query: 134 SGKLGPLKMIQMNFGSYKEYDMTNRFFSRDLAGGALLDIGVYALSCIRWFMSEAPHNITS 193
            G +G ++ +Q +FG    +   +R       GGALLD+GVY +S  +  + E P ++ +
Sbjct: 145 DGAIGEVRSLQADFGLAGPFPAAHRLRDPAQGGGALLDLGVYPVSFAQLLLGE-PTDVAA 203

Query: 194 QVTFAPTGVDEQVGILLTNPANEMATVSLSLHAKQP 229
           +   + GVD Q G LL+   + +A++  S+    P
Sbjct: 204 RAVLSEEGVDLQTGALLSYGNDALASIHCSITGGTP 239
```

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

Example 2253

A DNA sequence (GBSx2374) was identified in *S.agalactiae* <SEQ ID 6969> which encodes the amino acid sequence <SEQ ID 6970>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4957(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Example 2254

A DNA sequence (GBSx2375) was identified in *S.agalactiae* <SEQ ID 6971> which encodes the amino acid sequence <SEQ ID 6972>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Example 2255

A DNA sequence (GBSx2376) was identified in *S.agalactiae* <SEQ ID 6973> which encodes the amino acid sequence <SEQ ID 6974>. This protein is predicted to be a host cell surface-exposed lipoprotein. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.75   Transmembrane    9 -  25 (   5 -  28)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9005> which encodes amino acid sequence <SEQ ID 9006> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length of UR:   24
     Peak Value of UR:   2.84
     Net Charge of CR: 2
McG: Discrim Score:     10.29
GvH: Signal Score (-7.5): -4.34
     Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 1 value:  -7.75 threshold:  0.0
   INTEGRAL    Likelihood = -7.75   Transmembrane    5 -  21 (   1 -  24)
   PERIPHERAL  Likelihood = 13.31       86
 modified ALOM score:   2.05
icm1 HYPID: 7  CFP: 0.410

*** Reasoning Step: 3

----- 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:AAC03455 GB:AF020798 putative host cell surface-exposed
          lipoprotein [Streptococcus thermophilus bacteriophage TP-J34]
 Identities = 40/102 (39%), Positives = 63/102 (61%), Gaps = 10/102 (9%)

Query: 101 KNALISAKIYSKTMNLSKQSIFEQLYSESPDKATHSDKFTKEESQYAIDHLKVDFKENAL 160
           + A+  AK Y+ T+++SK+ +  QL S          DK++++ S YA+++  +D+ + AL
Sbjct: 51  RTAVSKAKQYASTVHMSKEELRSQLVS--------FDKYSQDASDYAVENSGIDYNKQAL 102

Query: 161 ETAKSYQSSSSLSKEEIYKQLTSTLGDKFTNDEAQYAVDHLK 202
           E AK YQ + S+S + I  QL S   DKFT +EA YAV +LK
Sbjct: 103 EKAKQYQDTLSMSPDAIRDQLVSF--DKFTQEEADYAVANLK 142
 Identities = 40/112 (35%), Positives = 64/112 (56%), Gaps = 9/112 (8%)

Query: 41  KKAKIKFNKTQKKIVKKAREYAKSGHMSKDSIIEKLKKDSKKYRQEDINFVINNLKVDYK 100
           + ++ K  K  +  V KA++YA + HMSK+ +  +L    K Y Q+  ++ + N  +DY
Sbjct: 40  QSSESKVPKEYRTAVSKAKQYASTVHMSKEELRSQLVSFDK-YSQDASDYAVENSGIDYN 98

Query: 101 KNALISAKIYSKTMNLSKQSIFEQLYSESPDKATHSDKFTKEESQYAIDHLK 152
           K AL  AK Y  T+++S  +I +QL S          DKFT+EE+ YA+ +LK
Sbjct: 99  KQALEKAKQYQDTLSMSPDAIRDQLVS--------FDKFTQEEADYAVANLK 142
```

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

SEQ ID 9006 (GBS122) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 6; MW 21.9kDa).

GBS122-His was purified as shown in Figure 202, lane 8.

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

Example 2256

A DNA sequence (GBSx2377) was identified in *S.agalactiae* <SEQ ID 6975> which encodes the amino acid sequence <SEQ ID 6976>. This protein is predicted to be transposase (orfA). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2830(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:CAB90833 GB:AJ250837 hypothetical protein [Streptococcus dysgalactiae]
 Identities = 91/96 (94%), Positives = 93/96 (96%)

Query: 1  MSRKVRRHFTDDFKQQIVDLYNVGRKRSSLIKVYELTPSTFDKWVRQAKTTGSFKSIDNL 60
          MSRK+RRHFTDDFKQQIVDLYN GRKRSSLIK YELTPSTFDKWVRQAKTTGSFKS+DNL
Sbjct: 1  MSRKIRRHFTDDFKQQIVDLYNAGRKRSSLIKEYELTPSTFDKWVRQAKTTGSFKSVDNL 60

Query: 61 TDEQRELIELRKHNKELEMQLDILKQAAVIMAQKGK 96
          TDEQRELIELRK NKELEMQLDILKQAAVIMAQKGK
Sbjct: 61 TDEQRELIELRKRNKELEMQLDILKQAAVIMAQKGK 96
```

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

Example 2257

A DNA sequence (GBSx2378) was identified in *S.agalactiae* <SEQ ID 6977> which encodes the amino acid sequence <SEQ ID 6978>. This protein is predicted to be transposase (orfB). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2618(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 9915> which encodes amino acid sequence <SEQ ID 9916> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9903> which encodes amino acid sequence <SEQ ID 9904> was also identified.

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

```
>GP:CAB90834 GB:AJ250837 putative transposase [Streptococcus dysgalactiae]
 Identities = 243/259 (93%), Positives = 250/259 (95%)

Query: 1   MCRWLNMPHSSYYYQAVESVSETEFEETIKRIFLDSESRYGSRKIKICLNNEGITLSRRR 60
            MCRWLN+P SSYYY+AVE VSE E EE+IK IFL+S++RYGSRKIKICLNNEGITLSRRR
Sbjct: 1   MCRWLNIPRSSYYYKAVEPVSEAELEESIKAIFLESKARYGSRKIKICLNNEGITLSRRR 60

Query: 61  IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFKQERPLQALVTDLTYVRVGNR 120
            IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFK ERPLQALVTDLTYVRVGNR
Sbjct: 61  IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFKPERPLQALVTDLTYVRVGNR 120

Query: 121 WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPYALTKVKMFHSDRGKEFDNQLID 180
            WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPY LTKVKMFHSDRGKEF+NQLID
Sbjct: 121 WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPYPLTKVKMFHSDRGKEFNNQLID 180

Query: 181 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFQLLEELALKTKDYVHWWNY 240
            EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFQ LEELALKTK YVHWWNY
Sbjct: 181 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFQSLEELALKTKAYVHWWNY 240

Query: 241 HRIHGSLNYQTPMTKRLIA 259
            HRIHGSLNYQTPMTKRLIA
Sbjct: 241 HRIHGSLNYQTPMTKRLIA 259
```

There is also homology to SEQ ID 32.

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

Example 2258

A DNA sequence (GBSx2379) was identified in *S.agalactiae* <SEQ ID 6979> which encodes the amino acid sequence <SEQ ID 6980>. This protein is predicted to be pXO1-128. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3684(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:AAD32432 GB:AF065404 pXO1-128 [Bacillus anthracis]
 Identities = 45/69 (65%), Positives = 52/69 (75%)

Query: 17 MKKAGKSNRVIMETLGIKNNSQIYTWMKWYENEELYRFHQGVGKQYTYGKGLEHLSEVEQ 76
           MKK   SNR IME LGIKN SQI TWMKWY  ++ YRF Q VGKQY+YGKG + LSE+EQ
Sbjct: 1  MKKESYSNRTIMEKLGIKNVSQIKTWMKWYRTDQTYRFQQPVGKQYSYGKGPKELSELEQ 60

Query: 77 LQLQVDLLK 85
           L+L+   LK
Sbjct: 61 LRLENKHLK 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 2259

A DNA sequence (GBSx2380) was identified in *S.agalactiae* <SEQ ID 6981> which encodes the amino acid sequence <SEQ ID 6982>. This protein is predicted to be transposase. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq

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

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 2260

A DNA sequence (GBSx2382) was identified in *S.agalactiae* <SEQ ID 6985> which encodes the amino acid sequence <SEQ ID 6986>. This protein is predicted to be Lmb. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> 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 DNA sequence was identified in *S.pyogenes* <SEQ ID 1595> which encodes the amino acid sequence <SEQ ID 1596>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> 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 = 302/306 (98%), Positives = 303/306 (98%)

Query: 1   MKKVFFLMAMVVSLVMIAGCDKSANPKQPTQGMSVVTSFYPMYAMTKEVSGDLNDVRMIQ 60
           MKK FFLMAMVVSLVMIAGCDKSANPKQPTQGMSVVTSFYPMYAMTKEVSGDLNDVRMIQ
Sbjct: 1   MKKGFFLMAMVVSLVMIAGCDKSANPKQPTQGMSVVTSFYPMYAMTKEVSGDLNDVRMIQ 60

Query: 61  SGAGIHSFEPSVNDVAAIYDADLFVYHSHTLEAWARDLDPNLKKSKVNVFEASKPLTLDR 120
           SGAGIHSFEPSVNDVAAIYDADLFVYHSHTLEAWARDLDPNLKKSKV+VFEASKPLTLDR
Sbjct: 61  SGAGIHSFEPSVNDVAAIYDADLFVYHSHTLEAWARDLDPNLKKSKVDVFEASKPLTLDR 120

Query: 121 VKGLEDMEVTQGIDPATLYDPHTWTDPVLAGEEAVNIAKELGHLDPKHKDSYTKKAKAFK 180
           VKGLEDMEVTQGIDPATLYDPHTWTDPVLAGEEAVNIAKELG LDPKHKDSYTK AKAFK
Sbjct: 121 VKGLEDMEVTQGIDPATLYDPHTWTDPVLAGEEAVNIAKELGRLDPKHKDSYTKNAKAFK 180

Query: 181 KEAEQLTEEYTQKFKKVRSKTFVTQHTAFSYLAKRFGLKQLGISGISPEQEPSPRQLKEI 240
```

```
            KEAEQLTEEYTQKFKKVRSKTFVTQHTAFSYLAKRFGLKQLGISGISPEQEPSPRQLKEI
Sbjct: 181 KEAEQLTEEYTQKFKKVRSKTFVTQHTAFSYLAKRFGLKQLGISGISPEQEPSPRQLKEI 240

Query: 241 QDFVKEYNVKTIFAEDNVNPKIAHAIAKSTGAKVKTLSPLEAAPSGNKTYLENLRANLEV 300
            QDFVKEYNVKTIFAEDNVNPKIAHAIAKSTGAKVKTLSPLEAAPSGNKTYLENLRANLEV
Sbjct: 241 QDFVKEYNVKTIFAEDNVNPKIAHAIAKSTGAKVKTLSPLEAAPSGNKTYLENLRANLEV 300

Query: 301 LYQQLK 306
            LYQQLK
Sbjct: 301 LYQQLK 306
```

There is also homology to SEQ ID 4.

SEQ ID 6986 (GBS189) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 2; MW 35.2kDa).

The GBS189-His fusion product was purified (Figure 204, lane 7) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 248A), FACS (Figure 248B), 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 2261

A DNA sequence (GBSx2383) was identified in *S.agalactiae* <SEQ ID 6987> which encodes the amino acid sequence <SEQ ID 6988>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4656(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:AAB41455 GB:U34956 phosphoribosylformylglycinamidine synthase
            [Mycobacterium tuberculosis]
 Identities = 73/237 (30%), Positives = 112/237 (46%), Gaps = 25/237 (10%)

Query: 43  GAGGVCVAIGELAD----GLEIDLDKVPLKYQGLNGTEIAISESQERMSVVVGPSDVDAF 98
           G  G+  A  ELA      G+ I LD VPL+ + +   E+  SESQERM  VV P +VDAF
Sbjct: 282 GGAGLSCATSELASAGDGGMTIQLDSVPLRAKEMTPAEVLCSESQERMCAVVSPKNVDAF 341

Query: 99  IAACNKENIDAVVVATVTEKPNLVMTWNGETIVDLERCFLDTNG------VRVVVDAKVV 152
           +A C K  + A V+  VT+   L +TW+GET+VD+     +  G      V       +
Sbjct: 342 LAVCRKWEVLATVIGEVTDGDRLQITWHGETVVDVPPRTVAHEGPVYQRPVARPDTQDAL 401

Query: 153 DKDLTVPEARTTSAETLEADMLKVLSDLNHASQKGLQTIFDSSVGRSTV--NHPIGGRYQ 210
           + D +   +R  + + L A +L +L   + S+  +   +D  V  +TV   H  GG  +
Sbjct: 402 NADRSAKLSRPVTGDELRATLLALLGSPHLCSRAFITEQYDRYVRGNTVLAEHADGGMLR 461

Query: 211 ITPTESSVQKLPVQYGVTTTASVMAQGYNPYIAEWSPYHGAAYAVIEATARLVATGA 267
           I   ES+ + + V    +   +++            PY GA  A+ EA   + TGA
Sbjct: 462 I--DESTGRGIAVSTDASGRYTLL----------DPYAGAQLALAEAYRNVAVTGA 505
```

There is also homology to SEQ ID 982.

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

Example 2262

A DNA sequence (GBSx2384) was identified in *S.agalactiae* <SEQ ID 6989> which encodes the amino acid sequence <SEQ ID 6990>. This protein is predicted to be 30S ribosomal protein S11 (rpsK). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0598(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 9281> which encodes amino acid sequence <SEQ ID 9282> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10919> which encodes amino acid sequence <SEQ ID 10920> was also identified.

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

```
>GP:CAB11918 GB:Z99104 ribosomal protein S11 (BS11) [Bacillus subtilis]
 Identities = 81/92 (88%), Positives = 87/92 (94%)

Query: 2   HGNALAWSSAGALGFKGSRKSTPFAAQMAAEAAAKSAQEHGLKTVEVTVKGPGSGRESAI 61
           HGNA++WSSAGALGF+GSRKSTPFAAQMAAE AAK + EHGLKT+EVTVKGPGSGRE+AI
Sbjct: 40  HGNAISWSSAGALGFRGSRKSTPFAAQMAAETAAKGSIEHGLKTLEVTVKGPGSGREAAI 99

Query: 62  RALAAAGLEVTAIRDVTPVPHNGARPPKRRRV 93
           RAL AAGLEVTAIRDVTPVPHNG RPPKRRRV
Sbjct: 100 RALQAAGLEVTAIRDVTPVPHNGCRPPKRRRV 131
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6991> which encodes the amino acid sequence <SEQ ID 6992>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0945(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 = 92/93 (98%), Positives = 93/93 (99%)

Query: 1   MHGNALAWSSAGALGFKGSRKSTPFAAQMAAEAAAKSAQEHGLKTVEVTVKGPGSGRESA 60
           +HGNALAWSSAGALGFKGSRKSTPFAAQMAAEAAAKSAQEHGLKTVEVTVKGPGSGRESA
Sbjct: 35  VHGNALAWSSAGALGFKGSRKSTPFAAQMAAEAAAKSAQEHGLKTVEVTVKGPGSGRESA 94

Query: 61  IRALAAAGLEVTAIRDVTPVPHNGARPPKRRRV 93
           IRALAAAGLEVTAIRDVTPVPHNGARPPKRRRV
Sbjct: 95  IRALAAAGLEVTAIRDVTPVPHNGARPPKRRRV 127
```

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

Example 2263

A DNA sequence (GBSx2385) was identified in *S.agalactiae* <SEQ ID 6993> which encodes the amino acid sequence <SEQ ID 6994>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2551(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:BAB03881 GB:AP001507 DNA-directed RNA polymerase alpha subunit
          [Bacillus halodurans]
 Identities = 190/314 (60%), Positives = 249/314 (78%), Gaps = 2/314 (0%)

Query: 1   MIEFEKPIITKIDENKD--YGRFVIEPLERGYGTTLGNSLRRVLLSSLPGAAVTSIKIDG 58
           MIE EKP+I  I+ ++D  YG+FV+EPLERGYGTTLGNSLRR+LLSSLPGAAVTS++IDG
Sbjct: 1   MIEIEKPVIETIEISEDAKYGKFVVEPLERGYGTTLGNSLRRILLSSLPGAAVTSVQIDG 60

Query: 59  VLHEFDTIPGVREDVMQIILNVKGLAVKSYVEDEKIIELDVEGPAEITAGDILTDSDIEI 118
           VLHEF TI GV EDV  I+LN+K LA+K Y +++K +E+D +G   +TAGD+  DSD+++
Sbjct: 61  VLHEFSTIEGVVEDVTTIVLNLKQLALKIYSDEDKTLEIDIQGEGVVTAGDLTHDSDVDV 120

Query: 119 VNPDHYLFTIAEGHSLKATMTVAKNRGYVPAEGNKKDDAPVGTLAVDSIYTPVKKVNYQV 178
           +NPD ++ T+  G  L+  +T  + RGYVPAEGNK D+  +G + +DSIYTPV +VNYQV
Sbjct: 121 LNPDLHIATLTTGAHLRMRITAKRGRGYVPAEGNKSDELAIGVIPIDSIYTPVSRVNYQV 180

Query: 179 EPARVGSNDGFDKLTIEIMTNGTIIPEDALGLSARVLIEHLNLFTDLTEVAKATEVMKET 238
           E  RVG    +DKLT+++ T+G+I PE+A+ L A++L EHLN+F  LT+ A+  E+M E
Sbjct: 181 ENTRVGQVTNYDKLTLDVWTDGSIRPEEAVSLGAKILTEHLNIFVGLTDQAQNAEIMVEK 240

Query: 239 EKVNDEKVLDRTIEELDLSVRSYNCLKRAGINTVFDLTEKTEPEMMKVRNLGRKSLEEVK 298
           E+   EKVL+ TIEELDLSVRSYNCLKRAGINTV +LT+KTE +MMKVRNLGRKSLEEV+
Sbjct: 241 EEDQKEKVLEMTIEELDLSVRSYNCLKRAGINTVQELTQKTEEDMMKVRNLGRKSLEEVQ 300

Query: 299 IKLADLGLGLKNDK 312
            KL +LGLGL+ ++
Sbjct: 301 EKLGELGLGLRKEE 314
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6995> which encodes the amino acid sequence <SEQ ID 6996>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2551(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 = 305/312 (97%), Positives = 311/312 (98%)

Query: 1   MIEFEKPIITKIDENKDYGRFVIEPLERGYGTTLGNSLRRVLLSSLPGAAVTSIKIDGVL 60
           MIEFEKPIITKIDENKDYGRFVIEPLERGYGTTLGNSLRRVLLSSLPGAAVTSIKIDGVL
Sbjct: 1   MIEFEKPIITKIDENKDYGRFVIEPLERGYGTTLGNSLRRVLLSSLPGAAVTSIKIDGVL 60

Query: 61  HEFDTIPGVREDVMQIILNVKGLAVKSYVEDEKIIELDVEGPAEITAGDILTDSDIEIVN 120
           HEFDTIPGVREDVMQIILNVKGLAVKSYVEDEKIIEL+VEGPAE+TAGDILTDSDIE+VN
Sbjct: 61  HEFDTIPGVREDVMQIILNVKGLAVKSYVEDEKIIELEVEGPAEVTAGDILTDSDIELVN 120
```

```
Query: 121 PDHYLFTIAEGHSLKATMTVAKNRGYVPAEGNKKDDAPVGTLAVDSIYTPVKKVNYQVEP 180
           PDHYLFTIAEGHSL+ATMTVAK RGYVPAEGNKKDDAPVGTLAVDSIYTPVKKVNYQVEP
Sbjct: 121 PDHYLFTIAEGHSLRATMTVAKKRGYVPAEGNKKDDAPVGTLAVDSIYTPVKKVNYQVEP 180

Query: 181 ARVGSNDGFDKLTIEIMTNGTIIPEDALGLSARVLIEHLNLFTDLTEVAKATEVMKETEK 240
           ARVGSNDGFDKLTIEIMTNGTIIPEDALGLSARVLIEHLNLFTDLTEVAKATEVMKETEK
Sbjct: 181 ARVGSNDGFDKLTIEIMTNGTIIPEDALGLSARVLIEHLNLFTDLTEVAKATEVMKETEK 240

Query: 241 VNDEKVLDRTIEELDLSVRSYNCLKRAGINTVFDLTEKTEPEMMKVRNLGRKSLEEVKIK 300
           VNDEKVLDRTIEELDLSVRSYNCLKRAGINTVFDLTEK+EPEMMKVRNLGRKSLEEVK+K
Sbjct: 241 VNDEKVLDRTIEELDLSVRSYNCLKRAGINTVFDLTEKSEPEMMKVRNLGRKSLEEVKVK 300

Query: 301 LADLGLGLKNDK 312
           LADLGLGLKNDK
Sbjct: 301 LADLGLGLKNDK 312
```

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

Example 2264

A DNA sequence (GBSx2386) was identified in *S.agalactiae* <SEQ ID 6997> which encodes the amino acid sequence <SEQ ID 6998>. This protein is predicted to be 50S ribosomal protein L17 (rplQ). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1609(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:CAB11920 GB:Z99104 ribosomal protein L17 (BL15) [Bacillus subtilis]
 Identities = 95/128 (74%), Positives = 105/128 (81%), Gaps = 8/128 (6%)

Query: 1   MAYRKLGRTSSQRKAMLRDLTTDLLINESIVTTEARAKEIRKTVEKMITLGKRGDLHARR 60
           M+YRKLGRTS+QRKAMLRDLTTDL+INE I TTE RAKE+R  VEKMITLGKRGDLHARR
Sbjct: 1   MSYRKLGRTSAQRKAMLRDLTTDLIINERIETTETRAKELRSVVEKMITLGKRGDLHARR 60

Query: 61  QAAAYVRNEIASENYDEASDKYTSTTALQKLFDDIAPRYAERNGGYTRILKTEPRRGDAA 120
           QAAAY+RNE+A+E  ++          ALQKLF DIA RY ER GGYTRI+K  PRRGD A
Sbjct: 61  QAAAYIRNEVANEENNQ--------DALQKLFSDIATRYEERQGGYTRIMKLGPRRGDGA 112

Query: 121 PMAIIELV 128
           PMAIIELV
Sbjct: 113 PMAIIELV 120
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6999> which encodes the amino acid sequence <SEQ ID 7000>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1609(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 = 125/128 (97%), Positives = 127/128 (98%)
```

```
Query: 1    MAYRKLGRTSSQRKAMLRDLTTDLLINESIVTTEARAKEIRKTVEKMITLGKRGDLHARR 60
            MAYRKLGRTSSQRKAMLRDLTTDLLINESIVTTEARAKEIRKTVEKMITLGKRGDLHARR
Sbjct: 1    MAYRKLGRTSSQRKAMLRDLTTDLLINESIVTTEARAKEIRKTVEKMITLGKRGDLHARR 60

Query: 61   QAAAYVRNEIASENYDEASDKYTSTTALQKLFDDIAPRYAERNGGYTRILKTEPRRGDAA 120
            QAAAYVRNEIASENYDEA+DKYTSTTALQKLF +IAPRYAERNGGYTRILKTEPRRGDAA
Sbjct: 61   QAAAYVRNEIASENYDEATDKYTSTTALQKLFSEIAPRYAERNGGYTRILKTEPRRGDAA 120

Query: 121  PMAIIELV 128
             PMAIIELV
Sbjct: 121  PMAIIELV 128
```

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

Example 2265

A DNA sequence (GBSx2396) was identified in *S.agalactiae* <SEQ ID 7001> which encodes the amino acid sequence <SEQ ID 7002>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2384(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:AAA83977 GB:AF138877 mercuric reductase MerA [Bacillus sp.
            RC607]
 Identities = 29/33 (87%), Positives = 32/33 (96%)

Query: 4    VGLTEEQAKEKGYDVKTSVLPLXAVPRAIVNRE 36
            VGLTE+QAKEKGY+VKTSVLPL AVPRA+VNRE
Sbjct: 520  VGLTEQQAKEKGYEVKTSVLPLDAVPRALVNRE 552
```

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 2266

A DNA sequence (GBSx2397) was identified in *S.agalactiae* <SEQ ID 7003> which encodes the amino acid sequence <SEQ ID 7004>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3016(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 = 146/194 (75%), Positives = 175/194 (89%)

Query: 2   PQISGLEKMDYLTSTTLLELKKIPKRLTVIGSGYIGMELGQLFHHLGSEITLMQRSERLL 61
           P I GL ++DYLTST+LLELKK+PKRL VIGSGYIGMELGQLFH+LGSE+TL+QRSERLL
Sbjct: 226 PNIPGLNEVDYLTSTSLLELKKVPKRLVVIGSGYIGMELGQLFHNLGSEVTLIQRSERLL 285

Query: 62  KEYDPEISESVEKALIEQGINLVKGATFERVEQSGEIKRVYVTVNGSREVIESDQLLVAT 121
           KEYDPEISESVEK+L+EQGINLVKGAT+ER+EQ+G+IK+V+V VNG + +IE+DQLLVAT
Sbjct: 286 KEYDPEISESVEKSLVEQGINLVKGATYERIEQNGDIKKVHVEVNGKKRIIEADQLLVAT 345

Query: 122 GRKPNTDSLNLSAAGVETGKNNEILINDFGQTSNEKIYAAGDVTLGPQFVYVAAYEGGII 181
           GR PNT +LNL AAGVE G   EI+I+D+ +T+N +IYAAGDVTLGPQFVYVAAY+GG+
Sbjct: 346 GRTPNTATLNLRAAGVEIGSRGEIIIDDYSRTTNTRIYAAGDVTLGPQFVYVAAYQGGVA 405

Query: 182 TDNAIGGLNKKIDL 195
             NAIGGLNKK++L
Sbjct: 406 APNAIGGLNKKLNL 419
```

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 2267

A DNA sequence (GBSx2398) was identified in *S.agalactiae* <SEQ ID 7005> which encodes the amino acid sequence <SEQ ID 7006>. This protein is predicted to be triacylglycerol acylhydrolase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3180(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 2268

A DNA sequence (GBSx2399) was identified in *S.agalactiae* <SEQ ID 7007> which encodes the amino acid sequence <SEQ ID 7008>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0544(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.

```
>GP:AAC74453 GB:AE000234 orf, hypothetical protein [Escherichia
          coli K12]
 Identities = 45/58 (77%), Positives = 51/58 (87%)
```

```
Query: 1  MPWQNLLHAGQENLFSGLTALTAEFTVGEGKLMTHDEPCSMAPDDKHDLISGTCSHLP 58
          +PWQNLLHAG+ENLFSGLTAL+AEFT+GEG+LM HD P   APD+  DLISGTCSHLP
Sbjct: 34 LPWQNLLHAGEENLFSGLTALSAEFTIGEGELMAHDVPLGCAPDEYDDLISGTCSHLP 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 2269

A DNA sequence (GBSx2400) was identified in *S.agalactiae* <SEQ ID 7009> which encodes the amino acid sequence <SEQ ID 7010>. This protein is predicted to be transposase for insertion sequence element is5. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2058(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:BAB15497 GB:AK026530 unnamed protein product [Homo sapiens]
 Identities = 297/299 (99%), Positives = 297/299 (99%)

Query: 1   MEQILPWQNMVEVIEPFYPKAGNGRRPYPLETMLRIHCMQHWYNLSDGAMEDALYEIASM 60
           MEQILPWQNMVEVIEPFYPKAGNGRRPYPLETMLRIHCMQHWYNLSDGAMEDALYEIASM
Sbjct: 40  MEQILPWQNMVEVIEPFYPKAGNGRRPYPLETMLRIHCMQHWYNLSDGAMEDALYEIASM 99

Query: 61  RLFARLSLDSALPDRTTIMNFRHLLEQHQLARQLFKTINRWLAEAGVMMTQGTLVDATII 120
           RLFARLSLDSALPDRTTIMNFRHLLEQHQLARQLFKTINRWLAEAGVMMTQGTLVDATII
Sbjct: 100 RLFARLSLDSALPDRTTIMNFRHLLEQHQLARQLFKTINRWLAEAGVMMTQGTLVDATII 159

Query: 121 EAPSSTKNKEQQRDPEMHQTKKGNQWHFGMKAHIGVDAKSGLTHSLVTTAANEHDLNQLX 180
           EAPSSTKNKEQQRDPEMHQTKKGNQWHFGMKAHIGVDAKSGLTHSLVTTAANEHDLNQL
Sbjct: 160 EAPSSTKNKEQQRDPEMHQTKKGNQWHFGMKAHIGVDAKSGLTHSLVTTAANEHDLNQLG 219

Query: 181 NLLHGEEQFVSADAXYQGAPQREELAEVDVDWLIAERPGKVRTLKQHPRKNKTAINIEYM 240
           NLLHGEEQFVSADA YQGAPQREELAEVDVDWLIAERPGKVRTLKQHPRKNKTAINIEYM
Sbjct: 220 NLLHGEEQFVSADAGYQGAPQREELAEVDVDWLIAERPGKVRTLKQHPRKNKTAINIEYM 279

Query: 241 KASIRARVEHPFRIIKRQFGFVKARYKGLLKNDNQLAMLFTLANLFRADQMIRQWERSH 299
           KASIRARVEHPFRIIKRQFGFVKARYKGLLKNDNQLAMLFTLANLFRADQMIRQWERSH
Sbjct: 280 KASIRARVEHPFRIIKRQFGFVKARYKGLLKNDNQLAMLFTLANLFRADQMIRQWERSH 338
```

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 2270

A DNA sequence (GBSx2401) was identified in *S.agalactiae* <SEQ ID 7011> which encodes the amino acid sequence <SEQ ID 7012>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> 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:CAB51958 GB:AL109661 putative eukaryotic-type serine/threonine
          protein kinase [Streptomyces coelicolor A3(2)]
 Identities = 49/169 (28%), Positives = 90/169 (52%), Gaps = 6/169 (3%)

Query: 23  PTTIRVPDVSNKTVAQAKMTLENSGLKVGAIRNIESDSVSEGLVVKTDPAAGRSRREGAK 82
           P T+++PDV+    + +A+  LE+ GL+ G +     SD V+ G V+ T P +G + R G+
Sbjct: 469 PDTVKLPDVTGYKLDKARTLLEDEGLEPGMVTRAFSDEVARGFVISTKPGSGTTVRAGSA 528

Query: 83  VNLYIATPNKSFTLGNYKEHNYKDILKDLQGKGVKKSLIKVKRKINNDYTTGTILAQSLP 142
           V L + +       + +     +  + +L+G G+K  +      ++N++Y +G + A+  P
Sbjct: 529 VAL-VVSKGSPVDVPDVTGDDLDEARAELEGAGLK--VKTADERVNSEYDSGRV-ARQTP 584

Query: 143 EGTSFNPDGNKKLTLTVAVNDPMI-MPDVTGMTVGEVIETLTDLGLDAD 190
           E      +G+  +TLTV+     MI +PDV G +V +  + L D G + D
Sbjct: 585 EPGGRAAEGD-TVTLTVSKGPRMIEVPDVVGDSVDDAKQKLEDAGFEVD 632
 Identities = 45/161 (27%), Positives = 80/161 (48%), Gaps = 4/161 (2%)

Query: 27  RVPDVSNKTVAQAKMTLENSGLKVGAIRNIESDSVSEGLVVKTDPAAGRSRREGAKVNLY 86
           +VP + +KT AQA+  L+++GL VG +R+  SD+V  G V+ TDP  G   R+   V+L
Sbjct: 405 KVPPLLSKTEAQARDRLDDAGLDVGKVRHAYSDTVERGKVISTDPGVGDRIRKNDSVSLT 464

Query: 87  IATPNKSFTLGNYKEHNYKDILKDLQGKGVKKSLIKVKRKINNDYTTGTILAQSLPEGTS 146
           ++    + L +   +        L+ +G++  + V R  +++   G +++      GT+
Sbjct: 465 VSDGPDTVKLPDVTGYKLDKARTLLEDEGLEPGM--VTRAFSDEVARGFVISTKPGSGTT 522

Query: 147 FNPDGNKKLTLTVAVNDPMIMPDVTGMTVGEVIETLTDLGL 187
                + L V+    P+ +PDVTG  + E    L   GL
Sbjct: 523 VR--AGSAVALVVSKGSPVDVPDVTGDDLDEARAELEGAGL 561
```

There is also homology to SEQ ID 3026.

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

Example 2271

A DNA sequence (GBSx2402) was identified in *S.agalactiae* <SEQ ID 7013> which encodes the amino acid sequence <SEQ ID 7014>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> 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 9311> which encodes amino acid sequence <SEQ ID 9312> was also identified.

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

```
>GP:AAB90561 GB:AE001058 glutamine ABC transporter, ATP-binding
          protein (glnQ) [Archaeoglobus fulgidus]
 Identities = 142/219 (64%), Positives = 178/219 (80%)

Query: 1   MDIHQGEVVVIIGPSGSGKSTFLRTMNLLEVPTKGTVTFEGIDITDKKNDIFKMREKMGM 60
           M + +GEVVVIIGPSGSGKST LR +N LE PT G +  +G+DIT+ K DI K+R+++G+
Sbjct: 24  MKVEKGEVVVIIGPSGSGKSTLLRCINRLEEPTSGKILLDGVDITNSKIDINKVRQRIGI 83
```

```
Query: 61  VFQQFNLFPNMTVLENITLSPIKTKGLSNLDAQTKAYELLEKVGLKEKANTYPASLSGGQ 120
           VFQQFNLFP++T L+N+TL+PIK K +S  +A+      LLEKVGL++KA+ YPA LSGGQ
Sbjct: 84  VFQQFNLFPHLTALQNVTLAPIKIKKMSKREAEELGMRLLEKVGLEDKADYYPAQLSGGQ 143

Query: 121 QQRIAIARGLAMNPDVLLFDEPTSALDPEMVGEVLTVMQDLAKSGMTMVIVTHEMGFARE 180
           QQR+AIAR LAMNP+V+LFDE TSALDPE+V EVL VM+ LA+ GMTMV+VTHEMGFARE
Sbjct: 144 QQRVAIARALAMNPEVMLFDEVTSALDPELVKEVLDVMKQLARDGMTMVVVTHEMGFARE 203

Query: 181 VADRVIFMDAGIIVEQGAPKEVFEQTKEIRTRDFLSKVL 219
           V DRVIFMD G+IVE+G P+++F   K  RTR FLS +L
Sbjct: 204 VGDRVIFMDGGVIVEEGKPEQIFSNPKHERTRKFLSMIL 242
```

There is also homology to SEQ ID 1186.

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

Example 2272

A DNA sequence (GBSx2403) was identified in *S.agalactiae* <SEQ ID 7015> which encodes the amino acid sequence <SEQ ID 7016>. This protein is predicted to be 4-hydroxy-2-oxoglutarate aldolase (kdgA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1479(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:CAB14127 GB:Z99115 deoxyphosphogluconate aldolase [Bacillus subtilis]
 Identities = 21/62 (33%), Positives = 38/62 (60%), Gaps = 4/62 (6%)

Query: 3  QLMQGKIVAVIRGNSQEEAFQAAQACIKGGISAIEIAYTNSKASQVIEQLVTQYTNQEQV 62
           +L + K++AVIR   ++EA Q  ++ +  GI A+E+ YT   AS +IE     + N+E +
Sbjct: 9  RLKEAKLIAVIRSKDKQEACQQIESLLDKGIRAVEVTYTTPGASDIIE----SFRNREDI 64

Query: 63 VV 64
           ++
Sbjct: 65 LI 66
```

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

Example 2273

A DNA sequence (GBSx2405) was identified in *S.agalactiae* <SEQ ID 7017> which encodes the amino acid sequence <SEQ ID 7018>. This protein is predicted to be H repeat-associated protein (rfbQRS) (b1458). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0207(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

There is homology to SEQ ID 504.

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

Example 2274

A DNA sequence (GBSx2406) was identified in *S.agalactiae* <SEQ ID 7019> which encodes the amino acid sequence <SEQ ID 7020>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.74   Transmembrane    2 -  18 (   1 -  21)
   INTEGRAL    Likelihood = -3.03   Transmembrane   73 -  89 (  73 -  92)

----- Final Results -----
               bacterial membrane --- Certainty=0.3697(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

There is also homology to SEQ ID 3376.

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

Example 2275

A DNA sequence (GBSx2407) was identified in *S.agalactiae* <SEQ ID 7021> which encodes the amino acid sequence <SEQ ID 7022>. This protein is predicted to be insertion element IS1 protein InsB (insB_5). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4280(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Example 2276

A DNA sequence (GBSx2409) was identified in *S.agalactiae* <SEQ ID 7023> which encodes the amino acid sequence <SEQ ID 7024>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3937(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 2277

A DNA sequence (GBSx2410) was identified in *S.agalactiae* <SEQ ID 7025> which encodes the amino acid sequence <SEQ ID 7026>. This protein is predicted to be triosephosphate isomerase (tpi). Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane   35 -  51 (  35 -  51)

----- 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:AAC43268 GB:U07640 triosephosphate isomerase [Lactococcus
          lactis]
 Identities = 50/75 (66%), Positives = 61/75 (80%)

Query: 6  IAGNWKMNKNPEEAKAFIEAVASKLPSSELVEAGIAAPALTLSTVLEAAKGSELKIAAQN 65
          IAGNWKMNK   EA+AF+EAV + LPSS+ VE+ I APAL L+ +     +GSELK+AA+N
Sbjct: 7  IAGNWKMNKTLSEAQAFVEAVKNNLPSSDNVESVIGAPALFLAPMAYLRQGSELKLAAEN 66

Query: 66 SYFENSGAFTGENSP 80
          SYFEN+GAFTGENSP
Sbjct: 67 SYFENAGAFTGENSP 81
```

There is also homology to SEQ ID 6838:.

```
Identities = 58/77 (75%), Positives = 68/77 (87%)

Query: 6  IAGNWKMNKNPEEAKAFIEAVASKLPSSELVEAGIAAPALTLSTVLEAAKGSELKIAAQN 65
          IAGNWKMNKNP+EAKAF+EAVASKLPS++LV+  +AAPA+ L T +EAAK S LK+AAQN
Sbjct: 7  IAGNWKMNKNPQEAKAFVEAVASKLPSTDLVDVAVAAPAVDLVTTIEAAKDSVLKVAAQN 66

Query: 66 SYFENSGAFTGENSPKV 82
           YFEN+GAFTGE SPKV
Sbjct: 67 CYFENTGAFTGETSPKV 83
```

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

Example 2278

A DNA sequence (GBSx2412) was identified in *S.agalactiae* <SEQ ID 7027> which encodes the amino acid sequence <SEQ ID 7028>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.39   Transmembrane   96 - 112 (  96 - 112)

----- Final Results -----
               bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:BAA14368 GB:D90354 surface protein antigen precursor
             [Streptococcus sobrinus]
 Identities = 60/129 (46%), Positives = 76/129 (58%), Gaps = 18/129 (13%)

Query: 3    ISFDNSFLETVSDDSAFQADVYLQMKRIAAGQVENTYLHTVNGYVISSNTVVTHTPQPEE 62
             ++F   FL +VS DSAFQA+VYLQMKRIA G   NTY++TVNG   SSNTV T TP+P++
Sbjct: 1442 VTFKEDFLRSVSVDSAFQAEVYLQMKRIAVGTFANTYVNTVNGITYSSNTVRTSTPEPKQ 1501

Query: 63   PSPNQP--------TPPQPPIETIEPPVPASILPNTGEQES----LLGLIG--AGILLGT 108
             PSP  P          P Q       PP  A  LP TG+  +    LLGL+   AG  L
Sbjct: 1502 PSPVDPKTTTTVVFQPRQGKAYQPAPPAGAQ-LPATGDSSNAYLPLLGLVSLTAGFSL-- 1558

Query: 109  AYGLKKKEE 117
               GL++K++
Sbjct: 1559 -LGLRRKQD 1566
```

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 2279

A DNA sequence (GBSx2413) was identified in *S.agalactiae* <SEQ ID 7029> which encodes the amino acid sequence <SEQ ID 7030>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3691(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 9359> which encodes amino acid sequence <SEQ ID 9360> was also identified.

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

```
>GP:CAB15793 GB:Z99123 phosphotransacetylase [Bacillus subtilis]
 Identities = 131/221 (59%), Positives = 169/221 (76%), Gaps = 2/221 (0%)

Query: 6   LVDPVILGKADEVHDSLARLGFVDQDYSIIDPEQYEKFEEMKEAFVEIRKGKATMEDADR 65
            +++P+++G  +E+      L        I DP  YE  E++ +AFVE RKGKAT E A +
Sbjct: 41  VLNPIVIGNENEIQAKAKELNLTLGGVKIYDPHTYEGMEDLVQAFVERRKGKATEEQARK 100

Query: 66  LLKDVNYFGVMLVKLGLADGMVSGAIHSTADTVRPALQIIKTKPGISRTSGVFLMNRENT 125
             L D NYFG MLV  GLADG+VSGA HSTADTVRPALQIIKTK G+ +TSGVF+M R
Sbjct: 101 ALLDENYFGTMLVYKGLADGLVSGAAHSTADTVRPALQIIKTKEGVKKTSGVFIMARG-- 158

Query: 126 QERYIFADCAINIDPNAQELAEIAVNTADTAKIFDIDPKIAMLSFSTKGSAKAPQAEKVQ 185
            +E+Y+FADCAINI P++Q+LAEIA+ +A+TAK+FDI+P++AMLSFSTKGSAK+ + EKV
Sbjct: 159 EEQYVFADCAINIAPDSQDLAEIAIESANTAKMFDIEPRVAMLSFSTKGSAKSDETEKVA 218

Query: 186 EAAKIAKDLSPELAVDGELQFDAAFVPETAEIKAPNSDVAG 226
            +A KIAK+ +PEL +DGE QFDAAFVP  AE KAP+S++ G
Sbjct: 219 DAVKIAKEKAPELTLDGEFQFDAAFVPSVAEKKAPDSEIKG 259
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7031> which encodes the amino acid sequence <SEQ ID 7032>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
```

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3182(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 = 181/227 (79%), Positives = 211/227 (92%)

Query: 1   MKFEGLVDPVILGKADEVHDSLARLGFVDQDYSIIDPEQYEKFEEMKEAFVEIRKGKATM 60
           +KFEGL++P+ILG+++EV + L +LGF DQDY+II+P +Y  F++MKEAFVE+RKGKAT+
Sbjct: 38  LKFEGLLEPIILGQSEEVRNLLTKLGFADQDYTIINPNEYADFDKMKEAFVEVRKGKATL 97

Query: 61  EDADRLLKDVNYFGVMLVKLGLADGMVSGAIHSTADTVRPALQIIKTKPGISRTSGVFLM 120
           EDAD++L+DVNYFGVMLVK+GLADGMVSGAIHSTADTVRPALQIIKTKPGISRTSGVFLM
Sbjct: 98  EDADKMLRDVNYFGVMLVKMGLADGMVSGAIHSTADTVRPALQIIKTKPGISRTSGVFLM 157

Query: 121 NRENTQERYIFADCAINIDPNAQELAEIAVNTADTAKIFDIDPKIAMLSFSTKGSAKAPQ 180
           NRENT ERY+FADCAINIDP AQELAEIAVNTA+TAKIFDIDPKIAMLSFSTKGS KAPQ
Sbjct: 158 NRENTSERYVFADCAINIDPTAQELAEIAVNTAETAKIFDIDPKIAMLSFSTKGSGKAPQ 217

Query: 181 AEKVQEAAKIAKDLSPELAVDGELQFDAAFVPETAEIKAPNSDVAGK 227
            +KV+EA +IA  L+P+LA+DGELQFDAAFVPETA IKAP+S VAG+
Sbjct: 218 VDKVREATEIATGLNPDLALDGELQFDAAFVPETAAIKAPDSAVAGQ 264
```

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

Example 2280

A DNA sequence (GBSx2414) was identified in *S.agalactiae* <SEQ ID 7033> which encodes the amino acid sequence <SEQ ID 7034>. This protein is predicted to be lipopolysaccharide biosynthesis protein-related 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=0.4076(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:AAG19110 GB:AE005009 Vng0600c [Halobacterium sp. NRC-1]
 Identities = 57/176 (32%), Positives = 86/176 (48%), Gaps = 20/176 (11%)

Query: 1   MKVLLYLEAEEYLKKSGIGRAIKHQEKALQIAGIDYTTNPT------------------ 41
           M+ L YLEA E L+  G+  A   Q  AL+   ++    P
Sbjct: 2   MRALNYLEAAEALR-GGMVTATNQQRAALETTDVEVVETPWRAGDPVRSIGSLAAGGSCF 60

Query: 42  DDFDLVHMNTYGIRSWLLMSKAKKTGKKVIMHGHSTEEDFRNSFIGSNLVSPLFKWYLCR 101
             FD+ H N  G  S  +   A++T   +++H H T EDF  SF GS+ ++P  + YL
Sbjct: 61  TAFDVAHCNLVGPGSVAVARHARRTDTPLVLHAHLTREDFAQSFRGSSTIAPALEPYLRW 120

Query: 102 FYQKADAIITPTDYSKQLIKAYGIKKPIFVLSNGIDLSRYQXSEKKESAFRHYFHL 157
           FY +AD ++ P++Y+K +++AY +  PI  LSNG+DL   Q  E   +  R  F L
Sbjct: 121 FYSQADLVLCPSEYTKDVLRAYPVDAPIRQLSNGVDLESMQGYESFRADTRARFDL 176
```

There is also homology to SEQ ID 1220.

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

Example 2281

A DNA sequence (GBSx2415) was identified in *S.agalactiae* <SEQ ID 7035> which encodes the amino acid sequence <SEQ ID 7036>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2625(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:AAC35010 GB:AF055987 intracellular a-amylase [Streptococcus mutans]
 Identities = 27/46 (58%), Positives = 33/46 (71%)

Query: 1   MEVGEIYAGKTFVDYLGNCEQEVVIGDDGWGDFLVESASISAWVPK 46
           M +GE     K FVDYL NC +EV++ D GWGDF V+ AS+SAWV K
Sbjct: 438 MNMGEFNRNKVFVDYLNNCTEEVILDDQGWGDFPVQEASLSAWVNK 483
```

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 2282

A DNA sequence (GBSx2416) was identified in *S.agalactiae* <SEQ ID 7037> which encodes the amino acid sequence <SEQ ID 7038>. This protein is predicted to be RopA. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2082(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

There is also homology to SEQ ID 6908:

```
Identities = 30/35 (85%), Positives = 33/35 (93%)

Query: 1   MEADQVRGLLSADMLKHDIAMKKAVDVITSSATVK 35
           M ADQVR LLSADMLKHDIAMKKAV+VITS+A+VK
Sbjct: 422 MPADQVRSLLSADMLKHDIAMKKAVEVITSTASVK 456
```

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

Example 2283

A DNA sequence (GBSx2417) was identified in *S.agalactiae* <SEQ ID 7039> which encodes the amino acid sequence <SEQ ID 7040>. This protein is predicted to be DNA-directed RNA polymerase, subunit delta. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2407(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:CAB15744 GB:Z99123 RNA polymerase (delta subunit) [Bacillus subtilis]
 Identities = 62/186 (33%), Positives = 102/186 (54%), Gaps = 15/186 (8%)

Query: 1   MELEVFAGQEKSELSMIEVARAILEQRGRDNEMYFSDLVNDIQTYLGKSDSAIRESLPFF 60
           M ++ ++ +E  E++++E+A  + E+  +   + F +L+N+I + LG     + + +  F
Sbjct: 1   MGIKQYSQEELKEMALVEIAHELFEEHKKP--VPFQELLNEIASLLGVKKEELGDRIAQF 58

Query: 61  YSDLNTDGSFIPLGENKWGLRSWYAIDEIDEEIITLEEDEDGAPKRKKKRVNAFMDGDED 120
           Y+DLN DG F+ L +  WGLRSWY  D++DEE           K KKK+    ++ D D
Sbjct: 59  YTDLNIDGRFLALSDQTWGLRSWYPYDQLDEE-------TQPTVKAKKKKAKKAVEEDLD 111

Query: 121 AIDYNDDDPEDEDFTEETPSLEYDEENPDDEKSEVESYDSEINEIIPDEDLDEDVEINEE 180
             ++ + D +D D  E    L+ + ++ D+E  + +  D EI E I DED DED
Sbjct: 112 LDEFEEIDEDDLDLDEVEEELDLEADDFDEEDLDEDDDDLEIEEDIIDED-DEDY----- 165

Query: 181 DDEEEE 186
           DDEEEE
Sbjct: 166 DDEEEE 171
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7041> which encodes the amino acid sequence <SEQ ID 7042>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2263(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 = 162/191 (84%), Positives = 181/191 (93%), Gaps = 1/191 (0%)

Query: 1   MELEVFAGQEKSELSMIEVARAILEQRGRDNEMYFSDLVNDIQTYLGKSDSAIRESLPFF 60
           ++L+VFAGQEKSELSMIEVARAILE+RGRDNEMYFSDLVN+IQ YLGKSD+ IR +LPFF
Sbjct: 12  LKLDVFAGQEKSELSMIEVARAILEERGRDNEMYFSDLVNEIQNYLGKSDAGIRHALPFF 71

Query: 61  YSDLNTDGSFIPLGENKWGLRSWYAIDEIDEEIITLEEDEDGAPKRKKKRVNAFMDGDED 120
           Y+DLNTDGSFIPLGENKWGLRSWYAIDEIDEEIITLEEDEDGA KRKKKRVNAFMDGDED
Sbjct: 72  YTDLNTDGSFIPLGENKWGLRSWYAIDEIDEEIITLEEDEDGAQKRKKKRVNAFMDGDED 131

Query: 121 AIDYNDDDPEDEDFTEETPSLEYDEENPDDEKSEVESYDSEINEIIPDEDLDEDVEINEE 180
           AIDY DDDPEDEDFTEE+  +EYDEE+PDDEKSEVESYDSE+NEIIP++D  E+V+INEE
Sbjct: 132 AIDYRDDDPEDEDFTEESAEVEYDEEDPDDEKSEVESYDSELNEIIPEDDF-EEVDINEE 190

Query: 181 DDEEEEEEEEV 191
           D+E+EE+EE V
Sbjct: 191 DEEDEEDEEPV 201
```

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

Example 2284

A DNA sequence (GBSx2418) was identified in *S.agalactiae* <SEQ ID 7043> which encodes the amino acid sequence <SEQ ID 7044>. This protein is predicted to be CTP synthetase (pyrG). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.11   Transmembrane    5 -  21 (   5 -  21)

----- 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:CAA09021 GB:AJ010153 CTP synthetase [Lactococcus lactis subsp.
          cremoris] (ver 2)
 Identities = 421/533 (78%), Positives = 481/533 (89%) .

Query: 2   TKYIFVTGGVVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPYINIDPGTMSPYQHGEVYV 61
           TKYIFVTGG  SS+GKGIVAASLGRLLKNRGLKVT+QKFDPY+NIDPGTMSPYQHGEV+V
Sbjct: 3   TKYIFVTGGGTSSMGKGIVAASLGRLLKNRGLKVTVQKFDPYLNIDPGTMSPYQHGEVFV 62

Query: 62  TDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVLKKERRGEYLGATVQVIPHVTDA 121
           TDDGAETDLDLGHYERFIDINLNKYSNVT+GK+YSE+L+KER+GEYLGATVQ++PHVT+
Sbjct: 63  TDDGAETDLDLGHYERFIDINLNKYSNVTSGKVYSEILRKERKGEYLGATVQMVPHVTNM 122

Query: 122 LKEKIKRAATTTDSDVIITEVGGTVGDIESLPFLEALRQMKADVGSDNVMYIHTTLLPYL 181
           LKEKIKRAATTTD+D+IITEVGGTVGD+ESLPF+EALRQMKA+VG+DNVMYIHT  + +L
Sbjct: 123 LKEKIKRAATTTDADIIITEVGGTVGDMESLPFIEALRQMKAEVGADNVMYIHTVPILHL 182

Query: 182 KAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTEQPAGQSIKNKLAQFCDVAPEAVIESLD 241
           +AAGE+KTK  Q++ K LR  GIQ NMLV+R+E P     +++K+A FCDVAPEAVI+SLD
Sbjct: 183 RAAGELKTKIAQNATKTLREYGIQANMLVLRSEVPITTEMRDKIAMFCDVAPEAVIQSLD 242

Query: 242 VDHIYQIPLNMQAQNMDQIVCDHLKLETPAADMTEWSAMVDKVMNLEKKVKIALVGKYVE 301
           V+H+YQIPLN+QAQNMDQIVCDHLKL+ P ADM EWSAMVD VMNL+KKVKIALVGKYVE
Sbjct: 243 VEHLYQIPLNLQAQNMDQIVCDHLKLDAPKADMAEWSAMVDHVMNLKKKVKIALVGKYVE 302

Query: 302 LPDAYLSVVEALKHSGYVNDVAIDLKWVNAAEVTEDNIKELVGDADGIIVPGGFGQRGSE 361
           LPDAY+SV EALKH+GY +D  +D+ WVNA +VT++N+ ELVGDA GIIVPGGFGQRG+E
Sbjct: 303 LPDAYISVTEALKHAGYASDAEVDINWVNANDVTDENVAELVGDAAGIIVPGGFGQRGTE 362

Query: 362 GKIEAIRYARENDVPMLGVCLGMQLTCVEFARNVLNLHGANSAELDPKTPFPIIDIMRDQ 421
           GKI AI+YARENDVPMLG+CLGMQLT VEFARNVL L GA+S ELDP+T +P+IDIMRDQ
Sbjct: 363 GKIAAIKYARENDVPMLGICLGMQLTAVEFARNVLGLEGAHSFELDPETKYPVIDIMRDQ 422

Query: 422 IDIEDMGGTLRLGLYPCKLKSGSRAAAAYNNQEVVQRRHRHRYEFNTKFREQFEAAGFVF 481
           +D+EDMGGTLRLGLYP KLK+GSRA AAYN+ EVVQRRHRHRYEFN K+RE FE AGFVF
Sbjct: 423 VDVEDMGGTLRLGLYPAKLKNGSRAKAAYNDAEVVQRRHRHRYEFNNKYREDFEKAGFVF 482

Query: 482 SGVSPDNRLMEVVELPEKKFFVAAQYHPELQSRPNHAEELYTAFVTAAVENMK 534
           SGVSPDNRL+E+VEL  KKFFVA QYHPELQSRPN  EELYT F+  AVEN K
Sbjct: 483 SGVSPDNRLVEIVELSGKKFFVACQYHPELQSRPNRPEELYTEFIRVAVENSK 535
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7045> which encodes the amino acid sequence <SEQ ID 7046>. Analysis of this protein sequence reveals the following:

```
     Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.11   Transmembrane    5 -  21 (   5 -  21)

----- 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 databases:

```
>GP:CAA09021 GB:AJ010153 CTP synthetase [Lactococcus lactis subsp.
        cremoris] (ver 2)
 Identities = 423/532 (79%), Positives = 483/532 (90%)

Query: 2   TKYIFVTGGVVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPYINIDPGTMSPYQHGEVYV 61
           TKYIFVTGG  SS+GKGIVAASLGRLLKNRGLKVT+QKFDPY+NIDPGTMSPYQHGEV+V
Sbjct: 3   TKYIFVTGGGTSSMGKGIVAASLGRLLKNRGLKVTVQKFDPYLNIDPGTMSPYQHGEVFV 62

Query: 62  TDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVLRKERKGEYLGATVQVIPHITDA 121
           TDDGAETDLDLGHYERFIDINLNKYSNVT+GK+YSE+LRKERKGEYLGATVQ++PH+T+
Sbjct: 63  TDDGAETDLDLGHYERFIDINLNKYSNVTSGKVYSEILRKERKGEYLGATVQMVPHVTNM 122

Query: 122 LKEKIKRAASTTDSDVIITEVGGTVGDIESLPFLEALRQMKADVGSENVMYIHTTLLPYL 181
           LKEKIKRAA+TTD+D+IITEVGGTVGD+ESLPF+EALRQMKA+VG++NVMYIHT  + +L
Sbjct: 123 LKEKIKRAATTTDADIIITEVGGTVGDMESLPFIEALRQMKAEVGADNVMYIHTVPILHL 182

Query: 182 KAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTEEPVEQGIKNKLAQFCDVNSEAVIESRD 241
           +AAGE+KTK  Q++ K LR  GIQ NMLV+R+E P+    +++K+A FCDV  EAVI+S D
Sbjct: 183 RAAGELKTKIAQNATKTLREYGIQANMLVLRSEVPITTEMRDKIAMFCDVAPEAVIQSLD 242

Query: 242 VEHLYQIPLNLQAQSMDQIVCDHLKLNAPQADMTEWSAMVDKVMNLRKTTKIALVGKYVE 301
           VEHLYQIPLNLQAQ+MDQIVCDHLKL+AP+ADM EWSAMVD VMNL+K  KIALVGKYVE
Sbjct: 243 VEHLYQIPLNLQAQNMDQIVCDHLKLDAPKADMAEWSAMVDHVMNLKKKVKIALVGKYVE 302

Query: 302 LPDAYLSVVEALKHSGYANDTAIDLKWVNANDVTVDNAADLLGDADGIIVPGGFGQRGTE 361
           LPDAY+SV EALKH+GYA+D  +D+ WVNANDVT +N A+L+GDA GIIVPGGFGQRGTE
Sbjct: 303 LPDAYISVTEALKHAGYASDAEVDINWVNANDVTDENVAELVGDAAGIIVPGGFGQRGTE 362

Query: 362 GKIQAIRYARENDVPMLGICLGMQLTCVEFARHVLNMEGANSFELEPSTKYPIIDIMRDQ 421
           GKI AI+YARENDVPMLGICLGMQLT VEFAR+VL +EGA+SFEL+P TKYP+IDIMRDQ
Sbjct: 363 GKIAAIKYARENDVPMLGICLGMQLTAVEFARNVLGLEGAHSFELDPETKYPVIDIMRDQ 422

Query: 422 IDIEDMGGTLRLGLYPCKLKPGSKAAMAYNNQEVVQRRHRHRYEFNNKFRPEFEAAGFVF 481
           +D+EDMGGTLRLGLYP KLK GS+A  AYN+ EVVQRRHRHRYEFNNK+R +FE AGFVF
Sbjct: 423 VDVEDMGGTLRLGLYPAKLKNGSRAKAAYNDAEVVQRRHRHRYEFNNKYREDFEKAGFVF 482

Query: 482 SGVSPDNRLVEIVELKEKKFFVAAQYHPELQSRPNRPEELYTAFVTAAIKNS 533
           SGVSPDNRLVEIVEL  KKFFVA QYHPELQSRPNRPEELYT F+  A++NS
Sbjct: 483 SGVSPDNRLVEIVELSGKKFFVACQYHPELQSRPNRPEELYTEFIRVAVENS 534
```

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

```
 Identities = 477/532 (89%), Positives = 503/532 (93%)

Query: 1   MTKYIFVTGGVVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPYINIDPGTMSPYQHGEVY 60
           MTKYIFVTGGVVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPYINIDPGTMSPYQHGEVY
Sbjct: 1   MTKYIFVTGGVVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPYINIDPGTMSPYQHGEVY 60

Query: 61  VTDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVLKKERRGEYLGATVQVIPHVTD 120
           VTDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVL+KER+GEYLGATVQVIPH+TD
Sbjct: 61  VTDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVLRKERKGEYLGATVQVIPHITD 120

Query: 121 ALKEKIKRAATTTDSDVIITEVGGTVGDIESLPFLEALRQMKADVGSDNVMYIHTTLLPY 180
           ALKEKIKRAA+TTDSDVIITEVGGTVGDIESLPFLEALRQMKADVGS+NVMYIHTTLLPY
Sbjct: 121 ALKEKIKRAASTTDSDVIITEVGGTVGDIESLPFLEALRQMKADVGSENVMYIHTTLLPY 180

Query: 181 LKAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTEQPAGQSIKNKLAQFCDVAPEAVIESL 240
           LKAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTE+P  Q IKNKLAQFCDV  EAVIES
Sbjct: 181 LKAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTEEPVEQGIKNKLAQFCDVNSEAVIESR 240

Query: 241 DVDHIYQIPLNMQAQNMDQIVCDHLKLETPAADMTEWSAMVDKVMNLEKKVKIALVGKYV 300
           DV+H+YQIPLN+QAQ+MDQIVCDHLKL  P ADMTEWSAMVDKVMNL K  KIALVGKYV
```

```
Sbjct: 241 DVEHLYQIPLNLQAQSMDQIVCDHLKLNAPQADMTEWSAMVDKVMNLRKTTKIALVGKYV 300

Query: 301 ELPDAYLSVVEALKHSGYVNDVAIDLKWVNAAEVTEDNIKELVGDADGIIVPGGFGQRGS 360
           ELPDAYLSVVEALKHSGY ND AIDLKWVNA +VT DN  +L+GDADGIIVPGGFGQRG+
Sbjct: 301 ELPDAYLSVVEALKHSGYANDTAIDLKWVNANDVTVDNAADLLGDADGIIVPGGFGQRGT 360

Query: 361 EGKIEAIRYARENDVPMLGVCLGMQLTCVEFARNVLNLHGANSAELDPKTPFPIIDIMRD 420
           EGKI+AIRYARENDVPMLG+CLGMQLTCVEFAR+VLN+ GANS EL+P T +PIIDIMRD
Sbjct: 361 EGKIQAIRYARENDVPMLGICLGMQLTCVEFARHVLNMEGANSFELEPSTKYPIIDIMRD 420

Query: 421 QIDIEDMGGTLRLGLYPCKLKSGSRAAAAYNNQEVVQRRHRHRYEFNTKFREQFEAAGFV 480
           QIDIEDMGGTLRLGLYPCKLK GS+AA AYNNQEVVQRRHRHRYEFN KFR +FEAAGFV
Sbjct: 421 QIDIEDMGGTLRLGLYPCKLKPGSKAAMAYNNQEVVQRRHRHRYEFNNKFRPEFEAAGFV 480

Query: 481 FSGVSPDNRLMEVVELPEKKFFVAAQYHPELQSRPNHAEELYTAFVTAAVEN 532
           FSGVSPDNRL+E+VEL EKKFFVAAQYHPELQSRPN  EELYTAFVTAA++N
Sbjct: 481 FSGVSPDNRLVEIVELKEKKFFVAAQYHPELQSRPNRPEELYTAFVTAAIKN 532
```

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

Example 2285

A DNA sequence (GBSx2419) was identified in *S.agalactiae* <SEQ ID 7047> which encodes the amino acid sequence <SEQ ID 7048>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.92   Transmembrane   13 -  29 (   3 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.4970(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 9285> which encodes amino acid sequence <SEQ ID 9286> was also identified.

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

```
>GP:CAB14296 GB:Z99116 yqkD [Bacillus subtilis]
 Identities = 79/289 (27%), Positives = 139/289 (47%), Gaps = 8/289 (2%)

Query: 1   MKKIRLSKFIKMIVVILFLISVAASFYFFHVAQVRDDKSFISNGQRKPGNSLYAYDKSFD 60
           MKKI L+  I  +V  +  I +  S     + +  D+   I   +  G+ ++   +SF+
Sbjct: 1   MKKILLA--IGALVTAVIAIGIVFSHMILFIKKKTDED--IIKRETDNGHDVF---ESFE 53

Query: 61  KLLKQKIEMTNQNIKQVAWYVPAVKKTHKTAVVVHGFANSKENMKAYGWLFHKLGYNVLM 120
           ++ K    + +     +  Y  A   T  T ++ HG   +  N   Y  LF  LG+NVL+
Sbjct: 54  QMEKTAFVIPSAYGYDIKGYHVAPHDTPNTIIICHGVTMNVLNSLKYMHLFLDLGWNVLI 113

Query: 121 PDNIAHGESHGQLIGYGWNDRENIIKWTEMIVDK-NPSSQITLFGVSMGGATVMMASGEK 179
            D+  HG+S G+   YG+ +++++ K   ++ +K N    I + G SMG  T ++ +G
Sbjct: 114 YDHRRHGQSGGKTTSYGFYEKDDLNKVVSLLKNKTNHRGLIGIHGESMGAVTALLYAGAH 173

Query: 180 LPSQVVNIIEDCGYSSVWDELKFQAKEMYGLPAFPLLYEVSTISKIRAGFSYGQASSVEQ 239
                I DC ++   ++L ++ +  Y LP++PLL       K+R G+   + S +
Sbjct: 174 CSDGADFYIADCPFACFDEQLAYRLRAEYRLPSWPLLPIADFFLKLRGGYRAREVSPLAV 233

Query: 240 LKKNNLPALFIHGDKDNFVPTSMVYDNYKATAGKKELYIVKGAKHAKSF 288
           + K   P LFIH   D+++P S     Y+   G K LYI +  +HA S+
Sbjct: 234 IDKIEKPVLFIHSKDDDYIPVSSTERLYEKKRGPKALYIAENGEHAMSY 282
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7049> which encodes the amino acid sequence <SEQ ID 7050>. Analysis of this protein sequence reveals the following:

```
      Possible site: 24
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -7.48   Transmembrane   10 -  26 (   3 -  32)

   ----- 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:

```
>GP:CAB14296 GB:Z99116 yqkD [Bacillus subtilis]
 Identities = 88/295 (29%), Positives = 145/295 (48%), Gaps = 4/295 (1%)

Query: 10  LGILFLLITLISVGASFYFFHVAQIREEKSFINNKKRSTNNPLYPAEQSFDALPYEKRQL 69
           L I  L+  +I++G    F H+    ++K+  +  KR T+N  +   +SF+ +      +
Sbjct: 6   LAIGALVTAVIAIG--IVFSHMILFIKKKTDEDIIKRETDNG-HDVFESFEQMEKTAFVI 62

Query: 70  TNRGLKQVGWYLPAAQKTKKTAIVVHGFTNDKEDMKPYAMLFHDLGYNVLMPDNEAHGES 129
            +     +  Y  A   T  T I+ HG T +  +   Y  LF DLG+NVL+ D+  HG+S
Sbjct: 63  PSAYGYDIKGYHVAPHDTPNTIIICHGVTMNVLNSLKYMHLFLDLGWNVLIYDHRRHGQS 122

Query: 130 EGNLIGYGWNDRLNVMAWTDQLI-KENPESQITLFGLSMGAATVMMASGERLPAQVTSLI 188
            G    YG+ ++ ++      L  K N    I + G SMGA T ++ +G          I
Sbjct: 123 GGKTTSYGFYEKDDLNKVVSLLKNKTNHRGLIGIHGESMGAVTALLYAGAHCSDGADFYI 182

Query: 189 EDCGYASVWDELKFQAKAMYNLPAFPLLYEVSALSKIRAGFSYGEASSVKQLAKNKRPTL 248
            DC +A   ++L ++ +A Y LP++PLL       K+R G+   E S +  + K ++P L
Sbjct: 183 ADCPFACFDEQLAYRLRAEYRLPSWPLLPIADFFLKLRGGYRAREVSPLAVIDKIEKPVL 242

Query: 249 FIHGDKDDFVPTKMVYDNYKATKGPKEILIVKGAKHAKSFETNPEQYQKKIAAFL 303
           FIH   DD++P       Y+  +GPK + I +  +HA S+  N   Y+K +  FL
Sbjct: 243 FIHSKDDDYIPVSSTERLYEKKRGPKALYIAENGEHAMSYTKNRHTYRKTVQEFL 297
```

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

```
 Identities = 203/294 (69%), Positives = 246/294 (83%)

Query: 1   MKKIRLSKFIKMIVVILFLISVAASFYFFHVAQVRDDKSFISNGQRKPGNSLYAYDKSFD 60
           MK IR++K++ ++ +++ LISV ASFYFFHVAQ+R++KSFI+N +R   N LY  ++SFD
Sbjct: 1   MKTIRIAKYLGILFLLITLISVGASFYFFHVAQIREEKSFINNKKRSTNNPLYPAEQSFD 60

Query: 61  KLLKQKIEMTNQNIKQVAWYVPAVKKTHKTAVVVHGFANSKENMKAYGWLFHKLGYNVLM 120
            L  +K ++TN+ +KQV WY+PA +KT KTA+VVHGF N KE+MK Y  LFH LGYNVLM
Sbjct: 61  ALPYEKRQLTNRGLKQVGWYLPAAQKTKKTAIVVHGFTNDKEDMKPYAMLFHDLGYNVLM 120

Query: 121 PDNIAHGESHGQLIGYGWNDRENIIKWTEMIVDKNPSSQITLFGVSMGGATVMMASGEKL 180
           PDN AHGES G LIGYGWNDR N++ WT+ ++ +NP SQITLFG+SMG ATVMMASGE+L
Sbjct: 121 PDNEAHGESEGNLIGYGWNDRLNVMAWTDQLIKENPESQITLFGLSMGAATVMMASGERL 180

Query: 181 PSQVVNIIEDCGYSSVWDELKFQAKEMYGLPAFPLLYEVSTISKIRAGFSYGQASSVEQL 240
           P+QV ++IEDCGY+SVWDELKFQAK MY LPAFPLLYEVS +SKIRAGFSYG+ASSV+QL
Sbjct: 181 PAQVTSLIEDCGYASVWDELKFQAKAMYNLPAFPLLYEVSALSKIRAGFSYGEASSVKQL 240

Query: 241 KKNNLPALFIHGDKDNFVPTSMVYDNYKATAGKKELYIVKGAKHAKSFETEPEK 294
            KN  P LFIHGDKD+FVPT MVYDNYKAT G KE+ IVKGAKHAKSFET PE+
Sbjct: 241 AKNKRPTLFIHGDKDDFVPTKMVYDNYKATKGPKEILIVKGAKHAKSFETNPEQ 294
```

SEQ ID 9286 (GBS662) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 136 (lane 8-10; MW 63kDa) and in Figure 187 (lane 4; MW 63kDa).

GBS662-GST was purified as shown in Figure 237, lane 7.

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

Example 2286

A DNA sequence (GBSx2420) was identified in *S.agalactiae* <SEQ ID 7051> which encodes the amino acid sequence <SEQ ID 7052>. This protein is predicted to be aspartate--ammonia ligase (asnA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2898(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 9309> which encodes amino acid sequence <SEQ ID 9310> was also identified.

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

```
>GP:AAC22222 GB:U32738 aspartate--ammonia ligase (asnA) [Haemophilus influenzae Rd]
 Identities = 246/300 (82%), Positives = 268/300 (89%)

Query: 1   MIDKLEIVEVQGPILSQVGDGMQDNLSGIEHPVSVKVLNIPEAEFEVVHSLAKWKRHTLA 60
           +I++L I+EVQGPILSQVG+GMQDNLSGIE  V V V  IP A FEVVHSLAKWKRHTLA
Sbjct: 23  LIEQLGIIEVQGPILSQVGNGMQDNLSGIEKAVQVNVKCIPNAVFEVVHSLAKWKRHTLA 82

Query: 61  RFGFNEGEGLFVHMKALRPDEDSLDPTHSVYVDQWDWEKVIPDGRRNLDYLKETVEKIYK 120
           RF F E EGLFVHMKALRPDEDSLDPTHSVYVDQWDWEKVIP+GRRN  YLKETV  IY+
Sbjct: 83  RFNFKEDEGLFVHMKALRPDEDSLDPTHSVYVDQWDWEKVIPEGRRNFAYLKETVNSIYR 142

Query: 121 AIRLTELAVEARFDIESILPKRITFIHTEELVEKYPDLSPKERENAIAKEYGAVFLIGIG 180
           AIRLTELAVEARFDI SILPK+ITF+H+E+LV++YPDLS KERENAI KEYGAVFLIGIG
Sbjct: 143 AIRLTELAVEARFDIPSILPKQITFVHSEDLVKRYPDLSSKERENAICKEYGAVFLIGIG 202

Query: 181 GELADGKPHDGRAPDYDDWTTPSENGFKGLNGDILVWNEQLGTAFELSSMGIRVDEDALK 240
           G+L+DGKPHDGRAPDYDDWTT SENG+KGLNGDILVWN+QLG AFELSSMGIRVDE AL+
Sbjct: 203 GKLSDGKPHDGRAPDYDDWTTESENGYKGLNGDILVWNDQLGKAFELSSMGIRVDESALR 262

Query: 241 RQVVLTGDEDRLEFEWHKTLLRGFFPLTIGGGIGQSRLAMFLLRKKHIGEVQSSVWPKEV 300
            QV LTGDED L+ +WH+ LL G  PLTIGGGIGQSRLAM LLRK HIGEVQSSVWPKE+
Sbjct: 263 LQVGLTGDEDHLKMDWHQDLLNGKLPLTIGGGIGQSRLAMLLLRKKHIGEVQSSVWPKEM 322
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7053> which encodes the amino acid sequence <SEQ ID 7054>. Analysis of this protein sequence reveals the following:

```
     Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16   Transmembrane  189 - 205 ( 189 - 205)

----- 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:AAC22222 GB:U32738 aspartate--ammonia ligase (asnA) [Haemophilus influenzae Rd]
 Identities = 255/330 (77%), Positives = 289/330 (87%)

Query: 1   MKKSFIHQQEEISFVKNTFTQYLIAKLDVVEVQGPILSRVGDGMQDNLSGTENPVSVNVL 60
           MKK+FI QQ+EISFVKNTFTQ LI +L ++EVQGPILS+VG+GMQDNLSG E  V VNV
```

```
Sbjct: 1   MKKTFILQQQEISFVKNTFTQNLIEQLGIIEVQGPILSQVGNGMQDNLSGIEKAVQVNVK 60

Query: 61  KIPNATFEVVHSLAKWKRHTLARFGFNEGEGLVVNMKALRPDEDSLDQTHSVYVDQWDWE 120
            IPNA FEVVHSLAKWKRHTLARF F E EGL V+MKALRPDEDSLD THSVYVDQWDWE
Sbjct: 61  CIPNAVFEVVHSLAKWKRHTLARFNFKEDEGLFVHMKALRPDEDSLDPTHSVYVDQWDWE 120

Query: 121 KVIPDGKRNLAYLKETVETIYKVIRLTELAVEARYDIEAVLPKKITFIHTEELVAKYPDL 180
            KVIP+G+RN AYLKETV +IY+ IRLTELAVEAR+DI ++LPK+ITF+H+E+LV +YPDL
Sbjct: 121 KVIPEGRRNFAYLKETVNSIYRAIRLTELAVEARFDIPSILPKQITFVHSEDLVKRYPDL 180

Query: 181 TPKERENAITKEFGAVFLIGIGGVLPDGKPHDGRAPDYDDWTTETENGYHGLNGDILVWN 240
            + KERENAI KE+GAVFLIGIGG L DGKPHDGRAPDYDDWTTE+ENGY GLNGDILVWN
Sbjct: 181 SSKERENAICKEYGAVFLIGIGGKLSDGKPHDGRAPDYDDWTTESENGYKGLNGDILVWN 240

Query: 241 DQLGSAFELSSMGIRVDEEALKRQVEMTGDQDRLGFDWHKSLLNGLFPLTIGGGIGQSRM 300
            DQLG AFELSSMGIRVDE AL+ QV +TGD+D L  DWH+ LLNG  PLTIGGGIGQSR+
Sbjct: 241 DQLGKAFELSSMGIRVDESALRLQVGLTGDEDHLKMDWHQDLLNGKLPLTIGGGIGQSRL 300

Query: 301 VMFLLRKQHIGEVQTSVWPQEVRDSYDNIL 330
            M LLRK+HIGEVQ+SVWP+E+ + + NIL
Sbjct: 301 AMLLLRKKHIGEVQSSVWPKEMLEEFSNIL 330
```

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

```
 Identities = 254/303 (83%), Positives = 280/303 (91%)

Query: 1   MIDKLEIVEVQGPILSQVGDGMQDNLSGIEHPVSVKVLNIPEAEFEVVHSLAKWKRHTLA 60
            +I KL++VEVQGPILS+VGDGMQDNLSG E+PVSV VL IP A FEVVHSLAKWKRHTLA
Sbjct: 23  LIAKLDVVEVQGPILSRVGDGMQDNLSGTENPVSVNVLKIPNATFEVVHSLAKWKRHTLA 82

Query: 61  RFGFNEGEGLFVHMKALRPDEDSLDPTHSVYVDQWDWEKVIPDGRRNLDYLKETVEKIYK 120
            RFGFNEGEGL V+MKALRPDEDSLD THSVYVDQWDWEKVIPDG+RNL YLKETVE IYK
Sbjct: 83  RFGFNEGEGLVVNMKALRPDEDSLDQTHSVYVDQWDWEKVIPDGKRNLAYLKETVETIYK 142

Query: 121 AIRLTELAVEARFDIESILPKRITFIHTEELVEKYPDLSPKERENAIAKEYGAVFLIGIG 180
             IRLTELAVEAR+DIE++LPK+ITFIHTEELV KYPDL+PKERENAI KE+GAVFLIGIG
Sbjct: 143 VIRLTELAVEARYDIEAVLPKKITFIHTEELVAKYPDLTPKERENAITKEFGAVFLIGIG 202

Query: 181 GELADGKPHDGRAPDYDDWTTPSENGFKGLNGDILVWNEQLGTAFELSSMGIRVDEDALK 240
            G L DGKPHDGRAPDYDDWTT +ENG+ GLNGDILVWN+QLG+AFELSSMGIRVDE+ALK
Sbjct: 203 GVLPDGKPHDGRAPDYDDWTTETENGYHGLNGDILVWNDQLGSAFELSSMGIRVDEEALK 262

Query: 241 RQVVLTGDEDRLEFEWHKTLLRGFFPLTIGGGIGQSRLAMFLLRKXHIGEVQSSVWPKEV 300
            RQV +TGD+DRL F+WHK+LL G FPLTIGGGIGQSR+ MFLLRK HIGEVQ+SVWP+EV
Sbjct: 263 RQVEMTGDQDRLGFDWHKSLLNGLFPLTIGGGIGQSRMVMFLLRKQHIGEVQTSVWPQEV 322

Query: 301 RDT 303
            RD+
Sbjct: 323 RDS 325
```

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

Example 2287

A DNA sequence (GBSx2421) was identified in *S.agalactiae* <SEQ ID 7055> which encodes the amino acid sequence <SEQ ID 7056>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3163(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 2288

A DNA sequence (GBSx2422) was identified in *S.agalactiae* <SEQ ID 7057> which encodes the amino acid sequence <SEQ ID 7058>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have a cleavable N-term signal seq.

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

A related GBS nucleic acid sequence <SEQ ID 9007> which encodes amino acid sequence <SEQ ID 9008> was also identified.

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

```
>GP:AAD56628 GB:AF165218 Bta [Streptococcus pneumoniae]
 Identities = 30/97 (30%), Positives = 50/97 (50%), Gaps = 3/97 (3%)

Query: 50  KALVSKSQQSEATIFIGRPTCQYCRAFLPKLLKSQATLHSKIYYLDSQKYKG-KRLKSFF 108
           +A  +  ++  AT FIGR TC YCR F   L    A   + IY+++S++      L++F
Sbjct: 18  RAQEALDKKETATFFIGRKTCPYCRKFAGTLSGVVAETKAHIYFINSEEASQLNDLQAFR 77

Query: 109 KKHHITTVPNLAHYQQGKMTKYLVQGSQATPQQIQTF 145
            ++ I TVP   H   G++   +   S  + Q+I+ F
Sbjct: 78  SRYGIPTVPGFVHITDGQIN--VRCDSSMSAQEIKDF 112
```

SEQ ID 9008 (GBS134) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 40 (lane 2; MW 17kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 4; MW 42kDa).

GBS134-GST was purified as shown in Figure 204, lane 10.

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

Example 2289

A DNA sequence (GBSx2423) was identified in *S.agalactiae* <SEQ ID 7059> which encodes the amino acid sequence <SEQ ID 7060>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.0735(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 9603> which encodes amino acid sequence <SEQ ID 9604> was also identified.

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

```
>GP:BAB06309 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 78/178 (43%), Positives = 115/178 (63%), Gaps = 3/178 (1%)

Query: 3   MRVVAGTFGGRPLKTLDGKTTRPTTDKVKGAIFNMIGPFFEGGRVLDLFSGSGSLAIEAI 62
           MRV+AG   G  LK + G  TRPTTDKVK AIFNMIGPFF+GG  LDL+ GSG L IEA+
Sbjct: 1   MRVIAGEQKGLTLKAVPGHKTRPTTDKVKEAIFNMIGPFFDGGIGLDLYGGSGGLGIEAL 60

Query: 63  SRGMDQAVLVEKDRRAQVVIQENIAMTKSPEQFQLLKMEANRALEQLTGQ---FDLVLLD 119
           SRG+++ + V++ +RA   I++N++      + ++ + +A RAL+ LT +   F  V LD
Sbjct: 61  SRGVERMIFVDQQKRAIETIKQNLSHCGLEGRAEVYRNDAKRALQVLTKRGIVFAYVFLD 120

Query: 120 PPYAKEEIVKQIQIMDSKGLLGDDIMIACETDKSVDLPEEIASFGIWKQKIYGISKVT 177
           PPYAK+ I   + I+ + GLL +  ++ CE D+   LP++I      K++ YG + +T
Sbjct: 121 PPYAKQTIKNDLAILANHGLLEEGGVVVCEHDRDTMLPDQIEYAVKHKEETYGDTMIT 178
```

There is also homology to SEQ ID 132.

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

Example 2290

A DNA sequence (GBSx2424) was identified in *S.agalactiae* <SEQ ID 7061> which encodes the amino acid sequence <SEQ ID 7062>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4984(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:CAB96619 GB:AJ400630 hypothetical protein [Streptococcus pneumoniae bacteriophage MM1]
 Identities = 175/254 (68%), Positives = 219/254 (85%)

Query: 2   LRRHIYSMLEEHXHLQPEIKYHQKTNLRKNRVYTVFIEEKVDVILADLKLADAFFGIETG 61
           L RH+Y        ++ EI++HQ++NLRKNRVYTVF +EKV  +L+DL LAD+FFG+ETG
Sbjct: 50  LARHLYESFLHFYEIKSEIRHHQRSNLRKNRVYTVFTDEKVQDLLSDLHLADSFFGLETG 109

Query: 62  IEHSILDNDENGRAYLRGAFLSTGTVREPDSGKYQLEIFSVYLDHAQDLANLMKKFMLDA 121
           I+ +IL ++E GRAYL GAFL+ G++R+P+SGKYQLEI SVYLDHAQ +A+L+++F+LDA
Sbjct: 110 IDEAILSDEEAGRAYLCGAFLANGSIRDPESGKYQLEISSVYLDHAQGIASLLQQFLLDA 169

Query: 122 KVIEHKHGAVTYLQKAEDIMDFLIVIDAMEARDAFEEIKMIRETRNDINRANNVETANIA 181
           KV+E K GAVTYLQ+AEDIMDFLIVI AM+ARD FE +K++RETRND+NRANN ETANIA
Sbjct: 170 KVLERKKGAVTYLQRAEDIMDFLIVIGAMQARDDFERVKILRETRNDLNRANNAETANIA 229

Query: 182 RTITASMKTINNIIKIMDTIGFDALPSDLRQVAQVRVAHPDYSIQQIADSLETPLSKSGV 241
           RT++ASMKTINNI KI D +G + LP DL++VAQ+R+ HPDYSIQQ+ADSL TPL+KSGV
Sbjct: 230 RTVSASMKTINNISKIKDIMGLENLPVDLQEVAQLRIQHPDYSIQQLADSLSTPLTKSGV 289

Query: 242 NHRLRKINKIADEL 255
           NHRLRKINKIADEL
Sbjct: 290 NHRLRKINKIADEL 303
```

There is also homology to SEQ ID 5540:

```
Identities = 186/254 (73%), Positives = 227/254 (89%)

Query: 2   LRRHIYSMLEEHXHLQPEIKYHQKTNLRKNRVYTVFIEEKVDVILADLKLADAFFGIETG 61
           + R+IYS++E+    + PEI+YHQKTNLRKNRVYTV++E+ V+ ILADLKLAD+FFG+ETG
```

```
Sbjct: 50  IARYIYSLIEDAYVIVPEIRYHQKTNLRKNRVYTVYVEQGVETILADLKLADSFFGLETG 109

Query: 62  IEHSILDNDENGRAYLRGAFLSTGTVREPDSGKYQLEIFSVYLDHAQDLANLMKKFMLDA 121
           IE  +L +D  GR+YL+GAFL+ G++R+P+SGKYQLEI+SVYLDHAQDLA LM+KFMLDA
Sbjct: 110 IEPQVLSDDNAGRSYLKGAFLAAGSIRDPESGKYQLEIYSVYLDHAQDLAQLMQKFMLDA 169

Query: 122 KVIEHKHGAVTYLQKAEDIMDFLIVIDAMEARDAFEEIKMIRETRNDINRANNVETANIA 181
           K IEHK GAVTYLQKAEDIMDFLI+I AM  ++ FE IK++RE RNDINRANN ETANIA
Sbjct: 170 KTIEHKSGAVTYLQKAEDIMDFLIIIGAMSCKEDFEAIKLLREARNDINRANNAETANIA 229

Query: 182 RTITASMKTINNIIKIMDTIGFDALPSDLRQVAQVRVAHPDYSIQQIADSLETPLSKSGV 241
           +TI+ASMKTINNIIKIMDTIG ++LP +L+QVAQ+RV HPDYSIQQ+AD+LE P++KSGV
Sbjct: 230 KTISASMKTINNIIKIMDTIGLESLPIELQQVAQLRVKHPDYSIQQVADALEFPITKSGV 289

Query: 242 NHRLRKINKIADEL 255
           NHRLRKINKIAD+L
Sbjct: 290 NHRLRKINKIADDL 303
```

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

Example 2291

A DNA sequence (GBSx2425) was identified in *S.agalactiae* <SEQ ID 7063> which encodes the amino acid sequence <SEQ ID 7064>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0297(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 2292

A DNA sequence (GBSx2428) was identified in *S.agalactiae* <SEQ ID 7065> which encodes the amino acid sequence <SEQ ID 7066>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2706(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:CAB54571 GB:AJ006393 response regulator [Streptococcus pneumoniae]
 Identities = 139/190 (73%), Positives = 166/190 (87%)

Query: 8   IKIVLVDDHEMVRLGLKSFLNLQADVEVIGEASNGLEGIKKALELRPDVVVMDLVMPEMD 67
           +KI+LVDDHEMVRLGLKS+ +LQ DVEV+GEASNG +GI  ALELRPDV+VMD+VMPEM+
Sbjct: 1   MKILLVDDHEMVRLGLKSYFDLQDDVEVVGEASNGSQGIDLALELRPDVIVMDIVMPEMN 60
```

```
Query: 68  GVEATLALLKDWPEAAILVLTSYLDNEKIYPVIEAGAKGYMLKTSSAAEILNAIRKVSRG 127
           G++ATLA+LK+WPEA IL++TSYLDNEKI PV++AGAKGYMLKTSSA E+L+A+ KV+ G
Sbjct: 61  GIDATLAILKEWPEAKILIVTSYLDNEKIMPVLDAGAKGYMLKTSSADELLHAVSKVAAG 120

Query: 128 EQAIENEVDKKIKAHDKCPALHEGLTARERDILNLLAKGYDNQRIADELFISLKTVKTHV 187
           E AIE EV KK++ H      LHE LTARERD+L L+AKGY+NQRIAD+LFISLKTVKTHV
Sbjct: 121 ELAIEQEVSKKVEYHRNHMELHEELTARERDVLQLIAKGYENQRIADDLFISLKTVKTHV 180

Query: 188 SNILGKLNGS 197
           SNIL KL  S
Sbjct: 181 SNILAKLEVS 190
```

There is also high homology to SEQ ID 2996:

```
Identities = 158/198 (79%), Positives = 176/198 (88%), Gaps = 1/198 (0%)

Query: 5   MDKIKIVLVDDHEMVRLGLKSFLNLQADVEVIGEASNGLEGIKKALELRPDVVVMDLVMP 64
           M KIK++LVDDHEMVR+GLKSFLNLQAD++V+GEASNG EG+  AL L+PDV+VMDLVMP
Sbjct: 3   MSKIKVILVDDHEMVRMGLKSFLNLQADIDVVGEASNGREGVDLALALKPDVLVMDLVMP 62

Query: 65  EMDGVEATLALLKDWPEAAILVLTSYLDNEKIYPVIEAGAKGYMLKTSSAAEILNAIRKV 124
           E+ GVEATL +LK W EA +LVLTSYLDNEKIYPVI+AGAKGYMLKTSSAAEILNAIRKV
Sbjct: 63  ELGGVEATLEVLKKWKEAKVLVLTSYLDNEKIYPVIDAGAKGYMLKTSSAAEILNAIRKV 122

Query: 125 SRGEQAIENEVDKKIKAHDKCPALHEGLTARERDILNLLAKGYDNQRIADELFISLKTVK 184
           S+GE AIE EVDKKIKAHD+ P LHE LTARE DIL+LLAKGYDNQ IADELFISLKTVK
Sbjct: 123 SKGELAIETEVDKKIKAHDQHPDLHEELTAREYDILHLLAKGYDNQTIADELFISLKTVK 182

Query: 185 THVSNILGKLN-GSRSNS 201
           THVSNIL KL  G R+ +
Sbjct: 183 THVSNILAKLEVGDRTQA 200
```

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

Example 2293

A DNA sequence (GBSx2429) was identified in *S.agalactiae* <SEQ ID 7067> which encodes the amino acid sequence <SEQ ID 7068>. This protein is predicted to be histidine kinase (narQ). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3944(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:CAB54570 GB:AJ006393 histidine kinase [Streptococcus pneumoniae]
 Identities = 32/55 (58%), Positives = 49/55 (88%)

Query: 1   MIDNGIGFDMDSVYDLSYGLKNIEDRVEDLAGNLQLLSQPGKGVAMDIRLPLVNQ 55
           ++DNGIGF + S+ DLSYGL+NI++RVED+AG +QLL+ P +G+A+DIR+PL++
Sbjct: 276 VVDNGIGFQLGSLDDLSYGLRNIKERVEDMAGTVQLLTAPKQGLAVDIRIPLLDK 330
```

There is also homology to SEQ ID 2992:

```
Identities = 44/59 (74%), Positives = 51/59 (85%)

Query: 1   MIDNGIGFDMDSVYDLSYGLKNIEDRVEDLAGNLQLLSQPGKGVAMDIRLPLVNQSEDK 59
           MID+G+GFDMD V DLSYGLKNIEDRV DLAGNL L+SQ GKGV+MDIRLP+V   +D+
Sbjct: 276 MIDDGVGFDMDQVRDLSYGLKNIEDRVNDLAGNLHLISQKGKGVSMDIRLPIVKGDDDE 334
```

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

Example 2294

A DNA sequence (GBSx2430) was identified in *S.agalactiae* <SEQ ID 7069> which encodes the amino acid sequence <SEQ ID 7070>. This protein is predicted to be RfbQRSO155-1. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1120(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

There is also homology to SEQ ID 7072:

```
Identities = 171/172 (99%), Positives = 172/172 (99%)

Query: 1   MGQVAVEEKSNEIVAIPQLLRTIDIRKSIVTIDAMGTQTAIVDTIIKGKADYCLAVKGNQ 60
           +GQVAVEEKSNEIVAIPQLLRTIDIRKSIVTIDAMGTQTAIVDTIIKGKADYCLAVKGNQ
Sbjct: 143 LGQVAVEEKSNEIVAIPQLLRTIDIRKSIVTIDAMGTQTAIVDTIIKGKADYCLAVKGNQ 202

Query: 61  ETLYDDIALYFSDVNLLEELQENAQYYQTVEKSRGQIEVREYWVSSDIKWLCQNHPKWHK 120
           ETLYDDIALYFSDVNLLEELQENAQYYQTVEKSRGQIEVREYWVSSDIKWLCQNHPKWHK
Sbjct: 203 ETLYDDIALYFSDVNLLEELQENAQYYQTVEKSRGQIEVREYWVSSDIKWLCQNHPKWHK 262

Query: 121 LRGIGMTRNTIDKDGQLSQENRYFIFSFKPDVLTFANCVRGHWQIESMHWLL 172
           LRGIGMTRNTIDKDGQLSQENRYFIFSFKPDVLTFANCVRGHWQIESMHWLL
Sbjct: 263 LRGIGMTRNTIDKDGQLSQENRYFIFSFKPDVLTFANCVRGHWQIESMHWLL 314
```

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

Example 2295

A DNA sequence (GBSx2431) was identified in *S.agalactiae* <SEQ ID 7073> which encodes the amino acid sequence <SEQ ID 7074>. This protein is predicted to be translation initiation factor if-3 homolog dsg (infC). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1787(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:CAA68920 GB:Y07640 translation initiation factor, IF3 [Listeria monocytogenes]
 Identities = 112/169 (66%), Positives = 134/169 (79%)

Query: 7   KDLFINDEIRVREVRLVGLEGEQLGIKPLSEAQAIADDANVDLVLIQPQATPPVAKIMDY 66
           KD+ +ND IR REVRL+  +GEQLG+K   +A  IA+ AN+DLVL+ P A PPVA+IMDY
Sbjct: 3   KDMLVNDGIRAREVRLIDQDGEQLGVKSKIDALQIAEKANLDLVLVAPTAKPPVARIMDY 62

Query: 67  GKFKFEYQKKQKEQRKKQSVVTVKEVRLSPVIDKGDFETKLRNGRKFLEKGNKVKVSIRF 126
           GKF+FE QKK KE RK Q V+ +KEVRLSP ID+ DF+TKLRN RKFLEKG+KVK SIRF
```

```
Sbjct: 63  GKFRFEQQKKDKEARKNQKVIVMKEVRLSPTIDEHDFDTKLRNARKFLEKGDKVKCSIRF 122

Query: 127 KGRMITHKEIGAKVLAEFAEATQDIAIIEQRAKMDGRQMFMQLAPIPDK 175
           KGR ITHKEIG KVL  FA+A +D+  IEQR KMDGR MF+ LAP+ +K
Sbjct: 123 KGRAITHKEIGQKVLDRFAKACEDLCTIEQRPKMDGRSMFLVLAPLHEK 171
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7075> which encodes the amino acid sequence <SEQ ID 7076>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2247(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 = 167/176 (94%), Positives = 173/176 (97%)

Query: 1   MKIIAKKDLFINDEIRVREVRLVGLEGEQLGIKPLSEAQAIADDANVDLVLIQPQATPPV 60
           +KIIAKKDLFINDEIRVREVRLVGLEGEQLGIKPLSEAQ++AD +NVDLVLIQPQA PPV
Sbjct: 1   VKIIAKKDLFINDEIRVREVRLVGLEGEQLGIKPLSEAQSLADASNVDLVLIQPQAVPPV 60

Query: 61  AKIMDYGKFKFEYQKKQKEQRKKQSVVTVKEVRLSPVIDKGDFETKLRNGRKFLEKGNKV 120
           AK+MDYGKFKFEYQKKQKEQRKKQSVVTVKEVRLSPVIDKGDFETKLRNGRKFLEKGNKV
Sbjct: 61  AKLMDYGKFKFEYQKKQKEQRKKQSVVTVKEVRLSPVIDKGDFETKLRNGRKFLEKGNKV 120

Query: 121 KVSIRFKGRMITHKEIGAKVLAEFAEATQDIAIIEQRAKMDGRQMFMQLAPIPDKK 176
           KVSIRFKGRMITHKEIGAKVLA+FAEATQDIAIIEQRAKMDGRQMFMQLAPI DKK
Sbjct: 121 KVSIRFKGRMITHKEIGAKVLADFAEATQDIAIIEQRAKMDGRQMFMQLAPISDKK 176
```

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

Example 2296

A DNA sequence (GBSx2432) was identified in *S.agalactiae* <SEQ ID 7077> which encodes the amino acid sequence <SEQ ID 7078>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1807(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:AAC45308 GB:U81957 RNA polymerase beta' subunit [Streptococcus gordonii]
 Identities = 262/286 (91%), Positives = 276/286 (95%)

Query: 1   MAAKVVKAGVEEVXIRSVFTCNTRHGVCRHCYGINLATGDAVEVGEAVGTIAAQSIGEPG 60
           MA +VV AGV EV IRSV TCNTRHGVCRHCYGINLATGDAVEVGEAVGTIAAQSIGEPG
Sbjct: 122 MARQVVNAGVTEVTIRSVLTCNTRHGVCRHCYGINLATGDAVEVGEAVGTIAAQSIGEPG 181

Query: 61  TQLTMRTFHTGGVASNTDITQGLPRIQEIFEARNPKGEAVITEVKGEVVAIEEDSSTRTK 120
           TQLTMRTFHTGGVAS++DITQGLPR+QEIFEARNPKGEAVITEVKGEV AIEED+STRTK
Sbjct: 182 TQLTMRTFHTGGVASSSDITQGLPRVQEIFEARNPKGEAVITEVKGEVTAIEEDASTRTK 241

Query: 121 KVFVKGQTGEGEYVVPFTARMKVEVGDEVARGAALTEGSIQPKRLLEVRDTLSVETYLLA 180
           KVFVKGQTGEGEYVVPFTARMKVEVGD+V+RGAALTEGSIQPK LL VRD LSVETYLLA
Sbjct: 242 KVFVKGQTGEGEYVVPFTARMKVEVGDQVSRGAALTEGSIQPKHLLAVRDVLSVETYLLA 301
```

```
Query: 181 EVQKVYRSQGVEIGDKHVEVMVRQMLRKVRVMDPGDTDLLPGTLMDISDFTDANKDIVIS 240
           EVQKVYRSQGVEIGDKH+EVMVRQM+RKVRVMDPGDTDLL GTLMDI+DFTDAN+D+VIS
Sbjct: 302 EVQKVYRSQGVEIGDKHIEVMVRQMIRKVRVMDPGDTDLLMGTLMDITDFTDANRDVVIS 361

Query: 241 GGIPATSRPVLMGITKASLETNSFLSAASFQETTRVLTDAAIRGKK 286
           GG+PAT+RPVLMGITKASLETNSFLSAASFQETTRVLTDAAIRGKK
Sbjct: 362 GGVPATARPVLMGITKASLETNSFLSAASFQETTRVLTDAAIRGKK 407
```

There is also homology to SEQ ID 384.

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

Example 2297

A DNA sequence (GBSx2434) was identified in *S.agalactiae* <SEQ ID 7079> which encodes the amino acid sequence <SEQ ID 7080>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0352(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 2298

A DNA sequence (GBSx2435) was identified in *S.agalactiae* <SEQ ID 7081> which encodes the amino acid sequence <SEQ ID 7082>. This protein is predicted to be acetoin dehydrogenase (TPP-dependent) beta chain (pdhB). Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.0266(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:BAB04496 GB:AP001509 acetoin dehydrogenase (TPP-dependent) beta
          chain [Bacillus halodurans]
 Identities = 37/57 (64%), Positives = 50/57 (86%)

Query: 1   MLEEFGAKRVRDTPISEAAIAGSAIGAAQTGLRPIVDLTFMDFVTIAMDAIVDDCIR 57
           M+EEFG++RVR+TPISEAAI+G+AIGAA TG+RPI++L F DF+TIAMD +V+   +
Sbjct: 44  MIEEFGSERVRNTPISEAAISGTAIGAALTGMRPILELQFSDFITIAMDNMVNQAAK 100
```

There is also homology to SEQ ID 4272.

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

Example 2299

A DNA sequence (GBSx2436) was identified in *S.agalactiae* <SEQ ID 7083> which encodes the amino acid sequence <SEQ ID 7084>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3015(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:AAB18706 GB:U38906 Structural protein [Bacteriophage r1t]
 Identities = 57/127 (44%), Positives = 83/127 (64%)

Query: 5   IKAGTLFKPELVTEIMSKVKGHSTLAKLSGQTPIPFNGVEQFVFNLDGNAQIVGEGEQKL 64
           +  GTLF P LVT+++SKV G S++A+LS Q PIPFNG + F F +D     +V E  +K
Sbjct: 3   LNKGTLFDPTLVTDLISKVAGKSSIARLSAQKPIPFNGEKVFTFTMDSEIDVVAESGKKT 62

Query: 65  GNTAKVTSKIIKPLKFVYQARMTDEFKYASEEKRLNFLKHYADGFAKKMAEAFDIAAIHG 124
                +  + + P+K  Y AR++DEF YAS+E+++N L+ + DGFAKK+A   D+ A HG
Sbjct: 63  HGGVTLAPQTMVPIKVEYGARISDEFMYASDEEKINILQEFNDGFAKKVARGIDLMAFHG 122

Query: 125 LEPRTMT 131
           + PR  T
Sbjct: 123 VNPRLGT 129
```

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

Example 2300

A DNA sequence (GBSx2439) was identified in *S.agalactiae* <SEQ ID 7085> which encodes the amino acid sequence <SEQ ID 7086>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Example 2301

A DNA sequence (GBSx2440) was identified in *S.agalactiae* <SEQ ID 7087> which encodes the amino acid sequence <SEQ ID 7088>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2227(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

Example 2302

A DNA sequence (GBSx2441) was identified in *S.agalactiae* <SEQ ID 7089> which encodes the amino acid sequence <SEQ ID 7090>. 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=0.2948(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 9319> which encodes amino acid sequence <SEQ ID 9320> was also identified.

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

```
>GP:CAB96616 GB:AJ400629 integrase [Streptococcus pneumoniae
           bacteriophage MM1]
 Identities = 84/238 (35%), Positives = 137/238 (57%), Gaps = 8/238 (3%)

Query: 1   MTLDKNSSQAQKKAGLILQEKIEDRLAIRNHSEMTYGELKKEYLKQWIPTVKDSTKRGYL 60
           +T++K + QA+ +A ++LQEKI  +L+ +    +T+ E+   + K W  TVK+STK
Sbjct: 30  VTMEKKTPQARNQAAILLQEKINKKLSTKQVESITFEEIYNLFYKSWAQTVKESTKHNCK 89

Query: 61  VSDSHIATVLPDDTIINKLTKRDIRLIIDKLLKHNSYHVTHKCRKRLHAIFSYAIQMDYM 120
             D  +  V+P DTI+  L +R ++  I+K+++ N Y    K R RL  IF+YA+Q  Y+
Sbjct: 90  SVDKKMKEVIPSDTILANLDRRFLQEAIEKIIESNGYITAKKVRHRLRGIFNYAVQYSYI 149

Query: 121 TSNPTENVLVP-KPK--DDYKPEKVLYLTSNEV---YDLCNRMIDNDEQTLADIVLFMFL 174
            +N  +   +P KPK  ++ + ++  +LT  E+    D+ NR      Q  AD+VL + L
Sbjct: 150 ENNEVDYTTIPQKPKTLEELEKKRNNFLTMQEIKALVDVLNRR--EYHQKYADMVLVLTL 207

Query: 175 TGVRYGELSCLTYDKIDFENKEILINATYDFNTRXITTTKTKKSTRKISVSDNILDIV 232
           TG+RYGEL+ L    IDFEN +I I   +D   +  T  KT  S R I VS+++++ +
Sbjct: 208 TGMRYGELTALQLKNIDFENNKIEITGNFDSVNKIKTLPKTTNSIRTIKVSESVIEAI 265
```

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 2303

A DNA sequence (GBSx2444) was identified in *S.agalactiae* <SEQ ID 7091> which encodes the amino acid sequence <SEQ ID 7092>. Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
        bacterial cytoplasm --- Certainty=0.2518(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.

There is also homology to SEQ ID 4212:

```
Identities = 92/144 (63%), Positives = 118/144 (81%), Gaps = 1/144 (0%)

Query: 1    MPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQE 60
            +PKYSLFELENGR+R+LASAGELQKGNELALP++++ FLYLAS Y + KG PE+ E+KQ
Sbjct: 1198 LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL 1257

Query: 61   FVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELANNIINLFTFT 120
            FV QH  Y D+I++ I++FSKRVILADANL+K+   Y  +++   + E A NII+LFT T
Sbjct: 1258 FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK-PIREQAENIIHLFTLT 1316

Query: 121  SLGAPAAFKFFDKIVDRKRYTSTQ 144
            +LGAPAAFK+FD  +DRKRYTST+
Sbjct: 1317 NLGAPAAFKYFDTTIDRKRYTSTK 1340
```

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

Example 2304

A DNA sequence (GBSx2445) was identified in *S.agalactiae* <SEQ ID 7093> which encodes the amino acid sequence <SEQ ID 7094>. This protein is predicted to be 0. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.57   Transmembrane  239 - 255 ( 236 - 256)

----- 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 GENPEPT database.

```
>GP:CAB15253 GB:Z99120 similar to opine catabolism [Bacillus subtilis]
 Identities = 88/257 (34%), Positives = 129/257 (49%), Gaps = 11/257 (4%)

Query: 1   MARLGADFYSKLVTDLQKDGFETKFYQQTGVFLLKKDESQLESLFALADKRRLESPLIGD 60
           +A+ GA +Y  L+  L+KDG      Y++ G   +  D S+L+ +   A KRR ++P IGD
Sbjct: 61  LAKGGARYYKDLIHQLEKDGESDTGYKRVGAISIHTDASKLDKMEERAYKRREDAPEIGD 120

Query: 61  LQILNKSEANTHFPEL-DGYEQLLYASGGARVEGADLTRILLEAS---GVNVIKDEVHF- 115
           +  L+ SE     FP L DGYE  ++ SG ARV G  L R LL A+   G  VIK
Sbjct: 121 ITRLSASETKKLFPILADGYES-VHISGAARVNGRALCRSLLSAAEKRGATVIKGNASLL 179

Query: 116 ----TITDNGFRVQGIDFDKLVLASGAWLAKILDEHNYQVDVRPQKGQLRDYYFSNINTG 171
               T+T      +    D +++ +GAW  +IL        V  QK Q+  +  ++ +TG
Sbjct: 180 FENGTVTGVQTDTKQFAADAVIVTAGAWANEILKPLGIHFQVSFQKAQIMHFEMTDADTG 239

Query: 172 KYPVVMPEGELDIIPFDNGKVSVGASHENDMAF-DLNIDFKVLDKFEEQAIGYFPQLKKQ 230
            +PVVMP  +  I+ FDNG++  GA+HEND    DL +      +  +A+   P L
Sbjct: 240 SWPVVMPPSDQYILSFDNGRIVAGATHENDAGLDDLRVTAGGQHEVLSKALAVAPGLADA 299

Query: 231 IRLLKRVEFVPIQVIFL 247
            +  RV F P    FL
```

```
Sbjct: 300 AAVETRVGFRPFTPGFL 316
```

There is also homology to SEQ ID 2656.

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

Example 2305

A DNA sequence (GBSx2446) was identified in *S.agalactiae* <SEQ ID 7095> which encodes the amino acid sequence <SEQ ID 7096>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2572(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 9315> which encodes amino acid sequence <SEQ ID 9316> was also identified.

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

```
>GP:AAC00337 GB:AF008220 YtqI [Bacillus subtilis]
 Identities = 119/256 (46%), Positives = 174/256 (67%), Gaps = 3/256 (1%)

Query: 6   QILDKIKEYDTIIIHRHMRPDPDALGSQIGLRDIIRHNFPKKKVLATGFDEPTLAWIAKM 65
           +++  I  YDTII+HRH+RPDPDA GSQ GL +I+R  +P+K + A G  EP+L+++  +
Sbjct: 4   ELIRTISLYDTIILHRHVRPDPDAYGSQCGLTEILRETYPEKNIFAVGTPEPSLSFLYSL 63

Query: 66  DQVTDQDYQGALVVVTDTANTPRIDDERYKKGDFLIKIDHHPNDEVYGDLSYVDTNASSA 125
           D+V ++ Y+GALV+V DTAN  RIDD+RY  G  L+KIDHHPN++ YGDL +VDT+ASS
Sbjct: 64  DEVDNETYEGALVIVCDTANQERIDDQRYPSGAKLMKIDHHPNEDPYGDLLWVDTSASSV 123

Query: 126 SEIVTDFAL---SCDLLLSTSAARVLYNGIVGDTGRFLYPATTSKTLKIASKLREFDFDF 182
           SE++ +  L        L+T AA ++Y GIVGDTGRFL+P TT KTLK A +L ++ F
Sbjct: 124 SEMIYELYLEGKEHGWKLNTKAAELIYAGIVGDTGRFLFPNTTEKTLKYAGELIQYPFSS 183

Query: 183 SAMARQMDSFPFKIAKLQGFIFEQLKIDKNGAACVTLTQEDLKRFDVTDAETAAIVGVPG 242
           S +  Q+       + KL GFIF+ + + +NGAA V + ++ L++F  T +E + +VG  G
Sbjct: 184 SELFNQLYETKLNVVKLNGFIFQNVSLSENGAASVFIKKDTLEKFGTTASEASQLVGTLG 243

Query: 243 KIDIVESWAIFVKQSD 258
            I  + +W  FV++ D
Sbjct: 244 NISGIRAWVFFVEEDD 259
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7097> which encodes the amino acid sequence <SEQ ID 7098>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2584(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 = 180/256 (70%), Positives = 215/256 (83%)

Query: 4   FQQILDKIKEYDTIIIHRHMRPDPDALGSQIGLRDIIRHNFPKKKVLATGFDEPTLAWIA 63
```

```
            F+ ILDKIK + TIIIHRH  PDPDALGSQ GL++II  NFP KKVL TGFDEP+LAWI+
Sbjct: 5    FETILDKIKAHQTIIIHRHQNPDPDALGSQAGLKEIIAQNFPDKKVLMTGFDEPSLAWIS 64

Query: 64   KMDQVTDQDYQGALVVVTDTANTPRIDDERYKKGDFLIKIDHHPNDEVYGDLSYVDTNAS 123
            +MDQVTD+DY+ ALV++TDTAN PRIDDERY  G  LIKIDHHPND+VYGD  YVDT+AS
Sbjct: 65   QMDQVTDKDYKEALVIITDTANRPRIDDERYTLGKCLIKIDHHPNDDVYGDFYYVDTSAS 124

Query: 124  SASEIVTDFALSCDLLLSTSAARVLYNGIVGDTGRFLYPATTSKTLKIASKLREFDFDFS 183
            SASEI+ DFA S +L LS  AA++LY GIVGDTGRFLY +TTSKTL IAS+LR F+FDF+
Sbjct: 125  SASEIIADFAFSQNLTLSDKAAKLLYTGIVGDTGRFLYASTTSKTLSIASQLRHFEFDFA 184

Query: 184  AMARQMDSFPFKIAKLQGFIFEQLKIDKNGAACVTLTQEDLKRFDVTDAETAAIVGVPGK 243
            A++RQMDSFP KIAKLQ ++FE L ID++GAA V ++QE LK FDVT AE++AIV  PGK
Sbjct: 185  AISRQMDSFPLKIAKLQSYVFEHLTIDESGAAYVLVSQETLKHFDVTLAESSAIVCAPGK 244

Query: 244  IDIVESWAIFVKQSDG 259
            ID V++WAIFV+ +DG
Sbjct: 245  IDNVQAWAIFVELTDG 260
```

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

Example 2306

A DNA sequence (GBSx2447) was identified in *S.agalactiae* <SEQ ID 7099> which encodes the amino acid sequence <SEQ ID 7100>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1846(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:CAB42949 GB:AL049863 putative adenosine deaminase [Streptomyces
           coelicolor A3(2)]
 Identities = 123/343 (35%), Positives = 175/343 (50%), Gaps = 26/343 (7%)

Query: 6    LKELAKAELHCHLDGSLSLPAIRKLANMADIILPSSDK-ELRKYVIAPAQTESLVDYLKT 64
            L+ L KA LH HLDG L    + +LA      LP++D  EL  +    A +  LV Y+ T
Sbjct: 11   LRRLPKAVLHDHLDGGLRPATVVELARSVGHTLPTTDPDELAAWYYEAANSGDLVRYIAT 70

Query: 65   FEFIRPLLQTKEALRFAAYDVARQAALENVIYIEIRFAPELSMDKGLTASDTVLAVLEGL 124
            FE    ++Q +E L  AA +     A + V+Y E+R+APEL+   GL+  + V  V EGL
Sbjct: 71   FEHTLAVMQNREGLLRAAEEYVLDLAADGVVYGEVRYAPELNTRGGLSMREVVETVQEGL 130

Query: 125  ADAQKEFNIVAR-----ALVCGMRQSSHKTTKDIIKHIVDLA----PKGLVGFDFAGDEF 175
            A    +             L+CGMR        D ++   DLA       G+VGFD AG E
Sbjct: 131  ATGMAKAAAAGTPVRVGTLLCGMRMF------DRVREAADLAVAFRDAGVVGFDIAGAED 184

Query: 176  SYPTDSLVDLIQEVKRSGYPMTLHAGECGCAKHIADSLNL-GIKRMGHVTALT------- 227
             +P    +D  + ++R   P T+HAGE        I  +L + G +R+GH    +T
Sbjct: 185  GFPPADHLDAFEHLRRENVPFTIHAGEAHGLPSIHQALQVCGAQRIGHGVRITDDIPDLA 244

Query: 228  -GQRDLIKRFVEEDAVA-EMCLTSNLQTKAASSIQSFPYQELYDAGGKITINTDNRTVSD 285
             G+   +  +V +  +A EMC TSNLQT AA+SI   P   L D G ++T+NTDNR VS
Sbjct: 245  AGKLGRLAAWVRDRRIALEMCPTSNLQTGAATSIAEHPITALKDLGFRVTLNTDNRLVSG 304

Query: 286  TNLTKEYSLFVTYFGTKIEDFLVFNQNAVKASFTSDSEKDTLL 328
            T +T+E SL V   G  +ED      NA+K++F    E+  L+
Sbjct: 305  TTMTREMSLLVEQAGWSVEDLRTVTVNALKSAFVPFDERTALI 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 2307

A DNA sequence (GBSx2448) was identified in *S.agalactiae* <SEQ ID 7101> which encodes the amino acid sequence <SEQ ID 7102>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2042(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 9639> which encodes amino acid sequence <SEQ ID 9640> was also identified.

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

```
>GP:CAB13290 GB:Z99111 similar to sulfite reductase [Bacillus subtilis]
 Identities = 63/146 (43%), Positives = 87/146 (59%), Gaps = 1/146 (0%)

Query: 5   MALAKIVYASMTGNTEEIADIVADKLRDLGLDVEVEECTMVDAAD-FEDADIAIVATYTY 63
           MA   +VYA+M+GNTE +AD++   L++   +V+  E   +D A  F D D  I+ TYT+
Sbjct: 1   MAKILLVYATMSGNTEAMADLIEKGLQEALAEVDRFEAMDIDDAQLFTDYDHVIMGTYTW 60

Query: 64  GDGDLPDEIVDFYEDLAEVDLSGKVYGVVGSGDTFYDYFCKSVDEFEAQFALTGAQKGAD 123
           GDGDLPDE +D  ED+ E+D SGK   V GSGDT Y++FC +VD  EA+    G
Sbjct: 61  GDGDLPDEFLDLVEDMEEIDFSGKTCAVFGSGDTAYEFFCGAVDTLEAKIKERGGDIVLP 120

Query: 124 CVKVDLAAEDEDIENLEAFAEEIASK 149
            VK++   E E+ E L  F  + A K
Sbjct: 121 SVKIENNPEGEEEEELINFGRQFAKK 146
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7103> which encodes the amino acid sequence <SEQ ID 7104>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1641(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 = 116/147 (78%), Positives = 136/147 (91%)

Query: 5   MALAKIVYASMTGNTEEIADIVADKLRDLGLDVEVEECTMVDAADFEDADIAIVATYTYG 64
           MALAKIVYASMTGNTEEIADIVA+KL++LG DV+++ECT VDA++FE+ADIA+VATYTYG
Sbjct: 1   MALAKIVYASMTGNTEEIADIVANKLQELGHDVDIDECTTVDASEFENADIAVVATYTYG 60

Query: 65  DGDLPDEIVDFYEDLAEVDLSGKVYGVVGSGDTFYDYFCKSVDEFEAQFALTGAQKGADC 124
           DGDLPDEIVDFYEDL ++DL GK+YGVVGSGDTFYDYFCKSVD+F  QFALTGA KGA+
Sbjct: 61  DGDLPDEIVDFYEDLQDLDLEGKIYGVVGSGDTFYDYFCKSVDDFSEQFALTGAIKGAEP 120

Query: 125 VKVDLAAEDEDIENLEAFAEEIASKLN 151
           VKVDLAAEDEDI+ LEAFAE+++  +N
Sbjct: 121 VKVDLAAEDEDIDRLEAFAEQLSQAVN 147
```

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

Example 2308

A DNA sequence (GBSx2449) was identified in *S.agalactiae* <SEQ ID 7105> which encodes the amino acid sequence <SEQ ID 7106>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3568(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:AAB98234 GB:U67480 chorismate mutase/prephenate dehydratase
          (pheA) [Methanococcus jannaschii]
 Identities = 26/85 (30%), Positives = 46/85 (53%), Gaps = 1/85 (1%)

Query: 2  ELEEIRQEIDEIDQQLVSLLETRMGLILEVIAFKKKHRLPVLDNNRENEVLNNVLKKVQN 61
          +L EIR++IDEID +++ L+  R  L  +V   K +  +P+ D  RE  + + + K  +
Sbjct: 4  KLAEIRKKIDEIDNKILKLIAERNSLAKDVAEIKNQLGIPINDPEREKYIYDRIRKLCKE 63

Query: 62 HQFDDVIRATFKDIMTE-SRVYQKE 85
          H  D+ I      I+ E ++  QK+
Sbjct: 64 HNVDENIGIKIFQILIEHNKALQKQ 88
```

There is also homology to SEQ ID 1568.

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

Example 2309

A DNA sequence (GBSx2450) was identified in *S.agalactiae* <SEQ ID 7107> which encodes the amino acid sequence <SEQ ID 7108>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1828(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:AAC34413 GB:AF158600 putative minor structural protein
          [Streptococcus thermophilus bacteriophage Sfi11]
 Identities = 39/65 (60%), Positives = 54/65 (83%)

Query: 1   MEVETDSQEVLMSTGLKDLKAHAYPAITYEVDGYVDLELGDVVRIQDDGYEPPLILTARV 60
           ME++TDS++VL+ST L++L+    YPAITYEVDG++DL++GD V+IQD G+ P L+L ARV
Sbjct: 707 MEIDTDSEDVLISTALRNLRKFCYPAITYEVDGFLDLDIGDTVKIQDTGFSPMLMLEARV 766

Query: 61  VEQDI 65
           EQ I
Sbjct: 767 SEQQI 771
```

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 2310

A DNA sequence (GBSx2451) was identified in *S.agalactiae* <SEQ ID 7109> which encodes the amino acid sequence <SEQ ID 7110>. This protein is predicted to be phosphomethylpyrimidine kinase (thiD). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2051(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:AAC22074 GB:U32725 phosphomethylpyrimidine kinase (thiD)
          [Haemophilus influenzae Rd]
 Identities = 29/78 (37%), Positives = 48/78 (61%), Gaps = 2/78 (2%)

Query: 4  RNVLAISGNDIFSGGGLHADLATYVVNKLHGFVAVTCLTAMSDKG-FEVIPIEASILKQQ 62
          + VL I+G+D   G G+ ADL T+ +  + G  A+T +TA +  G F++ PI    ++ Q
Sbjct: 5  KQVLTIAGSDSGGGAGIQADLKTFQMRGVFGTSAITAVTAQNTLGVFDIHPIPLKTIQAQ 64

Query: 63 LESLK-DVEFGSIKLGLL 79
          LE++K D +  S K+G+L
Sbjct: 65 LEAVKNDFQIASCKIGML 82
```

There is also homology to SEQ ID 4408.

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

Example 2311

A DNA sequence (GBSx2452) was identified in *S.agalactiae* <SEQ ID 7111> which encodes the amino acid sequence <SEQ ID 7112>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.43   Transmembrane  109 - 125 ( 102 - 129)
   INTEGRAL    Likelihood = -1.28   Transmembrane   84 - 100 (  84 - 100)

----- Final Results -----
               bacterial membrane --- Certainty=0.3972(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:CAA22372 GB:AL034446 putative transmembrane protein
          [Streptomyces coelicolor A3(2)]
 Identities = 25/93 (26%), Positives = 43/93 (45%), Gaps = 1/93 (1%)

Query: 62  SASVEILCRGWLLPVSATKYSKIVSVSISSIFFGLLHSANNHVSLISIFNLCL-FGLFLS 120
           +A+ E++ RG L  +        +++ ++ + FGL+H  N   +L     + +  G  L+
Sbjct: 143 AATEEVVFRGVLFRIIEEHIGTYLALGLTGLVFGLMHLLNEDATLWGALAIAIEAGFMLA 202
```

```
Query: 121 LYVILKGNIWGACGIHGAWNCVQGSVFGIEVSG 153
                N+W   G+H  WN   G VF   VSG
Sbjct: 203 AAYAATRNLWLTIGVHFGWNFAAGGVFSTVVSG 235
```

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 2312

A DNA sequence (GBSx2453) was identified in *S.agalactiae* <SEQ ID 7113> which encodes the amino acid sequence <SEQ ID 7114>. This protein is predicted to be pppL protein. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5796(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:CAA10712 GB:AJ132604 pppL protein [Lactococcus lactis]
 Identities = 38/64 (59%), Positives = 51/64 (79%)

Query: 1  MEISLLTDIGQRRSNNQDFINQFENKAGVPLIILADGMGGHRAGNIASEMTVTDLGSDWA 60
          ME S+L+DIG +RS NQD++  + N+AG  L +LADGMGGH+AGN+AS++TV DLG  W+
Sbjct: 1  MEYSILSDIGSKRSTNQDYVGTYVNRAGYQLFLLADGMGGHKAGNVASKLTVEDLGKLWS 60

Query: 61 ETDF 64
          ET F
Sbjct: 61 ETFF 64
```

There is also homology to SEQ ID 3022:

```
Identities = 58/74 (78%), Positives = 69/74 (92%)

Query: 1  MEISLLTDIGQRRSNNQDFINQFENKAGVPLIILADGMGGHRAGNIASEMTVTDLGSDWA 60
          M+ISL TDIGQ+RSNNQDFIN+F+NK G+ L+ILADGMGGHRAGNIASEMTVTDLG +W
Sbjct: 1  MKISLKTDIGQKRSNNQDFINKFDNKKGITLVILADGMGGHRAGNIASEMTVTDLGREWV 60

Query: 61 ETDFSELSEIRDWM 74
          +TDF+ELS+IRDW+
Sbjct: 61 KTDFTELSQIRDWL 74
```

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

Example 2313

A DNA sequence (GBSx2454) was identified in *S.agalactiae* <SEQ ID 7115> which encodes the amino acid sequence <SEQ ID 7116>. This protein is predicted to be sunL protein. Analysis of this protein sequence reveals the following:

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

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

```
bacterial cytoplasm --- Certainty=0.1631(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:CAA10711 GB:AJ132604 sunL protein [Lactococcus lactis]
 Identities = 48/81 (59%), Positives = 67/81 (82%)

Query: 1   MSILSSVCQTLRKGGIITYSTCTIFEEENFQVIEKFLENHPNFEQVELSHTQEDIVKRGC 60
           + IL+S  ++L+K GI+ YSTCTIF+EENF V+ +FLENHPNFEQVE+S+ + +++K GC
Sbjct: 342 LEILNSASKSLKKSGIMVYSTCTIFDEENFDVVHEFLENHPNFEQVEISNEKPEVIKEGC 401

Query: 61  ISISPEQYHTDGFFIGQVKRI 81
           + I+PE YHTDGFFI + K+I
Sbjct: 402 LFITPEMYHTDGFFIAKFKKI 422
```

There is also homology to SEQ ID 3018:

```
 Identities = 64/82 (78%), Positives = 74/82 (90%)

Query: 1   MSILSSVCQTLRKGGIITYSTCTIFEEENFQVIEKFLENHPNFEQVELSHTQEDIVKRGC 60
           + ILSSVCQTLRKGGIITYSTCTIF+EEN QVIE FL++HPNFEQV+L+HTQ DIVK G
Sbjct: 359 LEILSSVCQTLRKGGIITYSTCTIFDEENRQVIEAFLQSHPNFEQVKLNHTQADIVKDGY 418

Query: 61  ISISPEQYHTDGFFIGQVKRIL 82
           + I+PEQY TDGFFIGQV+R+L
Sbjct: 419 LIITPEQYQTDGFFIGQVRRVL 440
```

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

Example 2314

A DNA sequence (GBSx2455) was identified in *S.agalactiae* <SEQ ID 7117> which encodes the amino acid sequence <SEQ ID 7118>. This protein is predicted to be PTS permease for mannose subunit IIPMan. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.18   Transmembrane   32 -  48 (  30 -  58)
   INTEGRAL    Likelihood = -8.07   Transmembrane  127 - 143 ( 122 - 146)
   INTEGRAL    Likelihood = -2.07   Transmembrane   56 -  72 (  56 -  72)
   INTEGRAL    Likelihood = -1.44   Transmembrane   87 - 103 (  86 - 103)
   INTEGRAL    Likelihood = -0.53   Transmembrane  105 - 121 ( 105 - 121)

----- 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:AAF81084 GB:AF228498 AgaW [Escherichia coli]
 Identities = 38/122 (31%), Positives = 68/122 (55%), Gaps = 7/122 (5%)

Query: 25  KVPETKSIIRLTALAFLVCSILVVELVSMRELISSISFIGILVGSGPVNSFVHHIPQNLM 84
           ++P T  +  L A  +L        L+++       +F+ I  G+    + +  +PQ L+
Sbjct: 126 RMPRTPILAALNACNYLA-------LLALGNFYFLCAFLPIYFGAEHAKTIIDVLPQRLI 178

Query: 85  NGLSAAGGLLPAVGFAMLMKLLWINKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISS 144
           +GL  AGG++PA+GFA+L+K++  N    +++LGFV  A+LKLP +A+A     + +I
Sbjct: 179 DGLGVAGGIMPAIGFAVLLKIMMKNVYIPYFILGFVAAAWLKLPVLAIACPALAMALIDL 238

Query: 145 QR 146
```

```
              R
Sbjct: 239 LR 240
```

There is also homology to SEQ ID 1636:

```
Identities = 104/109 (95%), Positives = 108/109 (98%)

Query: 56  LISSISFIGILVGSGPVNSFVHHIPQNLMNGLSAAGGLLPAVGFAMLMKLLWTNKLAVFY 115
           +I+SISFIGILVGSGPVN+FV HIPQNLMNGLSAAGGLLPAVGFAMLMKLLWTNKLAVFY
Sbjct: 149 IIASISFIGILVGSGPVNAFVEHIPQNLMNGLSAAGGLLPAVGFAMLMKLLWTNKLAVFY 208

Query: 116 LLGFVLTAYLKLPAVAVAALGAVICVISSQRDIELDAITRGAISKQTTF 164
           LLGFVLTAYLKLPAVAVAALGAVICVISSQRD+ELDAITRGAISKQTTF
Sbjct: 209 LLGFVLTAYLKLPAVAVAALGAVICVISSQRDLELDAITRGAISKQTTF 257
```

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

Example 2315

A DNA sequence (GBSx2456) was identified in *S.agalactiae* <SEQ ID 7119> which encodes the amino acid sequence <SEQ ID 7120>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.12   Transmembrane  121 - 137 ( 118 - 144)
   INTEGRAL    Likelihood = -5.52   Transmembrane   91 - 107 (  89 - 111)
   INTEGRAL    Likelihood = -5.20   Transmembrane  166 - 182 ( 162 - 192)

----- Final Results -----
              bacterial membrane --- Certainty=0.4248(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:CAB15963 GB:Z99124 phosphotransferase system (PTS)
          beta-glucoside-specific enzyme IIABC component [Bacillus subtilis]
 Identities = 76/201 (37%), Positives = 122/201 (59%), Gaps = 3/201 (1%)

Query: 1   MIKALLALLLVFKILTPSSQTYILLNLFADGVFYFLPILIAITAAQKLKANPILALGTVV 60
           MIK L+AL + F  +   SQ +++L    DG FYFLP+L+A++AA+K  +NP +A
Sbjct: 121 MIKGLVALAVTFGWMAEKSQVHVILTAVGDGAFYFLPLLLAMSAARKFGSNPYVAAAIAA 180

Query: 61  MLLHPNWANLVASGKPVSLFHTIPFTLTNYASSVIPIILIICVQAYIEKYLKQIIPKSLR 120
            +LHP+   L+ +GKP+S F  +P T   Y+S+VIPI+L I + +Y+EK++ +    SL+
Sbjct: 181 AILHPDLTALLGAGKPIS-FIGLPVTAATYSSTVIPILLSIWIASYVEKWIDRFTHASLK 239

Query: 121 LVLVPMLIFLSMGILSFSILGPMGTIAGQYLAVIFTFLSKYASW-APAFLVGAFAPILIM 179
           L++VP    L +  L+   +GP+G I G+YL+     +L  +A   A  FL G F+ ++IM
Sbjct: 240 LIVVPTFTLLIVVPLTLITVGPLGAILGEYLSSGVNYLFDHAGLVAMIFLAGTFS-LIIM 298

Query: 180 FGVHSGIAALGITQLAKLGVD 200
           G+H     + I  +A+ G D
Sbjct: 299 TGMHYAFVPIMINNIAQNGHD 319
```

There is also homology to SEQ ID 2884.

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

Example 2316

A DNA sequence (GBSx2457) was identified in *S.agalactiae* <SEQ ID 7121> which encodes the amino acid sequence <SEQ ID 7122>. This protein is predicted to be glucose kinase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1180(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:CAB14416 GB:Z99116 glucose kinase [Bacillus subtilis]
 Identities = 32/57 (56%), Positives = 41/57 (71%)

Query: 1   MVIGGGVSAAGEFLRSRVEKYFVTFAFPQVKKSTKIKIAELGNDAGIIGAASLANQQ 57
           +V+GGGVS AGE LRS+VEK F   AFP+  ++  I IA LGNDAG+IG A +A  +
Sbjct: 258 IVLGGGVSRAGELLRSKVEKTFRKCAFPRAAQAADISIAALGNDAGVIGGAWIAKNE 314
```

There is also homology to SEQ ID 198. An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 50/56 (89%), Positives = 53/56 (94%)

Query: 1   MVIGGGVSAAGEFLRSRVEKYFVTFAFPQVKKSTKIKIAELGNDAGIIGAASLANQ 56
           +VIGGGVSAAGEFLRSR+EKYFVTF FPQV+ STKIKIAELGNDAGIIGAASLA Q
Sbjct: 264 VVIGGGVSAAGEFLRSRIEKYFVTFTFPQVRYSTKIKIAELGNDAGIIGAASLARQ 319
```

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

Example 2317

A DNA sequence (GBSx2458) was identified in *S.agalactiae* <SEQ ID 7123> which encodes the amino acid sequence <SEQ ID 7124>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have a cleavable N-term signal seq.

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

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

```
>GP:CAB14385 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 37/86 (43%), Positives = 51/86 (59%)

Query: 3   MSVILIIVILLAFVAWASWNYWRVRRAAKFLDNESFQKEMSRGQLIDIREAGAFHRKHIL 62
           MS +++++I  AF+ +   +Y   +R  K L  E F+     + QLID+RE    F   HIL
Sbjct: 1   MSNMIVLIIFPAFIIYMIASYVYQQRIMKTLTEEEFRAGYRKAQLIDVREPNEFEGGHIL 60

Query: 63  GARNIPASQFKVALSALRKDKPVLLY 88
           G RNIP SQ K   + +R DKPV LY
Sbjct: 61  GGRNIPLSQLKQRKNEIRTDKPVYLY 86
```

There is also homology to SEQ ID 202. An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 51/108 (47%), Positives = 70/108 (64%)
```

```
Query: 1   MDMSVILIIVILLAFVAWASWNYWRVRRAAKFLDNESFQKEMSRGQLIDIREAGAFHRKH 60
           M    +++ ++L+  V + +WNY+  R+ AK +DNE+F+  M +GQLID+RE  AF  KH
Sbjct: 1   MSPITLILWLLLVGIVGYYTWNYFSFRKMAKQVDNETFKDVMRQGQLIDLREPAAFRTKH 60

Query: 61  ILGARNIPASQFKVALSALRKDKPVLLYDASRGQSIPRIVLLLRKERF 108
           ILGARN PA QF  A+  LRKDKPVL+Y+  R Q     V  L+K  F
Sbjct: 61  ILGARNFPAQQFDAAIKGLRKDKPVLIYENMRPQYRVPAVKKLKKAGF 108
```

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

Example 2318

A DNA sequence (GBSx2459) was identified in *S.agalactiae* <SEQ ID 7125> which encodes the amino acid sequence <SEQ ID 7126>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Example 2319

A DNA sequence (GBSx2460) was identified in *S.agalactiae* <SEQ ID 7127> which encodes the amino acid sequence <SEQ ID 7128>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3522(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 2320

A DNA sequence (GBSx2461) was identified in *S.agalactiae* <SEQ ID 7129> which encodes the amino acid sequence <SEQ ID 7130>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2770(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:AAB18708 GB:U38906 ORF33 [Bacteriophage r1t]
 Identities = 56/85 (65%), Positives = 66/85 (76%), Gaps = 1/85 (1%)

Query: 1  MTNFATTDDVILLWRQLSVDEIKRAEALLETVSDTLRLEASKVGKNLDEMILETP-YFAT 59
          M  FAT DD+ +LWR L  DE +RAE LLE VSD+LR EA KVG++L  MI E P YFA+
Sbjct: 1  MNPFATVDDLTMLWRPLKGDEKERAEKLLEIVSDSLREEADKVGRDLYAMIAEKPSYFAS 60

Query: 60 VLKSVTVDIVARTLMTATQGEPMSQ 84
          V+KSVTVDIVARTLMT+T  EPM+Q
Sbjct: 61 VVKSVTVDIVARTLMTSTDQEPMTQ 85
```

There is also homology to SEQ ID 1432.

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

Example 2321

A DNA sequence (GBSx2462) was identified in *S.agalactiae* <SEQ ID 7131> which encodes the amino acid sequence <SEQ ID 7132>. This protein is predicted to be regulatory protein TypA (typA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2238(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:BAB06351 GB:AP001516 GTP-binding protein TypA/BipA (tyrosine
          phosphorylated protein A) [Bacillus halodurans]
 Identities = 175/237 (73%), Positives = 204/237 (85%), Gaps = 1/237 (0%)

Query: 1   MEDIFVGETVTPTDAIEPLPVLRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAE 60
           ME+I VGETV P D  +PLP+LRIDEPTLQMTFLVNNSPFAGREGK +TSRK+EERL AE
Sbjct: 281 MEEINVGETVCPVDHQDPLPILRIDEPTLQMTFLVNNSPFAGREGKHVTSRKLEERLRAE 340

Query: 61  LQTDVSLRVDPTDSPDKWTVSGRGELHLSILIETMRREGYELQVSRPEVIIKEIDGVQCE 120
           L+TDVSLRV+ TDSPD W VSGRGELHLSILIE MRREGYELQVS+PEVII+EIDGVQCE
Sbjct: 341 LETDVSLRVENTDSPDMWVVSGRGELHLSILIENMRREGYELQVSKPEVIIREIDGVQCE 400

Query: 121 PFERVQIDTPEEYQGAIIQSLSERKGDMLDMQMVGNGQTRLIFLIPARGLIGYSTEFLSM 180
           P ERVQID PEEY GA+++SL ERKG+ML+M   G+GQ RL F++PARGLIGY+TEFLS
Sbjct: 401 PVERVQIDVPEEYTGAVMESLGERKGEMLNMTNTGSGQVRLEFMVPARGLIGYTTEFLSQ 460

Query: 181 TRGYGIMNHTFDQYLPVVQGEIGGRHRGALVSIENGKATTYSIMRIEERGNLSFVNP 237
           TRGYGI+NH+FD Y PV  G++GGR +G LVS+E GKAT Y I+++E+RG + FV P
Sbjct: 461 TRGYGIINHSFDSYQPVTPGQVGGRRQGVLVSMETGKATQYGIIQVEDRGTI-FVEP 516
```

There is also homology to SEQ ID 206. An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 228/237 (96%), Positives = 233/237 (98%), Gaps = 1/237 (0%)

Query: 1   MEDIFVGETVTPTDAIEPLPVLRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAE 60
           MEDIFVGET+TPTD +E LP+LRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAE
Sbjct: 284 MEDIFVGETITPTDCVEALPILRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAE 343

Query: 61  LQTDVSLRVDPTDSPDKWTVSGRGELHLSILIETMRREGYELQVSRPEVIIKEIDGVQCE 120
```

```
            LQTDVSLRVDPTDSPDKWTVSGRGELHLSILIETMRREGYELQVSRPEVIIKEIDGV+CE
Sbjct: 344 LQTDVSLRVDPTDSPDKWTVSGRGELHLSILIETMRREGYELQVSRPEVIIKEIDGVKCE 403

Query: 121 PFERVQIDTPEEYQGAIIQSLSERKGDMLDMQMVGNGQTRLIFLIPARGLIGYSTEFLSM 180
            PFERVQIDTPEEYQGAIIQSLSERKGDMLDMQMVGNGQTRLIFLIPARGLIGYSTEFLSM
Sbjct: 404 PFERVQIDTPEEYQGAIIQSLSERKGDMLDMQMVGNGQTRLIFLIPARGLIGYSTEFLSM 463

Query: 181 TRGYGIMNHTFDQYLPVVQGEIGGRHRGALVSIENGKATTYSIMRIEERGNLSFVNP 237
            TRGYGIMNHTFDQYLPVVQGEIGGRHRGALVSIENGKATTYSIMRIEERG + FVNP
Sbjct: 464 TRGYGIMNHTFDQYLPVVQGEIGGRHRGALVSIENGKATTYSIMRIEERGTI-FVNP 519
```

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

Example 2322

A DNA sequence (GBSx2464) was identified in *S.agalactiae* <SEQ ID 7133> which encodes the amino acid sequence <SEQ ID 7134>. This protein is predicted to be pseudouridine synthase family 1 protein (rluB). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1950(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:CAB14248 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 59/105 (56%), Positives = 85/105 (80%)

Query: 5   VKERIYPVGRLDWDTTGLLILTNDGDFTDKMIHPRNEIDKVYLARVKGIATKENLRPLTR 64
            + +RIYP+GRLD+DT+GLL+LTNDG+F +K++HP+ EIDK Y+A+VKGI  KE LR L R
Sbjct: 91  IPQRIYPIGRLDYDTSGLLLLTNDGEFANKLMHPKYEIDKTYVAKVKGIPPKELLRKLER 150

Query: 65  GVVIDGKKTKPARYTIIKVDHEKNRSVVELTIHEGRNHQVKKMFE 109
            G+ ++  KT PA+  ++ +D +K  S+++LTIHEGRN QV++MFE
Sbjct: 151 GIRLEEGKTAPAKAKLLSLDKKKQTSIIQLTIHEGRNRQVRRMFE 195
```

There is also homology to SEQ ID 4728:

```
Identities = 96/109 (88%), Positives = 106/109 (97%)

Query: 1   MLPQVKERIYPVGRLDWDTTGLLILTNDGDFTDKMIHPRNEIDKVYLARVKGIATKENLR 60
            +LPQVKERIYPVGRLDWDT+G+LILTNDGDFTD MIHPRNEIDKVYLARVKGIATKENLR
Sbjct: 94  LLPQVKERIYPVGRLDWDTSGVLILTNDGDFTDTMIHPRNEIDKVYLARVKGIATKENLR 153

Query: 61  PLTRGVVIDGKKTKPARYTIIKVDHEKNRSVVELTIHEGRNHQVKKMFE 109
            PLTRG+VIDGKKTKPARY I++V+ +K+RS+VELTIHEGRNHQVKKMFE
Sbjct: 154 PLTRGIVIDGKKTKPARYNIVRVEADKSRSIVELTIHEGRNHQVKKMFE 202
```

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

Example 2323

A DNA sequence (GBSx2466) was identified in *S.agalactiae* <SEQ ID 7135> which encodes the amino acid sequence <SEQ ID 7136>. This protein is predicted to be L-ribulose 5-phosphate 4-epimerase. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2827(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:AAD45716 GB:AF160811 L-ribulose 5-phosphate 4-epimerase
          [Bacillus stearothermophilus]
 Identities = 68/103 (66%), Positives = 82/103 (79%)

Query: 2   QEMRERVCEANKSLPVHSLVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPENMVVTDLE 61
           +E+++ V EAN  LP + LV FTWGNVS +DRE GL+VIKPSGV YD+LT ++MVV DL
Sbjct: 3   EELKQAVLEANLQLPQYRLVTFTWGNVSGIDRERGLVVIKPSGVAYDKLTIDDMVVVDLT 62

Query: 62  GNIVEGDLNPSSDLPTHVQLYKAWPEVGGIVHTHSTEAVGWAQ 104
           GN+VEGDL PSSD PTH+ LYK +P +GGIVHTHST A  WAQ
Sbjct: 63  GNVVEGDLKPSSDTPTHLWLYKQFPGIGGIVHTHSTWATVWAQ 105
```

There is also homology to SEQ ID 4600:

```
Identities = 93/103 (90%), Positives = 96/103 (92%)

Query: 2   QEMRERVCEANKSLPVHSLVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPENMVVTDLE 61
           QEMRERVC ANKSLP H LVKFTWGNVSEV RE G IVIKPSGVDYD LTPENMVVTDL+
Sbjct: 6   QEMRERVCAANKSLPQHGLVKFTWGNVSEVCRELGRIVIKPSGVDYDLLTPENMVVTDLD 65

Query: 62  GNIVEGDLNPSSDLPTHVQLYKAWPEVGGIVHTHSTEAVGWAQ 104
           GN+VEGDLNPSSDLPTHV+LYKAWPEVGGIVHTHSTEAVGWAQ
Sbjct: 66  GNVVEGDLNPSSDLPTHVELYKAWPEVGGIVHTHSTEAVGWAQ 108
```

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

Example 2324

A DNA sequence (GBSx2467) was identified in *S.agalactiae* <SEQ ID 7137> which encodes the amino acid sequence <SEQ ID 7138>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3452(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:AAG05712 GB:AE004658 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 141/200 (70%), Positives = 162/200 (80%), Gaps = 1/200 (0%)

Query: 10  LSLGTDYETLANRFRPIFREISAGNVEREKARALPYEPIEWLKKAGFGAVRVPSEYGGAG 69
           LS G DYE LA RFRPIF  I+ G VERE+ R LP+E I WLK+AGFGAVRVP E+GGAG
Sbjct: 14  LSEGADYELLAQRFRPIFARIAEGAVERERQRELPHEAIAWLKQAGFGAVRVPREHGGAG 73

Query: 70  ASIGQLFQLLIELAEADSNIPQALRAHFAFVEDRLNAPPGVDRDTWFARFVAGDLVGNGW 129
           AS+ QL QLLIELAEADSNI QALR HFAFVEDRLNA PG  RD W  RFV GDLVG  W
Sbjct: 74  ASLPQLVQLLIELAEADSNITQALRGHFAFVEDRLNAEPGPGRDRWLRRFVEGDLVGCAW 133

Query: 130 TEVGTVKIGDVITKVSAQGDG-FVLNGTKFYSTGSIFADWIDVYAQRADNGADVIAVVNA 188
           TEVG+V++G+V+T+VS + DG +V+NG+K+YSTGS+F+DWID+YAQR D GADVIA +
```

```
Sbjct: 134 TEVGSVRLGEVLTRVSRKDDGRWVVNGSKYYSTGSLFSDWIDLYAQRDDTGADVIAAIRT 193

Query: 189 RHAGVRHSDDWDGFGQRTTG 208
               GVR SDDWDGFGQRTTG
Sbjct: 194 DQPGVRQSDDWDGFGQRTTG 213
```

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 2325

A DNA sequence (GBSx2468) was identified in *S.agalactiae* <SEQ ID 7139> which encodes the amino acid sequence <SEQ ID 7140>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1919(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 2326

A DNA sequence (GBSx2474) was identified in *S.agalactiae* <SEQ ID 7141> which encodes the amino acid sequence <SEQ ID 7142>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2978(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 2327

A DNA sequence (GBSx2476) was identified in *S.agalactiae* <SEQ ID 7143> which encodes the amino acid sequence <SEQ ID 7144>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5402(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 2328

A DNA sequence (GBSx2477) was identified in *S.agalactiae* <SEQ ID 7145> which encodes the amino acid sequence <SEQ ID 7146>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2755(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 = 190/247 (76%), Positives = 225/247 (90%)

Query: 1   MELGQLFHHLGSEITLMQRSERLLKEYDPEISESVEKALIEQGINLVKGATFERVEQSGE 60
           MELGQLFH+LGSE+TL+QRSERLLKEYDPEISESVEK+L+EQGINLVKGAT+ER+EQ+G+
Sbjct: 262 MELGQLFHNLGSEVTLIQRSERLLKEYDPEISESVEKSLVEQGINLVKGATYERIEQNGD 321

Query: 61  IKRVYVTVNGSREVIESDQLLVATGRKPNTDSLNLSAAGVETGKNNEILINDFGQTSNEK 120
           IK+V+V VNG + +IE+DQLLVATGR PNT +LNL AAGVE G   EI+I+D+ +T+N +
Sbjct: 322 IKKVHVEVNGKKRIIEADQLLVATGRTPNTATLNLRAAGVEIGSRGEIIIDDYSRTTNTR 381

Query: 121 IYAAGDVTLGPQFVYVAAYEGGIITDNAIGGLNKKIDLSVVPAVTFTNPTVATVGLTEEQ 180
           IYAAGDVTLGPQFVYVAAY+GG+   NAIGGLNKK++L VVP VTFT P +ATVGLTE+Q
Sbjct: 382 IYAAGDVTLGPQFVYVAAYQGGVAAPNAIGGLNKKLNLEVVPGVTFTAPAIATVGLTEQQ 441

Query: 181 AKEKGYDVKTSVLPLGAVPRAIVNRETTGVFKLVADAETLKVLGVHIVSENAGDVIYAAS 240
           AKE GY+VKTSVLPL AVPRA+VNRETTGVFKLVAD++T+KVLG H+V+ENAGDVIYAA+
Sbjct: 442 AKENGYEVKTSVLPLDAVPRALVNRETTGVFKLVADSKTMKVLGAHVVAENAGDVIYAAT 501

Query: 241 LAVKFGL 247
           LAVKFGL
Sbjct: 502 LAVKFGL 508
```

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 2329

A DNA sequence (GBSx2478) was identified in *S.agalactiae* <SEQ ID 7147> which encodes the amino acid sequence <SEQ ID 7148>. Analysis of this protein sequence reveals the following:

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

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

```
bacterial cytoplasm --- Certainty=0.3642(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 2330

A DNA sequence (GBSx2479) was identified in *S.agalactiae* <SEQ ID 7149> which encodes the amino acid sequence <SEQ ID 7150>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1936(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Example 2331

A DNA sequence (GBSx2480) was identified in *S.agalactiae* <SEQ ID 7151> which encodes the amino acid sequence <SEQ ID 7152>. This protein is predicted to be Nra. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1510(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 9383> which encodes amino acid sequence <SEQ ID 9384> 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 7153> which encodes the amino acid sequence <SEQ ID 7154>. Analysis of this protein sequence reveals the following:

```
     Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.64   Transmembrane   22 -  38 (  22 -  38)

----- 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>
```

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

```
Identities = 42/157 (26%), Positives = 78/157 (48%), Gaps = 2/157 (1%)

Query: 71  LLGREFIDSQHFKDINAYFLRHFICYCYYFIPDFYFLNTSRLSY--SKDLYHLLDKGLAD 128
           LLG   ++S  FK I   F R FI      +PD +  +  R     +K  Y+ L   + +
Sbjct: 8   LLGNNILNSLPFKRILVSFSRLFISNLQVLLPDIHLFHYLRRQQKRNKSFYNTLKTIVEE 67

Query: 129 IFNLKGGNLTFSKHETVLLTMQLSNLIETFLAPLSVYVISSSNIRLQTYQVMLNQYFTSK 188
             + +G       +  +L T+QL  L++T+L P+ VY+++++   L      L+ YF
Sbjct: 68  WMSAEGIVGKLPSYHLLLFTIQLEELLKTYLPPIPVYLLTNNTAALDLMTNALSIYFPPA 127

Query: 189 IAEFFFVNYQTTQIDEKLLKKADIIIAERRYISSLKN 225
           IA    VN +     + + +K  +IIA+R+Y++ +++
Sbjct: 128 IATVMPVNVEIIPFKDIVKEKQSVIIADRQYLNLIQH 164
```

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

Example 2332

A DNA sequence (GBSx2481) was identified in *S.agalactiae* <SEQ ID 7155> which encodes the amino acid sequence <SEQ ID 7156>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1383(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 2333

A DNA sequence (GBSx2482) was identified in *S.agalactiae* <SEQ ID 7157> which encodes the amino acid sequence <SEQ ID 7158>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4145(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 2334

A DNA sequence (GBSx2484) was identified in *S.agalactiae* <SEQ ID 7159> which encodes the amino acid sequence <SEQ ID 7160>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.02   Transmembrane   34 -  50 (  34 -  50)

----- Final Results -----
               bacterial membrane --- Certainty=0.1808(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 2335

A DNA sequence (GBSx2485) was identified in *S.agalactiae* <SEQ ID 7161> which encodes the amino acid sequence <SEQ ID 7162>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3488(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:CAB52002 GB:AL109663 hypothetical protein [Streptomyces
          coelicolor A3(2)]
 Identities = 61/141 (43%), Positives = 86/141 (60%), Gaps = 2/141 (1%)

Query: 3   TYFDNFLKTNQAYADLHGTAHLPIKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRNAG 62
            T  D  ++ N+ YA        +  +P  +VA+V CMD+RL +  ALGL LGD H +RNAG
Sbjct: 5   TVTDRLVEANERYAAAFADPGMDARPVQRVAVVACMDARLDLHAALGLKLGDCHTIRNAG 64

Query: 63  GRVTDDVLRSLVISQQQLGTREIVVLHHTDCGAQTFTNEAFAAQLQRDLGVDMHGHDFLP 122
            G VTDDV+RSL ISQ+ LGTR + ++HHT CG +T T E F   L+ ++G
Sbjct: 65  GVVTDDVIRSLTISQRALGTRSVALIHHTGCGMETITEE-FRHDLELEVG-QRPAWAVEA 122

Query: 123 FNDIEESVREDVAKLHASPFL 143
            F D ++ VR+ + ++  SPFL
Sbjct: 123 FRDADQDVRQSIERVRTSPFL 143
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6469> which encodes the amino acid sequence <SEQ ID 6470>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2295(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 = 109/146 (74%), Positives = 128/146 (87%)

Query: 1   MTTYFDNFLKTNQAYADLHGTAHLPIKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN 60
           + +YF++F+  NQAY  LHGTAHLP+KPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN
Sbjct: 1   LMSYFEHFMAANQAYVALHGTAHLPLKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN 60

Query: 61  AGGRVTDDVLRSLVISQQQLGTREIVVLHHTDCGAQTFTNEAFAAQLQRDLGVDMHGHDF 120
           AGGRVT+D++RSLVISQQQ+GTREIVVLHHTDCGAQTFTNE FA  +   LGVD+ G DF
Sbjct: 61  AGGRVTEDMIRSLVISQQQMGTREIVVLHHTDCGAQTFTNEGFAKHIHEHLGVDVSGQDF 120

Query: 121 LPFNDIEESVREDVAKLHASPFLREE 146
           LPF D+E+SVRED+AK+ AS  + ++
Sbjct: 121 LPFQDVEDSVREDMAKIRASSLISDD 146
```

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

Example 2336

A DNA sequence (GBSx2486) was identified in *S.agalactiae* <SEQ ID 7163> which encodes the amino acid sequence <SEQ ID 7164>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0932(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:AAG08811 GB:AE004955 phosphoribosylaminoimidazole carboxylase,
          catalytic subunit [Pseudomonas aeruginosa]
 Identities = 20/27 (74%), Positives = 26/27 (96%)

Query: 1  MFKHAEEARGRGIKIIIAGAGGAAHLP 27
          +F++AEEA GRG+++IIAGAGGAAHLP
Sbjct: 46 LFQYAEEAEGRGLEVIIAGAGGAAHLP 72
```

There is also homology to SEQ ID 910:

```
Identities = 27/27 (100%), Positives = 27/27 (100%)

Query: 1   MFKHAEEARGRGIKIIIAGAGGAAHLP 27
           MFKHAEEARGRGIKIIIAGAGGAAHLP
Sbjct: 87  MFKHAEEARGRGIKIIIAGAGGAAHLP 113
```

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

Example 2337

A DNA sequence (GBSx2488) was identified in *S.agalactiae* <SEQ ID 7165> which encodes the amino acid sequence <SEQ ID 7166>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.85   Transmembrane   58 -  74 (  53 -  80)
   INTEGRAL    Likelihood = -5.79   Transmembrane  103 - 119 ( 101 - 122)

----- 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>
```

There is also homology to SEQ IDs 880 and 9278.

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

Example 2338

A DNA sequence (GBSx2489) was identified in *S.agalactiae* <SEQ ID 7167> which encodes the amino acid sequence <SEQ ID 7168>. This protein is predicted to be short chain alcohol dehydrogenase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1742(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 9357> which encodes amino acid sequence <SEQ ID 9358> was also identified.

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

```
>GP:AAD06605 GB:AE001530 putative oxidoreductase [Helicobacter
          pylori J99]
 Identities = 68/94 (72%), Positives = 79/94 (83%)

Query: 4   IDLLVNNAGLALGLDKSYEADFGDWMTMINTNVVGLIYLTRCILPKMVEVNRGLIINLGS 63
           ID L+NNAGLALGL+K+YE +  DW  MI+TN+ GL++LTR ILP M+E ++G IINLGS
Sbjct: 76  IDALINNAGLALGLNKAYECELDDWEVMIDTNIKGLLHLTRLILPSMIEHDQGTIINLGS 135

Query: 64  XAGTIPYPGANVYGASKAFVKQFSLNLRADLAGT 97
            AGT  YPG NVYGASKAFVKQFSLNLRADLAGT
Sbjct: 136 IAGTYAYPGGNVYGASKAFVKQFSLNLRADLAGT 169
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7169> which encodes the amino acid sequence <SEQ ID 7170>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have an uncleavable N-term signal seq

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

A related sequence was also identified in GAS <SEQ ID 9121> which encodes the amino acid sequence <SEQ ID 9122>. Analysis of this protein sequence reveals the following:

```
     Possible site: 12
>>> 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 = 78/96 (81%), Positives = 87/96 (90%)

Query: 2   QSIDLLVNNAGLALGLDKSYEADFGDWMTMINTNVVGLIYLTRCILPKMVEVNRGLIINL 61
           Q I +LVNNAGLALGLDK+YEADF +WMTMINTN+VGLIYLTR +LP MV  + G+IINL
Sbjct: 82  QDITILVNNAGLALGLDKAYEADFENWMTMINTNIVGLIYLTRQLLPHMVSKDDGIIINL 141

Query: 62  GSXAGTIPYPGANVYGASKAFVKQFSLNLRADLAGT 97
           GS AGTIPYPGAN+YGASKAFVKQFSLNLRADLAG+
Sbjct: 142 GSTAGTIPYPGANIYGASKAFVKQFSLNLRADLAGS 177
```

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

Example 2339

A DNA sequence (GBSx2492) was identified in *S.agalactiae* <SEQ ID 7171> which encodes the amino acid sequence <SEQ ID 7172>. 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:CAC14663 GB:Y10855 mercuric reductase [Bacillus licheniformis]
 Identities = 68/104 (65%), Positives = 82/104 (78%)

Query: 1   MNKFKVNISGMTCTGCEKHVESALEKIGAKNIESSYRRGEAVFELPDDIEVESAIKAIDE 60
           M K++VN+ GMTCTGCE+HV  ALE +GAK IE  YRRGEAVFELP+ +EVE+A KAI E
Sbjct: 1   MKKYRVNVQGMTCTGCEEHVAVALENMGAKRIEVDYRRGEAVFELPNGLEVETAKKAIAE 60

Query: 61  ANYQAGEIEEVSSLENVALINEDNYDLLIIGSGAAAFSSAIKAI 104
           A YQ GE EEV S E + L +E +YD +IIGSG AAFSSAI+A+
Sbjct: 61  AKYQPGEAEEVQSQELIQLGDEGDYDYIIIGSGGAAFSSAIEAV 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 2340

A DNA sequence (GBSx2494) was identified in *S.agalactiae* <SEQ ID 7173> which encodes the amino acid sequence <SEQ ID 7174>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3341(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 2341

A DNA sequence (GBSx2495) was identified in *S.agalactiae* <SEQ ID 7175> which encodes the amino acid sequence <SEQ ID 7176>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4989(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 2342

A DNA sequence (GBSx2496) was identified in *S.agalactiae* <SEQ ID 7177> which encodes the amino acid sequence <SEQ ID 7178>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2569(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

Example 2343

A DNA sequence (GBSx2497) was identified in *S.agalactiae* <SEQ ID 7179> which encodes the amino acid sequence <SEQ ID 7180>. This protein is predicted to be DNA polymerase III alpha subunit (dnaE). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3124(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4095> which encodes the amino acid sequence <SEQ ID 4096>. Analysis of this protein sequence reveals the following:

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

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

```
bacterial cytoplasm --- Certainty=0.2600(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 = 186/237 (78%), Positives = 214/237 (89%)

Query: 10  DPVKHNLIFERFLNEERYSMPDIDIDLPDIYRGEFLRYVRNRYGSMHSAQIVTFSTFGAK 69
           DPV+H+L+FERFLN+ERYSMPDIDIDLPDIYR EFLRYVRNRYGS HSAQIVTFSTFG K
Sbjct: 321 DPVQHDLLFERFLNKERYSMPDIDIDLPDIYRSEFLRYVRNRYGSDHSAQIVTFSTFGPK 380

Query: 70  QAIRDVFKRFGASEYELTNITKKIHFRDNLTSVYNRNLAFRQIIDSKIEYQKAYDIAKRI 129
           QAIRDVFKRFG  EYELTN+TKKI F+D+L +VY ++++FRQ+I+S+ E+QKA+ IAKRI
Sbjct: 381 QAIRDVFKRFGVPEYELTNLTKKIGPKDSLATVYEKSISFRQVINSRTEFQKAFAIAKRI 440

Query: 130 EGNPRQTSIHAAGVVMSDDLLTDHIPLKNGEDMMITQYDASSVEDNGLLKMDFLGLRNLT 189
           EGNPRQTSIHAAG+VMSDD LT+HIPLK+G+DMMITQYDA +VE NGLLKMDFLGLRNLT
Sbjct: 441 EGNPRQTSIHAAGIVMSDDALTNHIPLKSGDDMMITQYDAHAVEANGLLKMDFLGLRNLT 500

Query: 190 FVQKMKEKVDKDYGISIQLETIDLEDKETLKLFAAGQTKGIFQFEQSGAINLLRRIR 246
           FVQKM+EKV KDYG  I +  IDLED +TL LFA G TKGIFQFEQ+GAINLL+RI+
Sbjct: 501 FVQKMQEKVAKDYGCQIDITAIDLEDPQTLALFAKGDTKGIFQFEQNGAINLLKRIK 557
```

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

Example 2344

A DNA sequence (GBSx2498) was identified in *S.agalactiae* <SEQ ID 7181> which encodes the amino acid sequence <SEQ ID 7182>. This protein is predicted to be a methylase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2121(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:AAG21729 GB:AF116907 putative methylase [Corynebacterium hoagii]
 Identities = 48/160 (30%), Positives = 85/160 (53%), Gaps = 6/160 (3%)

Query: 97   EPDDSENGHNDTDLEETDNQIPEEEVVETIPEIPVTDFYFPEDLTDFYPKTARDKVETNI 156
           EP+        +  E  +  + ++E       +P TDF    D+    P  A+ +V  NI
Sbjct: 1236 EPEAPTQPEAASAAETAEPAVEQQEPRAGPQSVPATDFALGTDV--HVPSGAKARVRANI 1293

Query: 157  VAIRLVKNLEVEHRNASPSEQELLAKYVGWGGLANEFFDD---YNPKFSKEREELKSLVT 213
            A RLV  L+ + R A+  EQ +LA++ GWG +  E FD+   +  +++ ER  L  L+
Sbjct: 1294 AAARLVLELDEQQRPATAEEQAVLAQWSGWGAVP-EVFDNRSKFLSEWADERAALLDLLG 1352

Query: 214  DKEYSDMKQSSLTAYYTDPSLIRQMWGIVERDGFTGWQIL 253
           +K +S  ++++L A+YTDP+++ ++W  V+R G     +L
Sbjct: 1353 EKGFSQARETTLNAHYTDPAIVGELWRAVQRAGLPDGALL 1392
```

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 2345

A DNA sequence (GBSx2499) was identified in *S.agalactiae* <SEQ ID 7183> which encodes the amino acid sequence <SEQ ID 7184>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1111(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 2346

A DNA sequence (GBSx2501) was identified in *S.agalactiae* <SEQ ID 7185> which encodes the amino acid sequence <SEQ ID 7186>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4752(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:CAA61516 GB:X89232 DNA-directed RNA polymerase [Pediococcus
          acidilactici]
 Identities = 48/53 (90%), Positives = 52/53 (97%)

Query: 5  KKPETINYRTLKPEREGLFDEVIFGPTKDWECACGKYKRIRYKGIICDRCGVE 57
          KKPETINYRTLKPE++GLFDE IFGPTKD+ECACGKYKRIRYKGI+CDRCGVE
Sbjct: 29 KKPETINYRTLKPEKDGLFDERIFGPTKDYECACGKYKRIRYKGIVCDRCGVE 81
```

There is also homology to SEQ ID 384.

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

Example 2347

A DNA sequence (GBSx2502) was identified in *S.agalactiae* <SEQ ID 7187> which encodes the amino acid sequence <SEQ ID 7188>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3080(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:AAC00282 GB:AF008220 YtlR [Bacillus subtilis]
 Identities = 61/216 (28%), Positives = 98/216 (45%), Gaps = 28/216 (12%)

Query: 8   IPCTYYPVGSGNDFARALKIPNL---------KETLTAIQTERLKEINCFIYDKGLIL-- 56
            I  ++ P G+ NDF+R   I  +         K  LT  +T  L  +N F+ DK  IL
Sbjct: 86  IELSFVPAGAYNDFSRGFSIKKIDLIQEIKKVKRPLT--RTFHLGSVN-FLQDKSQILYF 142

Query: 57  -NSLDLGFAAYVVWKASNSKIKNILNRYRLGKITYIVIAIKSLLHSSK------VQVLVE 109
             N + +GF AYV  KA     ++ +     RL  + Y +     S LH+S       +    E
Sbjct: 143 MNHIGIGFDAYVNKKAMEFPLRRVFLFLRLRFLVYPL----SHLHASATFKPFTLACTTE 198

Query: 110 GETGQQIKLNDLYFFALANNTYFGGGITIWPKASALTAELDMVYAKGHTFLKRLSILLSL 169
             ET +     +D++F  ++N+ ++GGG+   P A+      D+V  +   FLK+  +L  +
Sbjct: 199 DETRE---FHDVWFAVVSNHPFYGGGMKAAPLANPREKTFDIVIVENQPFLKKYWLLCLM 255

Query: 170 VFKRHTTSKSIKHQTFKAMTVYFPKNSLIEIDGEIV 205
             F +HT     +     K +T Y          DGEI+
Sbjct: 256 AFGKHTKMDGVTMFKAKDITFYTKDKIPFHADGEIM 291
```

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 2348

A DNA sequence (GBSx2503) was identified in *S.agalactiae* <SEQ ID 7189> which encodes the amino acid sequence <SEQ ID 7190>. This protein is predicted to be protease subunit HflC (hflC). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1809(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:AAG08326 GB:AE004907 protease subunit HflC [Pseudomonas aeruginosa]
 Identities = 182/202 (90%), Positives = 194/202 (95%)

Query: 1   MSQTERAVLLQFGKVVQTDVKPGLHVKVPYVNQVRKFDGRLLTLDAPTQRFLTLEKKAVM 60
            + QTERAV+L+FG+VV++DVKPGLH K+PYVNQVRKFD RLLTLDAPTQRFLTLEKKAVM
Sbjct: 26  VQQTERAVMLRFGRVVESDVKPGLHFKIPYVNQVRKFDARLLTLDAPTQRFLTLEKKAVM 85

Query: 61  VDAYAKWRVKDAERFYTATSGLKQIADERLSRRLESGLRDQFGKRTLHEVVSGERDALMA 120
            VDAYAKWRV DAERFYTATSGLKQIADERLSRRLE+GLRDQFGKRTLHEVVSGERDALM
Sbjct: 86  VDAYAKWRVADAERFYTATSGLKQIADERLSRRLEAGLRDQFGKRTLHEVVSGERDALMG 145

Query: 121 DITGSLNRMAEKELGIEVLDVRVKAIDLPKEVNRSVFERMSTERERREAREHRAKGNELGE 180
            DIT SLNRMA+KELGIEV+DVRVKAIDLPKEVNRSVFERMSTERERREAREHRAKG EL E
Sbjct: 146 DITASLNRMAQKELGIEVIDVRVKAIDLPKEVNRSVFERMSTERERREAREHRAKGRELAE 205

Query: 181 GIRADADRQRRVLLAEAYRESE 202
            GIRADADRQRRV++AEAYRESE
Sbjct: 206 GIRADADRQRRVIVAEAYRESE 227
```

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 2349

A DNA sequence (GBSx2504) was identified in *S.agalactiae* <SEQ ID 7191> which encodes the amino acid sequence <SEQ ID 7192>. 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 2350

A DNA sequence (GBSx2505) was identified in *S.agalactiae* <SEQ ID 7193> which encodes the amino acid sequence <SEQ ID 7194>. This protein is predicted to be ABC transporter (ATP-binding; daunorubicin resistance). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1846(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 = 88/231 (38%), Positives = 132/231 (57%), Gaps = 13/231 (5%)

Query: 10  QVIGYLPDVPKFYDYMTAQEYLQLC---AGLAQNKTSLPIADLLEQVGLADN-QQRISTY 65
           ++IGYLP  P FY +MTA E+L       +GL++ K    I ++LE VGL +   +RI  Y
Sbjct: 69  RLIGYLPQYPAFYSWMTANEFLTFAGRLSGLSKRKCQEKIGEMLEFVGLHEAAHKRIGGY 128

Query: 66  SRGMKQRLGLAQALIHXXKILICDEPTSALDPQGRQEILSIISQLRGQKTVIFSTHILSD 125
           S GMKQRLGLAQAL+H  K LI DEP SALDP GR E+L ++ +L+    V+FSTH+L D
Sbjct: 129 SGGMKQRLGLAQALLHKPKFLILDEPVSALDPTGRFEVLDMMRELKKHMAVLFSTHVLHD 188

Query: 126 VEKVCDQVLILTKSGIH---NLEDLRDKASASVNQLNLLIKVSDNEAQKLALRFPLNQKD 182
            E+VCDQV+I+    I     L++L+ +   +V  L++  K+      +K  +   + +
Sbjct: 189 AEQVCDQVVIMKNGEISWKGELQELKQQQQTNVFTLSVKEKLEGWLEEKPYVSAIVYKNP 248

Query: 183 QYYKVHLELSEANNREQALASFYRYLVEQEITPYFIELLEDSLEDFYLEVI 233
              +   EL + +     L+      + + +T    E   +SLED YL+V+
Sbjct: 249 S--QAVFELPDIHAGRSLLSD----CIRKGLTVTRFEQKTESLEDVYLKVV 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 2351

A DNA sequence (GBSx2506) was identified in *S.agalactiae* <SEQ ID 7195> which encodes the amino acid sequence <SEQ ID 7196>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0679(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with glycine-rich cell wall proteins (*e.g.* GB:AL161589 – the glycine-rich cell wall protein from *Arabidopsis thaliania*) and to SEQ ID 6882.

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

Example 2352

A DNA sequence (GBSx2507) was identified in *S.agalactiae* <SEQ ID 7197> which encodes the amino acid sequence <SEQ ID 7198>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2890(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 2353

A DNA sequence (GBSx2508) was identified in *S.agalactiae* <SEQ ID 7199> which encodes the amino acid sequence <SEQ ID 7200>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2410(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 9329> which encodes amino acid sequence <SEQ ID 9330> 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 9330 (GBS678) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 163 (lane 18; MW 53kDa), Figure 164 (lane 2 & 3; MW 53kDa) and Figure 188 (lane 7; MW 53kDa). Purified protein is shown in Figure 242, lanes 6 & 7.

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

Example 2354

A DNA sequence (GBSx2509) was identified in *S.agalactiae* <SEQ ID 7201> which encodes the amino acid sequence <SEQ ID 7202>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2025(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Example 2355

A DNA sequence (GBSx2510) was identified in *S.agalactiae* <SEQ ID 7203> which encodes the amino acid sequence <SEQ ID 7204>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Example 2356

A DNA sequence (GBSx2511) was identified in *S.agalactiae* <SEQ ID 7205> which encodes the amino acid sequence <SEQ ID 7206>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Example 2357

A DNA sequence (GBSx2512) was identified in *S.agalactiae* <SEQ ID 7207> which encodes the amino acid sequence <SEQ ID 7208>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0999(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 2358

A DNA sequence (GBSx2514) was identified in *S.agalactiae* <SEQ ID 7209> which encodes the amino acid sequence <SEQ ID 7210>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Example 2359

A DNA sequence (GBSx2515) was identified in *S.agalactiae* <SEQ ID 7211> which encodes the amino acid sequence <SEQ ID 7212>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2041(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 2360

A DNA sequence (GBSx2516) was identified in *S.agalactiae* <SEQ ID 7213> which encodes the amino acid sequence <SEQ ID 7214>. This protein is predicted to be 30S ribosomal protein S6 (rpsF). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3607(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 9423> which encodes amino acid sequence <SEQ ID 9424> was also identified.

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

```
>GP:CAB16128 GB:Z99124 ribosomal protein S6 (BS9) [Bacillus subtilis]
 Identities = 41/72 (56%), Positives = 58/72 (79%), Gaps = 1/72 (1%)

Query: 1  MVARFDSILSDNGATVVESKDWEKRRLAYEIQDFTEGLYHIVNVEAEDAVALNEFDRLSK 60
          ++ RF+++L+ NGA +  +KDW KRRLAYEI DF +G Y IVNV++ DA A+ EFDRL+K
Sbjct: 22 VIERFNNVLTSNGAEITGTKDWGKRRLAYEINDFRDGFYQIVNVQS-DAAAVQEFDRLAK 80

Query: 61 INGDILRHMIVK 72
          I+ DI+RH++VK
Sbjct: 81 ISDDIIRHIVVK 92
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7215> which encodes the amino acid sequence <SEQ ID 7216>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2720(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 = 66/74 (89%), Positives = 70/74 (94%)

Query: 1  MVARFDSILSDNGATVVESKDWEKRRLAYEIQDFTEGLYHIVNVEAEDAVALNEFDRLSK 60
          +VARFDSIL+DNGATVVESKDWEKRRLAYEI DF EGLYHIVN+EA DA ALNEFDRLSK
Sbjct: 22 LVARFDSILTDNGATVVESKDWEKRRLAYEINDFREGLYHIVNLEATDAAALNEFDRLSK 81

Query: 61 INGDILRHMIVKVD 74
          INGDILRHMIVK+D
Sbjct: 82 INGDILRHMIVKLD 95
```

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

Example 2361

A DNA sequence (GBSx2518) was identified in *S.agalactiae* <SEQ ID 7219> which encodes the amino acid sequence <SEQ ID 7220>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5289(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Example 2362

A DNA sequence (GASx1R) was identified in *S.pyogenes* <SEQ ID 7221> which encodes the amino acid sequence <SEQ ID 7222>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

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

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

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

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

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

Example 2363

A DNA sequence (GASx5R) was identified in *S.pyogenes* <SEQ ID 7223> which encodes the amino acid sequence <SEQ ID 7224>. 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>
```

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

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

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

Example 2364

A DNA sequence (GASx11) was identified in *S.pyogenes* <SEQ ID 7225> which encodes the amino acid sequence <SEQ ID 7226>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2614(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2365

A DNA sequence (GASx17) was identified in *S.pyogenes* <SEQ ID 7227> which encodes the amino acid sequence <SEQ ID 7228>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2849(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2366

A DNA sequence (GASx18) was identified in *S.pyogenes* <SEQ ID 7229> which encodes the amino acid sequence <SEQ ID 7230>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2099(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2367

A DNA sequence (GASx34) was identified in *S.pyogenes* <SEQ ID 7231> which encodes the amino acid sequence <SEQ ID 7232>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0801(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2368

A DNA sequence (GASx38) was identified in *S.pyogenes* <SEQ ID 7233> which encodes the amino acid sequence <SEQ ID 7234>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

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

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

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

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

```
>GP:CAB12617 GB:Z99108 similar to protein-tyrosine phosphatase
           [Bacillus subtilis]
 Identities = 57/155 (36%), Positives = 88/155 (56%), Gaps = 12/155 (7%)

Query: 1   MKKVCFVCLGNICRSPMAEFVMKSIVS----SDVMMIESRATSDWEHGNPIHSGTQSILK 56
           M  V FVCLGNICRSPMAE + + + +       +  +S     W  GNP H GTQ IL+
Sbjct: 1   MISVLFVCLGNICRSPMAEAIFRDLAAKKGLEGKIKADSAGIGGWHIGNPPHEGTQEILR 60

Query: 57  TYQINYDITKCSKQITITDFNTFDYIIGMDSDNVKNLKEMSQHQWDSKIYLFRE------ 110
              I++D    ++Q++  D + FDYII MD++N+ +L+ M+  +  S I    +
Sbjct: 61  REGISFD-GMLARQVSEQDLDDFDYIIAMDAENIGSLRSMAGFKNTSHIKRLLDYVEDSD 119

Query: 111 -GGVPDPWYTNDFEETYQLVRKGCQDWLSRLMSKE 144
              VPDP+YT +FEE  QL++ GC+  L+ +  ++
Sbjct: 120 LADVPDPYYTGNFEEVCQLIKTGCEQLLASIQKEK 154
```

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

Example 2369

A DNA sequence (GASx42R) was identified in *S.pyogenes* <SEQ ID 7235> which encodes the amino acid sequence <SEQ ID 7236>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4753(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2370

A DNA sequence (GASx47R) was identified in *S.pyogenes* <SEQ ID 7237> which encodes the amino acid sequence <SEQ ID 7238>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2014(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2371

A DNA sequence (GASx53R) was identified in *S.pyogenes* <SEQ ID 7239> which encodes the amino acid sequence <SEQ ID 7240>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.11   Transmembrane   56 -  72 (  56 -  72)

----- 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>
```

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

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

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

Example 2372

A DNA sequence (GASx67R) was identified in *S.pyogenes* <SEQ ID 7241> which encodes the amino acid sequence <SEQ ID 7242>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1610(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2373

A DNA sequence (GASx75) was identified in *S.pyogenes* <SEQ ID 7243> which encodes the amino acid sequence <SEQ ID 7244>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2803(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:CAA41942 GB:X59250 ribosomal protein B [Lactococcus lactis]
 Identities = 37/38 (97%), Positives = 37/38 (97%)

Query: 1  MKVRPSVKPICEYCKVIRRNGRVMVICPTNPKHKQRQG 38
          MKVRPSVKPICEYCKVIRRNGRVMVICP NPKHKQRQG
Sbjct: 1  MKVRPSVKPICEYCKVIRRNGRVMVICPANPKHKQRQG 38
```

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

Example 2374

A DNA sequence (GASx76) was identified in *S.pyogenes* <SEQ ID 7245> which encodes the amino acid sequence <SEQ ID 7246>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

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

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

```
bacterial cytoplasm --- Certainty=0.0824(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:AAB06824 GB:L47971 ribosomal protein S13 [Bacillus subtilis]
 Identities = 86/121 (71%), Positives = 103/121 (85%)

Query: 1   MARIAGVDIPNDKRVVISLTYVYGIGLATSKKILAAAGISEDIRVKDLTSDQEDAIRREV 60
            MARIAGVDIP DKRVVISLTY++GIG  T++++L  AG+SED RV+DLT ++   IR  +
Sbjct: 1   MARIAGVDIPRDKRVVISLTYIFGIGRTTAQQVLKEAGVSEDTRVRDLTEEELGKIRDII 60

Query: 61  DAIKVEGDLRREVNMNIKRLMEIGSYRGIRHRRGLPVRGQNTKNNARTRKGKAVAIAGKKK 121
            D +KVEGDLRREV++NIKRL+EIGSYRGIRHRRGLPVRGQN+KNNARTRKG    +A KKK
Sbjct: 61  DKLKVEGDLRREVSLNIKRLIEIGSYRGIRHRRGLPVRGQNSKNNARTRKGPRRTVANKKK 121
```

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

Example 2375

A DNA sequence (GASx81R) was identified in *S.pyogenes* <SEQ ID 7247> which encodes the amino acid sequence <SEQ ID 7248>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1842(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2376

A DNA sequence (GASx82) was identified in *S.pyogenes* <SEQ ID 7249> which encodes the amino acid sequence <SEQ ID 7250>. Analysis of this protein sequence reveals the following:

```
Possible site: 59

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3613(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2377

A DNA sequence (GASx83) was identified in *S.pyogenes* <SEQ ID 7251> which encodes the amino acid sequence <SEQ ID 7252>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1141(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2378

A DNA sequence (GASx85) was identified in *S.pyogenes* <SEQ ID 7253> which encodes the amino acid sequence <SEQ ID 7254>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2280(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2379

A DNA sequence (GASx89R) was identified in *S.pyogenes* <SEQ ID 7255> which encodes the amino acid sequence <SEQ ID 7256>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3040(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2380

A DNA sequence (GASx102) was identified in *S.pyogenes* <SEQ ID 7257> which encodes the amino acid sequence <SEQ ID 7258>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.75   Transmembrane   21 -  37 (  12 -  41)

----- Final Results -----
               bacterial membrane --- Certainty=0.6498(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:AAC45312 GB:U81957 ComYC [Streptococcus gordonii]
 Identities = 59/104 (56%), Positives = 85/104 (81%), Gaps = 1/104 (0%)

Query: 6   NNLRHKKLKGFTLLEMLLVILVISVLMLLFVPNLSKQKDRVTETGNAAVVKLVENQAELY 65
           N L+  ++K FTL+EML+V+L+ISVLMLLFVPNL+KQK+ V++TGNAAVVK+VE+QAELY
Sbjct: 2   NKLKKLRVKAFTLVEMLVVLLIISVLMLLFVPNLTKQKEAVSDTGNAAVVKVVESQAELY 61

Query: 66  EL-SQGSKPSLSQLKADGSITEKQEKAYQDYYDKHKNEKARLSN 108
           EL + G + +LS+L A G+I++KQ  +Y+ YY K+ +E   ++N
Sbjct: 62  ELKNTGDQATLSKLVAAGNISQKQADSYKAYYGKNNSETQAVAN 105
```

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

Example 2381

A DNA sequence (GASx103) was identified in *S.pyogenes* <SEQ ID 7259> which encodes the amino acid sequence <SEQ ID 7260>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

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

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

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

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

```
>GP:AAC23740 GB:AF052207 competence protein [Streptococcus pneumoniae]
 Identities = 52/131 (39%), Positives = 76/131 (57%)

Query: 8   IKAFTLLETLLSLSVMSFIILGLSVPVTKSYQKVEEHLFFSHFEHLYRHQQKLAILQQKQ 67
           IKAFT+LE+LL L ++S + LGLS  V  ++  VEE +FF  FE LYR  QK ++  Q++
Sbjct: 2   IKAFTMLESLLVLGLVSILALGLSGSVQSTFSAVEEQIFFMEFEELYRETQKRSVASQQK 61
```

```
Query: 68  RVLDISSTKIVTEGNSLTVPKSITVNHPYRLVIDQMGGNHSLAKIIFDMTDRRFKYQFYL 127
             L++    I      LTVPK I       +  D+ GGN SLAK+ F  +    +YQ YL
Sbjct: 62  TSLNLDGQMISNGSQKLTVPKGIQAPSGQSITFDRAGGNSSLAKVEFQTSKGAIRYQLYL 121

Query: 128 GSGNYQKTSQS 138
           G+G  ++  ++
Sbjct: 122 GNGKIKRIKET 132
```

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

Example 2382

A DNA sequence (GASx104) was identified in *S.pyogenes* <SEQ ID 7261> which encodes the amino acid sequence <SEQ ID 7262>. 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>
```

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

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

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

Example 2383

A DNA sequence (GASx109) was identified in *S.pyogenes* <SEQ ID 7265> which encodes the amino acid sequence <SEQ ID 7266>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.51   Transmembrane   37 -  53 (  28 -  58)
   INTEGRAL    Likelihood = -3.56   Transmembrane   61 -  77 (  60 -  77)

----- Final Results -----
              bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2384

A DNA sequence (GASx115R) was identified in *S.pyogenes* <SEQ ID 7267> which encodes the amino acid sequence <SEQ ID 7268>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
```

```
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.09   Transmembrane   20 -  36 (  13 -  40)

----- 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>
```

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

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

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

Example 2385

A DNA sequence (GASx124) was identified in *S.pyogenes* <SEQ ID 7269> which encodes the amino acid sequence <SEQ ID 7270>. Analysis of this protein sequence reveals the following:

```
Possible site: 52

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.17   Transmembrane   31 -  47 (  29 -  59)
   INTEGRAL    Likelihood = -5.63   Transmembrane  737 - 753 ( 734 - 756)

----- Final Results -----
               bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:AAC97148 GB:U49397 Cpa [Streptococcus pyogenes]
 Identities = 401/737 (54%), Positives = 517/737 (69%), Gaps = 25/737 (3%)

Query: 25  SKNSKR--FTVTLVGVFLMIFALVTSMVGAKTVFGLVESSTPNAINPDSSSEYRWYGYES 82
           S N+KR   T+ L+ VFL   AL+  +  +   FG  E S PN     S  +Y WYGY+S
Sbjct: 11  SANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPN--RQSSIQDYPWYGYDS 68

Query: 83  YVRGHPYYKQFRVAHDLRVNLEGSRSYQVYCFNLKKAFPLGSDSSVKKWYKKHDGISTKF 142
           Y +G+P Y   +  H+L+VNLEGS+ YQ YCFNL K FP  SDS    +WYKK +G +  F
Sbjct: 69  YPKGYPDYSPLKTYHNLKVNLEGSKDYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENF 128

Query: 143 EDYAMSPRITGDELNQKLRAVMYNGHPQNANGIMEGLEPLNAIRVTQEAVWYYSDNAPIS 202
             A  PRI   +L Q +  ++YNG+P N NGIM+G++PLNAI VTQ A+W Y+D+A I
Sbjct: 129 IKLADKPRIEDGQLQQNILRILYNGYPNNRNGIMKGIDPLNAILVTQNAIW-YTDSAQI- 186

Query: 203 NPDESFKRESESNLVSTSQLSLMRQALKQLIDPNLATKMPKQVPDDFQLSIFESEDKGDK 262
           NPDESFK E+ SN ++  QL LMR+ALK+LIDPNL +K   + P  ++L++FES D
Sbjct: 187 NPDESFKTEARSNGINDQQLGLMRKALKELIDPNLGSKYSNKTPSGYRLNVFESHD---- 242

Query: 263 YNKGYQNLLSGGLVPTKPPTPGDPPMPPNQPQTTSVLIRKYAIGDYSKLLEGATLQLTGD 322
             K +QNLLS   VP  PP PG+   PP + + TSV+IRKYA GD SKLLEGATL+L+
Sbjct: 243 --KPFQNLLSAEYVPDTPPKPGEE--PPAKTEKTSVIIRKYAEGD-SKLLEGATLKLSQI 297

Query: 323 NVNSFQARVFSSNDIGERIELSDGTYTLTELNSPAGYSIAEPITFKVEAGKVYTI-IDGK 381
             + FQ + F SN +GE +EL +GTYTLTE +SP GY IAEPI F+VE  KV+ +  DG
Sbjct: 298 EGSGFQEKDFQSNSLGETVELPNGTYTLTETSSPDGYKIAEPIKFRVENKKVFIVQKDGS 357

Query: 382 QIENPNKEIVEPYSVEAYNDFEEFSVLT-TQNYAKFYYAKNKNGSSQVVYCFNADLKSPP 440
           Q+ENPNKE+ EPYSVEAYNDF +  VL+    Y KFYYA NK+ SSQVVYCFNADL SPP
```

```
Sbjct: 358 QVENPNKEVAEPYSVEAYNDFMDEEVLSGFTPYGKFYYATNKDKSSQVVYCFNADLHSPP 417

Query: 441 DSEDGGKTMTPDFTT-GEVKYTHIAGRDLFKYTVKPRDTDPDTFLKHIKKVIEKGYREKG 499
           DS D G+T+ PD +T  EVKYTH AG DLFKY ++PRDT+P+ FLKHIKKVIEKGY++KG
Sbjct: 418 DSYDSGETINPDTSTMKEVKYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYKKKG 477

Query: 500 QAIEYSGLTETQLRAATQLAIYYFTDSAELDKDKL----KDYHGFGDMNDSTLAVAKILV 555
            +  Y+GLTETQ RAATQLAIYYFTDSA+L   K     K YHGF  M++ TLAV K L+
Sbjct: 478 DS--YNGLTETQFRAATQLAIYYFTDSADLKTLKTYNNGKGYHGFESMDEKTLAVTKELI 535

Query: 556 EYAQDSNPPQLTDLDFFIPNNNKYQSLIGTQWHPEDLVDIIRMEDKK-EVIPVTHNLTLR 614
            YAQ+ + PQLT+LDFF+PNN+K QSLIGT+ HP+DLVD+IRMEDKK EVIPVTH+LT++
Sbjct: 536 TYAQNGSAPQLTNLDFFVPNNSKDQSLIGTECHPDDLVDVIRMEDKKQEVIPVTHSLTVK 595

Query: 615 KTVTGLAGDRTKDFHFEIELKNNKQELLSQTVKTDKTNLEFKDGKATINLKHGESLTLQG 674
           KTV G  GD+TK F FE+ELK+   + +  T+KT+  +L  KDGK + NLKHG+++ ++G
Sbjct: 596 KTVVGELGDKTKGFQFELELKDKTGQPIVNTLKTNNQDLVAKDGKYSFNLKHGDTIRIEG 655

Query: 675 LPEGYSYLVKETDSEGYKVKVNSQEVANATVSKTGITSDETLAFENNKEPVVPTGVDQKI 734
           LP GYSY +KE +++ Y V V+++    A      IT D+ + FEN K+ V PTG+
Sbjct: 656 LPTGYSYTLKEAEAKDYIVTVDNKVSQEAQSVGKDITEDKKVTFENRKDLVPPTGLTTDG 715

Query: 735 NGYLALIVIAGISLGIW 751
             YL L+++  + L +W
Sbjct: 716 AIYLWLLLLVPLGLLVW 732
```

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

Example 2386

A DNA sequence (GASx125R) was identified in *S.pyogenes* <SEQ ID 7271> which encodes the amino acid sequence <SEQ ID 7272>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2604(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2387

A DNA sequence (GASx126) was identified in *S.pyogenes* <SEQ ID 7273> which encodes the amino acid sequence <SEQ ID 7274>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1537(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:AAC97149 GB:U49397 LepA [Streptococcus pyogenes]
 Identities = 59/132 (44%), Positives = 84/132 (62%), Gaps = 5/132 (3%)

Query: 1   MIIKRNDMAPSVKAGDAILFYRLSQTYKVEEAVVYEDSKTSITKVGRIIAQAGDEVDLTE 60
           MII  NDM+P++ AGD +L+YRL+    + + VVYE   T   KVGRI AQAGDEV+ T+
Sbjct: 42  MIINTNDMSPALSAGDGVLYYRLADRSHINDVVVYEVDNT--LKVGRIAAQAGDEVNFTQ 99

Query: 61  QGELKINGHIQNEG---LTFIKSREANYPYRIADNSYLILNDYYSQESENYLQDAIAKDA 117
           +G L INGH   +    LT+  S   N+PY++   +Y ILNDY  +  ++    A+  +
Sbjct: 100 EGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGTYFILNDYREERLDSRYYGALPINQ 159

Query: 118 IKGTINTLIRLR 129
           IKG I+TL+R+R
Sbjct: 160 IKGKISTLLRVR 171
```

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

Example 2388

A DNA sequence (GASx127) was identified in *S.pyogenes* <SEQ ID 7275> which encodes the amino acid sequence <SEQ ID 7276>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -3.93   Transmembrane  312 - 328 ( 311 - 337)

----- Final Results -----
              bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:AAC97152 GB:U49397 unknown [Streptococcus pyogenes]
 Identities = 125/355 (35%), Positives = 191/355 (53%), Gaps = 26/355 (7%)

Query: 1   MKLRHLLLLTGAALTSFA-----ATTVHGET--VVNGAKLTVTKNL-DLVNSNALIPNTDF 52
           MK   LLL  A L +       +  +  ET  V++G+ L V K      + N L+P  D+
Sbjct: 1   MKKNKLLLATAILATALGMASMSQNIKAETAGVIDGSTLVVKKTFPSYTDDNVLMPKADY 60

Query: 53  TFKIEPDTTVN---EDGNKFK-GVALNTPMTK-VTYTNSDKGGSNTKTAEFDFSEVTFEK 107
           +FK+E D        +DG   K GV      TK + Y+NSDK  +  K+  F+F+ V F
Sbjct: 61  SFKVEADDNAKGKTKDGLDIKPGVIDGLENTKTIRYSNSDKITAKEKSVNFEFANVKFPG 120

Query: 108 PGVYYYKVTEEKIDKVPGVSYDTTSYTVQVHVLWNEEQQKPVATYIVGYKEGS--KVPIQ 165
            GVY Y V E   +K  G++YD+  +TV V+V+ N+E       YIV  + G   K P+
Sbjct: 121 VGVYRYTVAEVNGNKA-GITYDSQQWTVDVYVV-NKEGGGFEVKYIVSTEVGQSEKKPVL 178

Query: 166 FKNSLDSTTLTVKKKVSGTGGDRSKDFNFGLTLKANQYYKASEKVMIEKTTKGGQAPVQT 225
           FKNS D+T+L ++K+V+G  G+  + F+F L L  N+ +   EK  +    +GG+
Sbjct: 179 FKNSFDTTSLKIEKQVTGNTGEHQRLFSFTLLLTPNECF---EKGQVVNILQGGETK--- 232

Query: 226 EASIDQLYHFTLKDGESIKVTNLPVGVDYVVTEDDYKSEKYTTNVEVSPQDGAVKNIAGN 285
           +  I + Y FTLKD  S+ ++ LPVG++Y +TE+D   + Y T+  +   + +     G
Sbjct: 233 KVVIGEEYSFTLKDKGSVTLSQLPVGIEYKLTEEDVTKDGYKTSATLKDGEQSSTYELGK 292

Query: 286 STEQETSTDKDMTITFTNKKDFEVPTGVAMTVAPYIALGIVAVGGALYFVKKKNA 340
             + + S D+   I  TNK+D +VPTGV  T+AP+  L IVA+GG +Y  K+K A
```

```
Sbjct: 293 DHKTDKSADE---IVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA 344
```

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

Example 2389

A DNA sequence (GASx128) was identified in *S.pyogenes* <SEQ ID 7277> which encodes the amino acid sequence <SEQ ID 7278>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

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

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

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

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

```
>GP:AAC97152 GB:U49397 unknown [Streptococcus pyogenes]
 Identities = 115/240 (47%), Positives = 178/240 (73%), Gaps = 3/240 (1%)

Query: 1   MIVRLIKLLDKLINVIVLCFFFLCLLIAALGIYDALTVYQGANATNYQQYKKKGVQ--FD 58
           M++ ++++++K I+ ++L F  + L +A  G++D+  +YQ A+A+N++++K    Q  F+
Sbjct: 351 MMMTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFE 410

Query: 59  DLLAINSDVMAWLTVKGTHIDYPIVQGENNLEYINKSVEGEYSLSGSVFLDYRNKVTFED 118
           DLLA+N DV+ WL + GTHIDYP+VQG+ NLEYINK+V+G  ++SGS+FLD RN   F D
Sbjct: 411 DLLALNEDVIGWLNIPGTHIDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTD 470

Query: 119 KYSLIYAHHMAGNVMFGELPNFRKKSFFNKHKEFSIETKTKQKLKINIFACIQTDAFDSL 178
            YSLIY HHMAGN MFGE+P F KK+FFNKH +  IETK ++KL + IFAC++TDAFD L
Sbjct: 471 DYSLIYGHHMAGNAMFGEIPKFLKKNFFNKHNKAIIETKERKKLTVTIFACLKTDAFDQL 530

Query: 179 LFNPIDV-DISSKNEFLNHIKQKSVQYREILTTNESRFVALSTCEDMTTDGRIIVIGQIE 237
           +FNP  + +   + + +++I ++S Q++ +   + ++FVA STCE+ +TD R+IV+G I+
Sbjct: 531 VFNPNAITNQDQQRQLVDYISKRSKQFKPVKLKHHTKFVAFSTCENFSTDNRVIVVGTIQ 590
```

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

Example 2390

A DNA sequence (GASx129) was identified in *S.pyogenes* <SEQ ID 7279> which encodes the amino acid sequence <SEQ ID 7280>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.05   Transmembrane    5 -  21 (   4 -  22)
   INTEGRAL    Likelihood = -5.04   Transmembrane  191 - 207 ( 186 - 209)

----- Final Results -----
             bacterial membrane --- Certainty=0.3421(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 181-186
```

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

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

```
>GP:AAC97151 GB:U49397 unknown [Streptococcus pyogenes]
 Identities = 64/213 (30%), Positives = 106/213 (49%), Gaps = 20/213 (9%)

Query: 1   MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI 60
           M+K    + ++  +L       +V A++ T   +I V N ++ A +         F   +
Sbjct: 1   MRKYWKMLFSVVMMLTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP------FSIAL 54

Query: 61  EALDKESPLPNSVTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV 120
           E++D    +      ++ G+GK SF  L F+ VGQY Y+++Q   +N  Y  D TV++V+
Sbjct: 55  ESIDAMKTIEE---ITIAGSGKASFSPLNFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL 111

Query: 121 IYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEPHQPDTTEKEKPQKKRNGI 180
           +YV Y+E  G L   ++S + G+ EKS + FK +   K   P QPD  +
Sbjct: 112 VYVTYDE-DGTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPDIPKTP--------- 161

Query: 181 LPSTGEMVSYVSALGIVLVATITLYSIYKKLKT 213
           LP  GE+ S +  L IVL+  + L  + KKLK+
Sbjct: 162 LPLAGEVKSLLGILSIVLLGLLVLLYV-KKLKS 193
```

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

Example 2391

A DNA sequence (GASx130R) was identified in *S.pyogenes* <SEQ ID 7281> which encodes the amino acid sequence <SEQ ID 7282>. Analysis of this protein sequence reveals the following:

```
Possible site: 57

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1614(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:CAB54046 GB:AJ245436 hypothetical protein, 57.8 kD [Pseudomonas
           putida]
 Identities = 128/388 (32%), Positives = 204/388 (51%), Gaps = 21/388 (5%)

Query: 4   IGSVVQRQELVFIPAQLKRINHVQHAYKCQTCSDNSLSDKIIKAPVPKAPLAHSLGSASI 63
           IG  V  Q L  +P Q++ I HV+  Y C+ C    ++      A  P   +  S+ S S+
Sbjct: 126 IGEEVSEQ-LEIVPMQIRVIKHVRKVYGCRDCESAPVT-----ADKPAQMIEKSMASPSV 179

Query: 64  IAHTVHQKFTLKVPNYRQEEDWNKLGLSISRKEIANWHIKSSQYYFEPLYDLLRDILLSQ 123
           +A  +  K+   +P +R E+   + G+ I R+ +A W I+ S++ F+PL +L+R+ LL+
Sbjct: 180 LAMLLTTKYVDGLPLHRFEKVLGRHGIDIPRQTLARWVIQCSEH-FQPLLNLMRESLLNS 238

Query: 124 EVIHADETSYRVLESD----TQLTYYWTFLSGKHEKKGITLYHHDKRRSGLVTQEVLGDY 179
            +IH DET  +VL+      +  ++ W    G  ++  + L+ +   R+  V   +L  Y
Sbjct: 239 RIIHCDETRVQVLKEPGREPSSQSWMWVQTGGPPDRP-VILFDYATSRAQEVPVRLLDGY 297

Query: 180 SGYVHCDMHGAYRQL---EHAKLVGCWAHVRRKFFEATPKQAD-KTSLGRKGLVYCDKLF 235
           GYV  D +  Y  L   +  + +GCWAH RRKF EA   Q   KT      L   +KL+
Sbjct: 298 RGYVMTDDYAGYNALAAQDGLERLGCWAHARRKFVEAQKVQPKGKTGRADIALNLINKLY 357

Query: 236 ALEAEWCELPPQERLVKRKEILTPLMTTFFDWCR--EQVVLSGSKLGLAIAYSLKHERTF 293
           +E +  +   ++R V R E   PL+T   +W    +  V + + LG AI Y   +
```

```
Sbjct: 358 GVERDLKDSDDEDRKVARMERSLPLLTQLKNWVEKTQPQVTTQNALGKAIGYLASNWSKL 417

Query: 294 RTVLEDGHIVLSNNMAERAIKSLVMGRKNWLFSQSFEGAKAAAIIMSLLETAKRHGLNSE 353
              +E G++ + NN AERAI+  V+GRKNWLFS + +GA A+A + SL+ETAK +G
Sbjct: 418 ERYVEHGYLPMDNNAAERAIRPFVIGRKNWLFSDTPKGATASAQLYSLVETAKANGQEPY 477

Query: 354 KYISYLLDRLPNEETLAKREVLEAYLPW 381
            ++ + L+RLP   ++   E  EA LPW
Sbjct: 478 AWLRHALERLPQACSV---EDYEALLPW 502
```

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

Example 2392

A DNA sequence (GASx131R) was identified in *S.pyogenes* <SEQ ID 7283> which encodes the amino acid sequence <SEQ ID 7284>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4465(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2393

A DNA sequence (GASx132R) was identified in *S.pyogenes* <SEQ ID 7285> which encodes the amino acid sequence <SEQ ID 7286>. Analysis of this protein sequence reveals the following:

```
Possible site: 46

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1529(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:BAA84885 GB:AB024946 orf50 [Escherichia coli]
 Identities = 37/91 (40%), Positives = 53/91 (57%)

Query: 10  QVYLVCGKTDMRQGIDSLAYLVKSQHELDLFSGAVYLFCGGRRDRFKALYWDGQGFWLLY 69
           +++LV G TDMR G + LA  V++  + D FSG +++F G R D+ K L+ D  G  L
Sbjct: 9   RIWLVAGITDMRNGFNGLASKVQNVLKDDPFSGHLFIFRGRRGDQIKVLWADSDGLCLFT 68

Query: 70  KRFENGKLAWPRNRDEVKCLTAVQVDWLMKG 100
           KR E G+  WP  RD     LT  Q+  L++G
Sbjct: 69  KRLERGRFVWPVTRDGKVHLTPAQLSMLLEG 99
```

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

Example 2394

A DNA sequence (GASx133R) was identified in *S.pyogenes* <SEQ ID 7287> which encodes the amino acid sequence <SEQ ID 7288>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1979(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2395

A DNA sequence (GASx135R) was identified in *S.pyogenes* <SEQ ID 7289> which encodes the amino acid sequence <SEQ ID 7290>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

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

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

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

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

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

Example 2396

A DNA sequence (GASx136) was identified in *S.pyogenes* <SEQ ID 7291> which encodes the amino acid sequence <SEQ ID 7292>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.73   Transmembrane  222 - 238 ( 212 - 242)
   INTEGRAL    Likelihood =-10.88   Transmembrane   37 -  53 (  32 -  57)
   INTEGRAL    Likelihood = -9.87   Transmembrane  462 - 478 ( 456 - 478)
   INTEGRAL    Likelihood = -4.25   Transmembrane  119 - 135 ( 117 - 137)
   INTEGRAL    Likelihood = -2.60   Transmembrane  308 - 324 ( 306 - 324)
   INTEGRAL    Likelihood = -1.28   Transmembrane  164 - 180 ( 164 - 180)
   INTEGRAL    Likelihood = -0.06   Transmembrane  137 - 153 ( 137 - 153)
   INTEGRAL    Likelihood = -0.06   Transmembrane  343 - 359 ( 343 - 359)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:BAB04077 GB:AP001508 short-chain fatty acids transporter
          [Bacillus halodurans]
 Identities = 158/465 (33%), Positives = 248/465 (52%), Gaps = 41/465 (8%)

Query: 15  IKTKKRFMDRYIDGFMKWMPESLFICFILTFLVVIMSVLMTDSPFIGTEKTGGIIYGWVN 74
           I    R M RY+       P+      +LTFLV  +S++ T+S    T  T  I+  W
Sbjct: 5   ISLSNRLMQRYL-------PDPFLFVVLLTFLVFALSLIFTES----TPLT--IVQYWGE 51

Query: 75  GFWGLLSFAMQMTILLATGNAVASSPPAHKMFKSLAKLPQTRTQIFIFSIVVGSIFGFLH 134
           GFWGLLSF+MQM ++L TG+ +ASSP   K   +LA LP +  Q  +   VV  +  F++
Sbjct: 52  GFWGLLSFSMQMVLVLVTGHVLASSPLFKKGLGALAGLPASPGQAILLVTVVSLVASFIN 111

Query: 135 WGLGMMVAIVFGKELLVQARQKGIKVHTPLFVATLFFTFLPATSGLSGAAVLYSATPDYL 194
           WG G+++  +F KEL     +K   V   L +A+ +  F+    GLSG+  L  ATPD+
Sbjct: 112 WGFGLVIGALFAKELA----KKVDNVDYRLLIASAYSGFMIWHGGLSGSVPLTIATPDHF 167

Query: 195 RNSVADAYKQVVPESVPLTESVL---NLPFISLLVVCMLVPLCFALLAHPKDETKIME-- 249
             +           +P +E++    NL  +  L +   +PL   L+   K +T  ++
Sbjct: 168 AQDMIGV--------IPTSETIFAPYNLAIVFALFIA--IPLANRLMMPGKSDTVTVDRS 217

Query: 250 -LDDEIYHHSLDTASHVVIARNTPAEKMNASRLVMYLVGGAIVSYSLYHFSVVGLSGLDL 308
            LDD       L  AS + +   TP++++  SR++  LVG   + +  Y+F+  G   L+L
Sbjct: 218 LLDDG---RDLQAAS-LELEAMTPSDRLENSRMISLLVGVLGLVFLGYYFATNGFE-LNL 272

Query: 309 NCFNFLFLGLGLLLCGQQGPEYYGSLFKDGVMSSWGLVLQFPFYAGIFGIIQSTGLGLEI 368
           +  N LFL LG+L  G    P+ +        V  + G+++QFPFYAG+ GI+ S+GL   +
Sbjct: 273 DIVNSLFLFLGILFHGT--PKLFLKAVTSAVKGASGIIIQFPFYAGLMGIMVSSGLATVM 330

Query: 369 SHFFVAISNGTTWPVFAYLYSALLNIAVPSGGSKFVIEAPYIVPATIEVGNDLGKILQAY 428
           S  FV+ SN  T+P+F +L + ++N++ VPSGG ++ ++AP ++ A   +G    K   A
Sbjct: 331 SEAFVSFSNEVTFPLFVFLSAGIVNVFVPSGGGQWAVQAPVVLEAAQSLGVPAAKAAMAV 390

Query: 429 QLGDATTNLIVPFWALSYLSNFKLKFNQIVAYTIPCVLVVTGIAI 473
             GDA TN+I PFWAL  L+   LK   I+ + +  +LVV+G+ I
Sbjct: 391 AWGDAWTNMIQPFWALPALAIAGLKAKDIMGFCV-MILVVSGVVI 434
```

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

Example 2397

A DNA sequence (GASx137R) was identified in *S.pyogenes* <SEQ ID 7293> which encodes the amino acid sequence <SEQ ID 7294>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2591(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:AAC22434 GB:U32761 transcriptional regulator [Haemophilus influenzae Rd]
 Identities = 37/107 (34%), Positives = 56/107 (51%), Gaps = 1/107 (0%)

Query: 21  LHRQNLVTFDKTFMINHQLTTLFEEANSLPVVKCYSASWDFLLNCTRYS-SYLTILPRPI 79
           LH+Q +  FD+TFMI+H L   FE  N  P +   S+ WDFLL+  + +   LTILP P+
Sbjct: 205 LHQQKMAIFDQTFMIHHHLKEAFERNNCYPDIVLDSSCWDFLLSAVKTNKELLTILPLPM 264

Query: 80  THFAHMDGLVEVQLTEHPKWEVVLASLKHNKTSHLKHYIKHTILDYF 126
              H    +  ++     W+V L   +    +HL+ YI   +L+ F
Sbjct: 265 AELYHSKEFLCRKIESPVPWKVTLCRQRKTVYTHLEEYIFDKLLEAF 311
```

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

Example 2398

A DNA sequence (GASx140) was identified in *S.pyogenes* <SEQ ID 7295> which encodes the amino acid sequence <SEQ ID 7296>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3351(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
!GB:U32761 acetate CoA-transferase, alpha subunit [H...    215  4e-55
 Identities = 105/213 (49%), Positives = 146/213 (68%)

Query: 22  ENKRIAIAEAISHIKDGDTIMVGGFMANGTPEALIDALVDKGTKDLTLICNDAGFVDRGV 81
           + K + + +A    +DG TIMVGGFM  GTP  L++AL++ G +DLTLI ND  FVD G+
Sbjct: 2   KTKLMTLQDATGFFRDGMTIMVGGFMGIGTPSRLVEALLESGVRDLTLIANDTAFVDTGI 61

Query: 82  GKMVANHQFKTIYATHIGLNKEAGRQMTAGETTIELIPQGTFAEKIRIGAYGIGGFYTPT 141
           G ++ N + + + A+HIG N E GR+M +GE  + L+PQGT  E+IR G  G+GGF TPT
Sbjct: 62  GPLIVNGRVRKVIASHIGTNPETGRRMISGEMDVVLVPQGTLIEQIRCGGAGLGGFLTPT 121

Query: 142 GVGTLVAEGKETKTIKGKTYLLEYPFEADVALIFANQADEMGNLQYSGSENNFNQLMAAC 201
           GVGT+V EGK+T T+ GKT+LLE P  AD+ALI A++ D +GNL Y  S  NFN L+A
Sbjct: 122 GVGTVVEEGKQTLTLDGKTWLLERPLRADLALIRAHRCDTLGNLTYQLSARNFNPLIALA 181

Query: 202 AKTTIVQAREIVPVGTIQPECVHTPHIFVDYIV 234
           A  T+V+  E+V  G +QP+ + TP   +D+I+
Sbjct: 182 ADITLVEPDELVETGELQPDHIVTPGAVIDHII 214
           subunit (EC 2.8.3.-). [Escherichia coli]
```

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

Example 2399

A DNA sequence (GASx141) was identified in *S.pyogenes* <SEQ ID 7297> which encodes the amino acid sequence <SEQ ID 7298>. Analysis of this protein sequence reveals the following:

```
Possible site: 41

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

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4941(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:AAF12248 GB:AE001862 CoA transferase, subunit B [Deinococcus radiodurans]
 Identities = 114/203 (56%), Positives = 158/203 (77%), Gaps = 3/203 (1%)

Query: 11  QNRIAKRVAKELEDGTLVNLGIGLPTKVANFVPEEMTVYFQSENGFIGLGP--KSDDPNS 68
           ++ +A R A+EL+DG  VNLGIGLPT VAN +P  M+V+ QSENG +G+GP    D+ +
Sbjct: 5   RDEMAARAAQELQDGYYVNLGIGLPTLVANHIPAGMSVWLQSENGLLGIGPFPTEDEVDP 64

Query: 69  TIVNAGGQPVTVYPGAAFFNSADSFGIIRGGHVDLTVLGALEIAENGDIANYLIPGKMVP 128
            ++NAG Q VT  PGA+FF+SADSF +IRGGHV+L +LGA++++E GD+AN++IPGKMV
Sbjct: 65  DLINAGKQTVTALPGASFFSSADSFAMIRGGHVNLAILGAMQVSETGDLANWMIPGKMVK 124

Query: 129 GMGGAMDLLVGAKKVIVAMEHTNKG-KHKLLKECTLPLTAKGVVDLIITEMGVFKVTPDG 187
           GMGGAMDL+ G ++V+V MEH  KG  HK+L+ECTLPLT +GVVD IIT++GV  VTP G
Sbjct: 125 GMGGAMDLVAGVQRVVVLMEHVAKGDAHKILRECTLPLTGQGVVDRIITDLGVLDVTPQG 184

Query: 188 IQVIEISEGFTFDEVQAATGVPL 210
           ++++E++ G T DE++  TG  +
Sbjct: 185 LKLVELAPGVTLDELRQKTGADI 207
```

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

Example 2400

A DNA sequence (GASx144) was identified in *S.pyogenes* <SEQ ID 7299> which encodes the amino acid sequence <SEQ ID 7300>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3227(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:BAA29948 GB:AP000003 137aa long hypothetical protein [Pyrococcus
           horikoshii]
 Identities = 49/113 (43%), Positives = 71/113 (62%), Gaps = 1/113 (0%)

Query: 5   PEPMGPYSTYTIEGHFLYTAGQLPLNPVTGQLSDG-FEAQCRQVFVNLQSILAEQKLDLN 63
           P+P+GPYS     G+FL+ AGQ+P++P TG++  G  + Q RQV  N+++IL      LN
Sbjct: 22  PKPIGPYSQAIKAGNFLFIAGQIPIDPKTGEIVKGDIKDQTRQVLENIKAILEAAGYSLN 81

Query: 64  HIYKLNVYLTDVTNVEILNHVMTDLFEEPYPVRTAVQVSALPLQALIEVEAVA 116
            + K+ VYL D+ +   +N V  + F E  P R AV+VS LP   LIE+EA+A
Sbjct: 82  DVIKVTVYLKDMNDFAKMNEVYAEYFGESKPARVAVEVSRLPKDVLIEIEAIA 134
```

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

Example 2401

A DNA sequence (GASx146) was identified in *S.pyogenes* <SEQ ID 7301> which encodes the amino acid sequence <SEQ ID 7302>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1238(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2402

A DNA sequence (GASx147) was identified in *S.pyogenes* <SEQ ID 7303> which encodes the amino acid sequence <SEQ ID 7304>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.46   Transmembrane  456 - 472 ( 452 - 481)
   INTEGRAL    Likelihood = -8.17   Transmembrane  603 - 619 ( 595 - 623)
   INTEGRAL    Likelihood = -6.85   Transmembrane  495 - 511 ( 491 - 518)
   INTEGRAL    Likelihood = -5.31   Transmembrane  420 - 436 ( 418 - 443)
   INTEGRAL    Likelihood = -4.99   Transmembrane  396 - 412 ( 392 - 413)
   INTEGRAL    Likelihood = -1.59   Transmembrane  522 - 538 ( 522 - 538)
   INTEGRAL    Likelihood = -0.64   Transmembrane  577 - 593 ( 577 - 593)
   INTEGRAL    Likelihood = -0.43   Transmembrane  377 - 393 ( 377 - 393)

----- Final Results -----
              bacterial membrane --- Certainty=0.5585(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:BAA04270 GB:D17462 Na+ -ATPase subunit I [Enterococcus hirae]
 Identities = 232/681 (34%), Positives = 370/681 (54%), Gaps = 40/681 (5%)

Query: 1   MAISQMKKLAMVFEKDYLDLVLKTLQQSQLVEVRDMKQLKH---WQDAFNKGNVKLPQIV 57
           MA+++M+K+ ++ +K   +++L+ +Q    VE+RD+ Q      W + F       P+++
Sbjct: 1   MAVTKMEKVTLISDKKNREILLQAVQGLHAVEIRDLFQESENNQWVETF----FPEPEMI 56

Query: 58  QYDLTHQKPLLDDEALQYLLQSQQELENGLASLSAFLPPIGKLTALRQ--KTPSLSFKQF 115
             D    K       L Y L         + +   F+   G+ +  +Q  K   LS
Sbjct: 57  DKDKELAK-------LSYKLTD-------IRTAIQFIEHHGEKSQKKQHLKRRELSLDTL 102

Query: 116 EERHRQQAAQTALKMMSQKIERLEQLQSKIDQLTEYCQELEKWRSLTVLPQDLAQFHFLS 175
           E+ + ++A    L+ +    E+ EQL  +  QL +    L  W++L + P+
Sbjct: 103 EKNYSEEAFSKKLEEVLLLKEQWEQLVDERQQLEDQENWLLNWQNLDLAPKAFDS-QMTK 161

Query: 176 ARVGTIPSTANNHFYHQLKQHKGLFIEEVYH----TEFEYGLVLFWQAQDTIHLQKYQFK 231
             +GT+ +    F  ++ +   ++EE+      T F Y ++   +++      +Y F
Sbjct: 162 LVIGTVNAKNAESFKAEVAEINEAYLEEINSSPTTTYFAYIVLRADESRMEEIASRYGFV 221
```

```
Query: 232 PLLYKEQLLPSEQLRINKELLTNWLAEKDSLLKELRQSQKILAQLQVEIDYVLSQYQRQQ 291
             Y  +  P +QL   K+ L     ++  L   +       +   +   +  L++ +R+
Sbjct: 222 KEDYLYEGTPQQQLVAAKQSLQEIKDQQKKLSSAIGACSGYIKDFEWTEEIFLARSEREA 281

Query: 292 TKKQLLGTRHLIALEGWIEADSVNQLKGLMTKTLGDMFYLDSYDVTPDDW--EDVPIKLR 349
             K +++ T +LI ++GW++ +   +L  ++   L       ++D   D+   E+VP KL+
Sbjct: 282 IKDRIIHTPYLILIQGWVDHEEKQELIHMLQNILASEEVYLTFDEPTDNEIAEEVPTKLK 341

Query: 350 NHRYIAPFELVTEMYALPKYQEKDPTPFLAPLYLTFFGMMVADLGYGLLLYAVTLAALVF 409
           NH  +APFE++TEMY+LPKY+E DPTP++ P YL FFGMMVAD+GYGLL++
Sbjct: 342 NHPIVAPFEMLTEMYSLPKYEEVDPTPWMMPFYLVFFGMMVADIGYGLLMFLGAFLLQKL 401

Query: 410 FNLQKTSKRLVTFFNILAISVAIWGLIYGSFFG---------FDLPVALLSTKTDVITIL 460
             L +  +R   FF ILAI   IWG IY SFFG           LP  +LST  DV TIL
Sbjct: 402 VVLPRGMQRFAKFFEILAIPSIIWGFIYSSFFGAALPKEIFGIHLPFPILSTTDDVNTIL 461

Query: 461 VVSLLFGFVTLIFGLLLGAWQQVRMKAYATAYTSSLAWTFILLGLLLFILGKNVSGLAYL 520
           ++S++FG + ++ GL + A + ++ KAY  A     AW +ILLG++L +LG
Sbjct: 462 ILSVIFGLIQILVGLFIAAKEHIKRKAYVDAVNDGFAWQWILLGIILILLGTMTLKNNAF 521

Query: 521 SVIGKWLALGNAFGILVVSLLKSKSLL-GLGSGLYNLYGISSYLSDLVSFTRLMALGLSG 579
             +G  LA+ +A  IL++ + +S S   G+  G YNLYG++ Y+ DLVS+TRLMALG+SG
Sbjct: 522 VYLGGALAVLSAVCILIIPVFQSSSKAKGIAKGAYNLYGLTGYIGDLVSYTRLMALGISG 581

Query: 580 ASIGAAFNMIVGIFPPVTRFTVGIFIFILLHAINIFLSMLSGYVHGARLIFVEFFGKFYE 639
            SI AAFNM+V   PP  RF+VGI + I+L A+N+FL++LS YVHGARL +VEFFGKFY
Sbjct: 582 GSIAAAFNMLVAFMPPAARFSVGILLIIVLQALNMFLTLLSAYVHGARLQYVEFFGKFYT 641

Query: 640 GGGKAFNPLKLADNYVNVNEE 660
           GGG++F PLK  + YVN+N +
Sbjct: 642 GGGRSFKPLKTVEKYVNINHK 662
```

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

Example 2403

A DNA sequence (GASx148) was identified in *S.pyogenes* <SEQ ID 7305> which encodes the amino acid sequence <SEQ ID 7306>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.80   Transmembrane   28 -  44 (  21 -  51)
   INTEGRAL    Likelihood = -6.85   Transmembrane  148 - 164 ( 146 - 170)
   INTEGRAL    Likelihood = -2.81   Transmembrane  105 - 121 ( 105 - 123)

----- 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>
```

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

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

```
>GP:BAA03841 GB:D16334 Na+-ATPase K subunit [Enterococcus hirae]
 Identities = 85/150 (56%), Positives = 107/150 (70%)

Query: 20  HYFTAHGGVFFAALGIVLAVALSGMGSAYGVGKGGQAAAALLKEEPEKFTSALILQLLPG 79
           +  T +GG+ FA L +  A   SG+GSA GVG  G+AAAAL   +PEKF  ALILQLLPG
Sbjct: 4   YLITQNGGMVFAVLAMATATIFSGIGSAKGVGMTGEAAAALTTSQPEKFGQALILQLLPG 63

Query: 80  SQGIYGFAIGILIWMKLTPELSVNQGLAYFLVSLPIAIVGYFSAKHQGNVSVAGMQILAK 139
           +QG+YGF I  LI++ L  ++SV QGL +   SLPIA  G FS   QG V+ AG+QILAK
```

```
Sbjct: 64  TQGLYGFVIAFLIFINLGSDMSVVQGLNFLGASLPIAFTGLFSGIAQGKVAAAGIQILAK 123

Query: 140 RPKDFMKGVILAAMVETYAILAFVVSFILL 169
           +P+   KG+I AAMVETYAIL FV+SF+L+
Sbjct: 124 KPEHATKGIIFAAMVETYAILGFVISFLLV 153
```

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

Example 2404

A DNA sequence (GASx149) was identified in *S.pyogenes* <SEQ ID 7307> which encodes the amino acid sequence <SEQ ID 7308>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4510(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:BAA04272 GB:D17462 Na+ -ATPase subunit E [Enterococcus hirae]
 Identities = 43/193 (22%), Positives = 95/193 (48%), Gaps = 2/193 (1%)

Query: 1   VNDITQLRQNVLEKAHQEGQQCLKIATDSLDTDFKERQQQGLHDLKAKRQKELKALEQQF 60
           V+ I ++   + E A E     ++    +D  F+ ++ Q   D + ++  +L+ +E+ +
Sbjct: 3   VDAIDKIITQINETAQLERASFEEMKRKEIDQKFEVKKWQIEADFQKEKASKLEEIERSY 62

Query: 61  QVAQQQLKNQERQALLALKQDSIKELFEASLEKMTNFSKEEELAFLKQVLSKYP-EQPLQ 119
           +  + + K Q +Q +L . KQ+ ++ LF  +  ++ N  KEE+LA +KQ++   P     +
Sbjct: 63  RQLRNKQKMQVKQEILNAKQEVLQRLFTEATLQLENEPKEEQLALMKQMIQTLPINGTAR 122

Query: 120 VTFGEKTGQKFSSYDCAELRLAFPQLSYNQELIPQ-EAGFLVSLDQVDDNYLYRYLLESV 178
           + GEK+    +    AE     P     ++  + +AG ++    +  N+L+ +L++ +
Sbjct: 123 LIPGEKSADILTPAVIAEWNEELPFELIREDFTEKAQAGLIIDDAGIQYNFLFSHLIKEI 182

Query: 179 LKEESSRIIDMLF 191
            +  S+ I   LF
Sbjct: 183 QETMSAEIAKELF 195
```

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

Example 2405

A DNA sequence (GASx150) was identified in *S.pyogenes* <SEQ ID 7309> which encodes the amino acid sequence <SEQ ID 7310>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3095(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:BAA04273 GB:D17462 Na+ -ATPase subunit C [Enterococcus hirae]
 Identities = 94/326 (28%), Positives = 167/326 (50%), Gaps = 5/326 (1%)

Query: 6   ELNTTISVKEKELLTKEQFDKLLQAPNTTTLARLLHQSVYHLTVDDLNDLDRLESILMAE 65
           ELN  I  +E EL++K+ F++++Q  +  +L  +L  ++Y   + D  D D  E+ L  E
Sbjct: 5   ELNPLIRGRELELISKDTFEQMIQTDSIDSLGEILQSTIYQPYIYDGFDKD-FEANLSQE 63

Query: 66  LTKTYRWAFAETPQPDIVQLFTLRYTYHNVKVLLKAKASQADLSHLLLPIGDKPLVALEH 125
            +K ++W     P+P+IV ++T+RYT+HN+KVL KA+ +  +L HL +  G   L  L+
Sbjct: 64  RSKLFQWLKESAPEPEIVWIYTMRYTFHNLKVLTKAEITGQNLDHLYIHDGFYSLEVLKD 123

Query: 126 LIRTMTSDEFPKEVVTEIQSIWAEYQDYQDIRVLEIGTDLAYFKALKQIAQRLEDPVFQQ 185
            I T  S E P  ++  I+ +    ++   ++ +++  D  +    +++ ++L  P   +
Sbjct: 124 AIHTQVSVELPDSLMDYIREVHEYCEESTILQGIDVIYDRCFLTEQRRLGEQLGYPELLE 183

Query: 186 AVLIVIDLYNLITVRRAKSQNKPISFMMQLLSDEASRPSKTFITLEDDKDLMTWFENVTP 245
            ++  IDL N+ T  R   Q++   FM  ++S   S P  T ++     ++ ++ + +
Sbjct: 184 EIIAFIDLTNITTTARGILQHRSAGFMTTVISSSGSIPKDTLLSFVRG-EMASFTQFLLT 242

Query: 246 DSYMTALKPYSEKLRQGTLQTTELEYLVDECLYHLFAKAKYQVDGPYVLARFLLAKSFEV 305
             Y   LK   + + +  +    LE L D+ L   +  A+ Q  GP  L  FL AK  E
Sbjct: 243 TDYSELLK---QVIHEEQIDLVSLEQLKDDYLSSFYQVAQTQAFGPLPLLAFLNAKEVES 299

Query: 306 KNLRLLAAALANDLPKERVIERMRPI 331
           KNLRLL     N    E++ ERMR +
Sbjct: 300 KNLRLLIIGKRNHFSLEQLKERMRQV 325
```

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

Example 2406

A DNA sequence (GASx151) was identified in *S.pyogenes* <SEQ ID 7311> which encodes the amino acid sequence <SEQ ID 7312>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0484(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:BAA04274 GB:D17462 Na+ -ATPase subunit G [Enterococcus hirae]
 Identities = 45/101 (44%), Positives = 65/101 (63%)

Query: 6   YKVGVIGNRDVILPFQMIGFQTFPVIKPQDAINQLRQLAMEDFGIIYITEDIAAAIPEAL 65
           YK+GV+G++D + PF++ GF         +    + ++A  ++G+IYITE  A  +PE +
Sbjct: 3   YKIGVVGDKDSVSPFRLFGFDVQHGTTKTEIRKTIDEMAKNEYGVIYITEQCANLVPETI 62

Query: 66  THYDNQVLPAVIPLPTHQGAQGIGLSRIQAMVEKAVGQNIL 106
             Y  Q+ PA+I +P+HQG  GIGL  IQ  VEKAVGQNIL
Sbjct: 63  ERYKGQLTPAIILIPSHQGTLGIGLEEIQNSVEKAVGQNIL 103
```

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

Example 2407

A DNA sequence (GASx152R) was identified in *S.pyogenes* <SEQ ID 7313> which encodes the amino acid sequence <SEQ ID 7314>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1048(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2408

A DNA sequence (GASx156) was identified in *S.pyogenes* <SEQ ID 7315> which encodes the amino acid sequence <SEQ ID 7316>:

```
EYSIIPQLKETIHYIELKLEEAERASLVRIMKITS
```

Analysis of this protein sequence reveals the following:

```
Possible site: 16

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5026(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:BAA04277 GB:D17462 Na+ -ATPase subunit D [Enterococcus hirae]
 Identities = 119/201 (59%), Positives = 151/201 (74%), Gaps = 2/201 (0%)

Query: 10  RLNVKPTRMELSNLKNRLKTATRGHKLLKDKRDELMRRFVDLIRENNELRQTIEKELAAN 69
           RLNV PTRMEL+ LK +L TATRGHKLLKDK+DELMR+F+ LIR+NNELRQ IEKE
Sbjct: 2   RLNVNPTRMELTRLKKQLTTATRGHKLLKDKQDELMRQFILLIRKNNELRQAIEKETQTA 61

Query: 70  MKEFVLAKASENSLMVEELFAVPVHEVTLWIDIENIMSVNVPKFHVQSNTAREQEQGEFA 129
           MK+FVLAK++     ++EL A+P   V++ +  +NIMSV VP  + Q +    +   E
Sbjct: 62  MKDFVLAKSTVEEAFIDELLALPAENVSISVVEKNIMSVKVPLMNFQYDETLNETPLE-- 119

Query: 130 YSYLSSNSEMDNTIQKTKELLEKLLRLAEVEKTCQLMADDIEKTRRRVNGLEYSIIPQLK 189
           Y YL SN+E+D +I    +LL KLL+LAEVEKTCQLMA++IEKTRRRVN LEY  IPQL+
Sbjct: 120 YGYLHSNAELDRSIDGFTQLLPKLLKLAEVEKTCQLMAEEIEKTRRRVNALEYMTIPQLE 179

Query: 190 ETIHYIELKLEEAERASLVRI 210
           ETI+YI++KLEE ERA + R+
```

```
Sbjct: 180 ETIYYIKMKLEENERAEVTRL 200
```

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

Example 2409

A DNA sequence (GASx161R) was identified in *S.pyogenes* <SEQ ID 7317> which encodes the amino acid sequence <SEQ ID 7318>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

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

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

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

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

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

Example 2410

A DNA sequence (GASx164) was identified in *S.pyogenes* <SEQ ID 7319> which encodes the amino acid sequence <SEQ ID 7320>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.06   Transmembrane    9 -  25 (   9 -  25)

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

A related sequence was also identified <SEQ ID 9091> which encodes the amino acid sequence <SEQ ID 9092>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 33
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty= 0.300(Affirmative) < succ>
               bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

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

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

Example 2411

A DNA sequence (GASx165) was identified in *S.pyogenes* <SEQ ID 7321> which encodes the amino acid sequence <SEQ ID 7322>. Analysis of this protein sequence reveals the following:

```
Possible site: 59

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2251(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2412

A DNA sequence (GASx166) was identified in *S.pyogenes* <SEQ ID 7323> which encodes the amino acid sequence <SEQ ID 7324>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

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

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

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

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

Example 2413

A DNA sequence (GASx167) was identified in *S.pyogenes* <SEQ ID 7325> which encodes the amino acid sequence <SEQ ID 7326>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

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

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

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

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

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

Example 2414

A DNA sequence (GASx168R) was identified in *S.pyogenes* <SEQ ID 7327> which encodes the amino acid sequence <SEQ ID 7328>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

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

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

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

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

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

Example 2415

A DNA sequence (GASx169R) was identified in *S.pyogenes* <SEQ ID 7329> which encodes the amino acid sequence <SEQ ID 7330>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

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

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

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

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

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

Example 2416

A DNA sequence (GASx170) was identified in *S.pyogenes* <SEQ ID 7331> which encodes the amino acid sequence <SEQ ID 7332>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.34   Transmembrane  154 - 170 ( 153 - 170)
   INTEGRAL    Likelihood = -1.12   Transmembrane   20 -  36 (  19 -  36)
   INTEGRAL    Likelihood = -0.69   Transmembrane   52 -  68 (  52 -  68)
   INTEGRAL    Likelihood = -0.53   Transmembrane  399 - 415 ( 399 - 415)

----- 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>
```

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

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

```
>GP:BAB05347 GB:AP001512 cystathionine beta-lyase [Bacillus halodurans]
 Identities = 200/384 (52%), Positives = 262/384 (68%), Gaps = 3/384 (0%)

Query: 79  IAEVYEMRENTTLLHGYTVIDEFTGAASVPIYQTSTFHNSELYCPSQKHLYTRFSNPTTE 138
           ++E Y ++  T LLH    +D+ TGA SVPI   STFH  + +     + Y+R  NPT +
Sbjct: 1   MSEQYSLQ--TKLLHNEHKVDQATGAVSVPIQHASTFHQFD-FDTFGTYDYSRSGNPTRD 57

Query: 139 ALEDGLACLEKATYAVAYASGMAAISTVLMLLKAGDHVIFPLEVYGGTCQFATAILPNYQ 198
           ALE  +A LE   +  A+ASGMAAIST  MLL  GDHV+   +VYGGT +  T +L
Sbjct: 58  ALEAAIAELEGGNHGFAFASGMAAISTAFMLLSKGDHVVLTKDVYGGTFRLVTEVLTRLG 117

Query: 199 IETSFVDMADLATVKASIRPNTRMIYLETPSNPLLKICDISELVQLAKAYGVLTVADNTF 258
           IE +FVDM +LA V A+IRPNTR++Y+ETPSNP L I DI  +V LAK +  LT  DNTF
Sbjct: 118 IEHTFVDMTNLAEVAAAIRPNTRVLYMETPSNPTLNITDIRGVVSLAKEHECLTFLDNTF 177

Query: 259 MTSLYQEPLAMGVDIVVESVTKFINGHSDVVAGLAATNNEAIYNQLKLFQKNFGAIVGVE 318
           +T   Q PL +GVD+V+ S TKFI GHSDVVAGLA T NE +  +L   Q +FGAI+GV+
Sbjct: 178 LTPALQRPLELGVDVVLHSATKFIGGHSDVVAGLAVTKNEELGKKLAFLQNSFGAILGVQ 237

Query: 319 DAWLILRGMKTMGIRMEQAVKNAQQLANYLAKHPKVLKVHYPGLDSHPNHDTHLQQAKNG 378
           D WL+LRG+KT+ +RME   K AQQ+A +L   P+V +V+YPGL  HP H+   +QA+
Sbjct: 238 DVWLVLRGLKTLHVRMEHGEKGAQQIAEWLQGVPEVKRVYYPGLKDHPGHELQKRQAEGF 297

Query: 379 GAVLSFELASKEELMTFTHRIQLPILAVSLGGVESILSHPATMSHACLSPQARLEQGVVD 438
           GAVLSFEL ++E +  F   ++LP+ AVSLG VESILS+PA MSHA +  + R  +G+ D
Sbjct: 298 GAVLSFELENEEAVRRFVEHVKLPVFAVSLGAVESILSYPAKMSHAAMPKEEREARGIRD 357

Query: 439 GLLRLSCGVENIEDLLADFEQALA 462
           GLLRLS G+E  E+L+ADF+ A A
Sbjct: 358 GLLRLSVGLEKPEELMADFKAAFA 381
```

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

Example 2417

A DNA sequence (GASx178) was identified in *S.pyogenes* <SEQ ID 7333> which encodes the amino acid sequence <SEQ ID 7334>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1492(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2418

A DNA sequence (GASx182) was identified in *S.pyogenes* <SEQ ID 7335> which encodes the amino acid sequence <SEQ ID 7336>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
```

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2584(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2419

A DNA sequence (GASx187) was identified in *S.pyogenes* <SEQ ID 7337> which encodes the amino acid sequence <SEQ ID 7338>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2084(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

Example 2420

A DNA sequence (GASx188) was identified in *S.pyogenes* <SEQ ID 7339> which encodes the amino acid sequence <SEQ ID 7340>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2060(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:AAG05515 GB:AE004640 conserved hypothetical protein [Pseudomonas aeruginosa]
 Identities = 140/442 (31%), Positives = 208/442 (46%), Gaps = 73/442 (16%)

Query: 2   KKYLNQNVYDALIERLHFLFNDFPIVYISFSGGKDSGLLLNILLDFRDKYYPDREIG--- 58
           K Y + +V+ A + RL  +F +F  V ++FSGGKDS + L + LD        RE+G
Sbjct: 4   KHYQDADVHAATLSRLRLVFRNFERVCVAFSGGKDSSVTLQLALDVA------RELGRSP 57

Query: 59  --VFHQDFEAQYSLTTKYVQETFTSLEGRKKVSLYWVCLPMATRTALSSYEMFWYPWDDK 116
             V   D E QY  T  +V E      GR  V  +WVCLP+  R A S  E +W  W+
Sbjct: 58  VDVLFIDLEGQYQATIDHVSEML----GRPDVRPWWVCLPLNLRNASSLEEPYWCCWEPG 113

Query: 117 TEDIWVRPMPSQDYVINLENNSITTYRYKMNQEDLAKQFGRWYKQIHGNQKTVCILGNRA 176
```

```
            E  WVRP+P Q  VI+ +      YRY+M  E+    F  W  +    + T  ++G R+
Sbjct: 114 AEADWVRPLPKQRGVIS-DPAFFPFYRYRMEFEEFVAGFNAWLAR---EEPTAFLVGIRS 169

Query: 177 SESLHRYSGFINKKYGYQKEC----------WITKQFKDVWTAS--PLYDWSVEDIWH 222
            ESL+RY     K+    K+C            W  +   +    S  P+YDW  ED+W
Sbjct: 170 DESLNRYLAV--KRRSRAKQCAWTPPGGSAPLAWSARDRANPQAVSFFPIYDWRFEDLWR 227

Query: 223 AYYKFSYSYNELYDLFYKAGLKPSQMRVASPFQDYAVDSLNLYRIIDQETWVKLLGRVQG 282
                Y+YN LYD  Y+AG+  SQMR+  P+ D     L+L+  I+  TW K++ RV G
Sbjct: 228 CVADHGYAYNRLYDQMYRAGVPFSQMRICQPYGDDQRKGLDLFHRIEPRTWFKVVRRVAG 287

Query: 283 VNFSNIYGRTKAMGYK-SIALPKGH-SWKSYTQFLLSTLPVRLRNNYVRKFNKSIDFWHK 340
            N+   Y R + +GY+  + LP    +W+ Y+QFLL ++P  LR  Y R+  + I +W +
Sbjct: 288 ANYGARYCRQRFLGYRGGLGLPPSFGTWREYSQFLLRSMPPPLRGIYQRRIERFILWWKQ 347

Query: 341 TGGGLAEETINELIEKGYRIARNGISNYTSFKHSRVIFLDQ-IPDDTDDIVTTKDIPSWK 399
               LA                              I+ D  IP     +  + PSW+
Sbjct: 348 HDYPLA------------------------------IWPDAGIP----ALENRRKQPSWR 373

Query: 400 RMCFCILKNDHICRTMGFGLTR 421
            R+   +LK D + R++ FG ++
Sbjct: 374 RIALSLLKQD-MARSLSFGFSQ 394
```

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

Example 2421

A DNA sequence (GASx189) was identified in *S.pyogenes* <SEQ ID 7341> which encodes the amino acid sequence <SEQ ID 7342>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

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

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

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

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

```
>GP:AAC73702 GB:AE000165 orf, hypothetical protein [Escherichia
          coli]
 Identities = 79/162 (48%), Positives = 110/162 (67%), Gaps = 1/162 (0%)

Query: 7   PVYEIKSIPIEKISPNDYNPNSVAPPEMKLLYDSIKSDGYTMPIVCYYDKEEDRYSIVDG 66
           PV  +  +   ++ PNDYNPN+VAPPE KLL  SI+ DG+T PIV  +  +++   IVDG
Sbjct: 46  PVDCVLWVKNSQLMPNDYNPNNVAPPEKKLLQKSIEIDGFTQPIVVTHT-DKNAMEIVDG 104

Query: 67  FHRYRIMLDYSDIYERESGRLPVSVIDKSLDYRMASTIRHNRARGSHDVDLMSQIVKDLH 126
           FHR+ I    S +  R  G LPV+ ++ + + R+A+TIRHNRARG H +  MS+IV++L
Sbjct: 105 FHRHEIGKGSSSLKLRLKGYLPVTCLEGTRNQRIAATIRHNRARGRHQITAMSEIVRELS 164

Query: 127 ECGRSDNWIAKHLGMDKDEILRLKQITGLASLFKDHEFNQSW 168
           + G  DN I K LGMD DE+LRLKQI  L  LF D +++++W
Sbjct: 165 QLGWDDNKIGKELGMDSDEVLRLKQINGLQELFADRQYSRAW 206
```

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

Example 2422

A repeated DNA sequence (GASx192R) was identified in *S.pyogenes* <SEQ ID 7343> which encodes the amino acid sequence <SEQ ID 7344>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4301(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:CAA63509 GB:X92946 transposase [Lactococcus lactis]
 Identities = 23/36 (63%), Positives = 28/36 (76%)

Query: 1   MQDKLVTEAFNQAYNREKPKEGVIVHTDQGSQYTGA 36
           MQDKLV + F QA  +E P+ G+IVHTDQGSQYT +
Sbjct: 134 MQDKLVRDCFLQACGKEHPQPGLIVHTDQGSQYTSS 169
```

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

Example 2423

A DNA sequence (GASx194R) was identified in *S.pyogenes* <SEQ ID 7345> which encodes the amino acid sequence <SEQ ID 7346>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

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

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

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

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

```
>GP:CAA63508 GB:X92946 hypothetical protein [Lactococcus lactis]
 Identities = 64/96 (66%), Positives = 78/96 (80%)

Query: 1   MPRKTFDKAFKLSAVKLILEEEQSVKMVSSTLEIHPNSLYQWIQEYEKYGESAFPGHGSA 60
           M R+ FDK FK SAVKLILEE  SVK VS  LE+H NSLY+W+QE E+YGESAFPG+G+A
Sbjct: 1   MARRKFDKQFKNSAVKLILEEGYSVKEVSQELEVHANSLYRWVQEVEEYGESAFPGNGTA 60

Query: 61  LRHAQFETKKLEKEHKLLQEELALLKKFQVFLKPNR 96
           L +AQ + K LEKE++ LQEEL LLKKF+VFLK ++
Sbjct: 61  LANAQHKIKLLEKENRYLQEELELLKKFRVFLKRSK 96
```

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

Example 2424

A DNA sequence (GASx195R) was identified in *S.pyogenes* <SEQ ID 7347> which encodes the amino acid sequence <SEQ ID 7348>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-11.30   Transmembrane  179 - 195 ( 173 - 201)
   INTEGRAL    Likelihood = -8.86   Transmembrane  229 - 245 ( 224 - 254)
   INTEGRAL    Likelihood = -8.39   Transmembrane  289 - 305 ( 280 - 307)
   INTEGRAL    Likelihood = -8.23   Transmembrane  417 - 433 ( 410 - 435)
   INTEGRAL    Likelihood = -5.89   Transmembrane  324 - 340 ( 323 - 349)
   INTEGRAL    Likelihood = -4.73   Transmembrane  260 - 276 ( 256 - 278)
   INTEGRAL    Likelihood = -4.51   Transmembrane   96 - 112 (  91 - 113)
   INTEGRAL    Likelihood = -4.25   Transmembrane   24 -  40 (  20 -  43)
   INTEGRAL    Likelihood = -2.44   Transmembrane  344 - 360 ( 342 - 360)

----- Final Results -----
              bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:CAB75191 GB:AL139075 putative integral membrane protein
          [Campylobacter jejuni]
 Identities = 177/430 (41%), Positives = 274/430 (63%), Gaps = 8/430 (1%)

Query: 5   IIISAIALAIGIGYRTKINIGLLAIAFSYLIATTLMGLSPKELLHFWPTSLFFTIFSVSL 64
           +IIS+I +AI +GY T+ N+G+ A+ F+Y+I    M L+PK+++ FWP S+FF IF+VSL
Sbjct: 6   LIISSIIVAIILGYITRHNVGIFAMIFAYIIGAFFMDLAPKKIIAFWPISIFFVIFAVSL 65

Query: 65  FYNVATTNGTLDVLAQHILYRTRTHPNALYMILYLIATLLSALGAGFFTTMAVCCPLAIT 124
           FYN AT NGTL+ LA H++YR   HP  L  ++++++ +++ALGAGF+T +A    PL
Sbjct: 66  FYNFATVNGTLEKLAGHLMYRFANHPYLLPFVIFVVSAIIAALGAGFYTVLAFMAPLTFL 125

Query: 125 LCQKADKHPLIGAQAVNWGASGGANLITSGSGIVFQGLFKQMGWE-EQAFSLGNHIFIVS 183
           LC K     + GA A+N+GA GGAN ITS SGI+F+GL +  G E  +AF+  + IF  +
Sbjct: 126 LCDKIGLSKIAGAMAINYGALGGANFITSQSGIIFRGLMENSGIEANEAFANSSIIFAFT 185

Query: 184 IIYPLIVLLLLSCYIRYSKGRTNSSLT-IDQPPVLSKVQRQTTLLMISSMVLVWLFPLLL 242
           II P++VL     +  ++  + N  ++ I +P       Q+ T +LM    +V+V +FP+L
Sbjct: 186 IILPIVVL----SFFVFNAFKNNIKISVISKPDPFDYKQKTTLILMFMMIVVVLIFPVLN 241

Query: 243 LIFPNIAWIATYRQTFDIGFVSILMVCLALRLKLGKQEAILAKVPWAIIIMLCGMSLLMS 302
           +IFP+   I+ + +  DI  ++++ V +AL LKL  ++ ++A +PW  +IM+CG+ +L+S
Sbjct: 242 IIFPHNETISYFNKKIDIAMIAMIFVAIALFLKLADEKQVVALIPWGTLIMICGVGMLIS 301

Query: 303 LAVKSGLVTLIGHLITTTIPHFWLPLFFCVIAGVMSLFSSTLSVVAPTLFPIIATISAQS 362
           +AV++G + L   L+  I   ++PL  C IA  MSLFSSTL VV P LFPI+ +I+A S
Sbjct: 302 IAVEAGAIKLFSDLVENEINVIFIPLIMCAIAAFMSLFSSTLGVVTPALFPIVPSIAASS 361

Query: 363 PHIDIRLLTTATIIGALSTNISPFSSAGSLIQLSLPHIEERSLAFKKQILLGVPISLSLA 422
             +   LL +  ++GA ++ ISPFSS GSLI  S P   +  L FK  ++  VPI    A
Sbjct: 362 -GLSEALLFSCIVVGAQASAISPFSSGGSLILGSCPDKYKEKL-FKDLLIKAVPIGFIAA 419

Query: 423 LLTIWILMLL 432
           +L   I+  +
Sbjct: 420 ILATIIMSFI 429
```

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

Example 2425

A DNA sequence (GASx196) was identified in *S.pyogenes* <SEQ ID 7349> which encodes the amino acid sequence <SEQ ID 7350>. Analysis of this protein sequence reveals the following:

```
Possible site: 57

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0563(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:AAC45128 GB:U65510 nicotinate-nucleotide pyrophosphorylase
          [Rhodospirillum rubrum]
 Identities = 116/277 (41%), Positives = 170/277 (60%), Gaps = 4/277 (1%)

Query: 17  LTPFQIDDTLKAALREDV-HSEDYSTNAIFDHHGQAKVSLFAKEAGVLAGLTVFQRVFTL 75
           L+PF ID+ ++ AL ED+  + D ++ A      +A     A++ G+LAGL   +  F L
Sbjct: 10  LSPFAIDEAVRRALAEDLGRAGDITSTATIPAATRAHARFVARQPGILAGLGCARSAFAL 69

Query: 76  FDTEVTFQNPHQFKDGDRLTSGDLVLEIIGSVRSLLTCERVALNFLQHLSGIASMTAAYV 135
            D  VTF  P   +DG  + +G  V E+ G+ R++L  ER ALNFL HLSGIA+ T  +
Sbjct: 70  LDDTVTFTTP--LEDGAEIAAGQTVAEVAGAARTILAAERTALNFLGHLSGIATRTRRFG 127

Query: 136 EALGDDRIKVFDTRKTTPNLRLFEKYAVRVGGGYNHRFNLSDAIMLKDNHIAAVGSVQKA 195
           +A+   R ++  TRKTTP LR  EKYAVR GGG NHRF L DA+++KDNHIA  G V  A
Sbjct: 128 DAIAHTRARLTCTRKTTPGLRGLEKYAVRCGGGSNHRFGLDDAVLIKDNHIAVAGGVSAA 187

Query: 196 IAQARAYAPFVKMVEVEVESL-AAAEEAAAAGVDIIMLDNMSLEQIEQAITLIAGRSRIE 254
           +++ARA    +  +E+EV++L   AE  A  G ++++LDNM    + +A+ ++AGR   E
Sbjct: 188 LSRARAGVGHMVRIEIEVDTLEQLAEVLAVGGAEVVLLDNMDAPTLTRAVDMVAGRLVTE 247

Query: 255 CSGNIDMTTISRFRGLAIDYVSSGSLTHSAKSLDFSM 291
            SG + + TI+       +DY+S G+LTHS  +LD  +
Sbjct: 248 ASGGVSLDTIAALAESGVDYISVGALTHSVTTLDIGL 284
```

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

Example 2426

A DNA sequence (GASx199) was identified in *S.pyogenes* <SEQ ID 7351> which encodes the amino acid sequence <SEQ ID 7352>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1649(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2427

A DNA sequence (GASx201) was identified in *S.pyogenes* <SEQ ID 7353> which encodes the amino acid sequence <SEQ ID 7354>. 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>
```

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

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

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

Example 2428

A DNA sequence (GASx203) was identified in *S.pyogenes* <SEQ ID 7355> which encodes the amino acid sequence <SEQ ID 7356>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

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

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

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

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

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

Example 2429

A DNA sequence (GASx210) was identified in *S.pyogenes* <SEQ ID 7357> which encodes the amino acid sequence <SEQ ID 7358>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

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

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

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

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

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

Example 2430

A DNA sequence (GASx211) was identified in *S.pyogenes* <SEQ ID 7359> which encodes the amino acid sequence <SEQ ID 7360>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

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

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

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

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

Example 2431

A DNA sequence (GASx213) was identified in *S.pyogenes* <SEQ ID 7361> which encodes the amino acid sequence <SEQ ID 7362>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4430(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2432

A DNA sequence (GASx219) was identified in *S.pyogenes* <SEQ ID 7363> which encodes the amino acid sequence <SEQ ID 7364>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

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

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

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

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

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

Example 2433

A DNA sequence (GASx220) was identified in *S.pyogenes* <SEQ ID 7365> which encodes the amino acid sequence <SEQ ID 7366>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0530(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2434

A DNA sequence (GASx231R) was identified in *S.pyogenes* <SEQ ID 7367> which encodes the amino acid sequence <SEQ ID 7368>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

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

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

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

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

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

Example 2435

A DNA sequence (GASx237) was identified in *S.pyogenes* <SEQ ID 7369> which encodes the amino acid sequence <SEQ ID 7370>. Analysis of this protein sequence reveals the following:

```
Possible site: 52

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4961(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:CAB49143 GB:AJ248283 hypothetical protein [Pyrococcus abyssi]
 Identities = 79/229 (34%), Positives = 131/229 (56%), Gaps = 11/229 (4%)

Query: 18  MRFTIDQNMQFPLVEIDLEHGGSVYLQQGSMVYHTENVTLNTKLNGKGSGLGKLVGAIGR 77
           M + I+    F L+E++L  G +V  + G+MVY    V++ TK  G       L+GA+ R
Sbjct: 1   MEYRIEHRPSFSLLEVNLREGEAVQAEAGAMVYMDPTVSIETKARGG------LLGALKR 54

Query: 78  SMVSGESMFITQAMSNGDGKLALAPNTPGQIVALELGEKQYRLNDGAFLALDGSAQYKME 137
           S++ GES F+   +  G G++  AP  PG I++LEL    Y    GAFL          ++
Sbjct: 55  SVLGGESFFMN--VFRGPGRVGFAPGYPGDIISLELNGTLYA-QSGAFLVASEGIDIDVK 111

Query: 138 RQNIGKALFGGQGGLFVMTTEGLGTLLANSFGSIKKITLDGGTMTIDNAHVVAWSRELDY 197
               GK +FG +G +F++  +G G +  +S+G+I+KITL G ++ +D  H+VA++  +D+
Sbjct: 112 FGG-GKTIFGREG-VFLLELKGKGIVFLSSYGAIEKITLRGESVIVDTGHMVAFTEGIDF 169

Query: 198 DIHLENGFMQSIGTGEGVVNTFRGHGEIYIQSLNLEQFAGTLKRYLPTS 246
           I    G   ++ +GEG+V  F GHG++YIQ+ +L+ F   +  +LP S
Sbjct: 170 RIRKIGGLKATLFSGEGLVFEFSGHGDVYIQTRSLDGFLSWILPHLPKS 218
```

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

Example 2436

A DNA sequence (GASx240R) was identified in *S.pyogenes* <SEQ ID 7371> which encodes the amino acid sequence <SEQ ID 7372>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2745(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2437

A DNA sequence (GASx241) was identified in *S.pyogenes* <SEQ ID 7373> which encodes the amino acid sequence <SEQ ID 7374>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.14   Transmembrane  196 - 212 ( 187 - 215)
   INTEGRAL    Likelihood = -8.01   Transmembrane  160 - 176 ( 156 - 179)
   INTEGRAL    Likelihood = -5.89   Transmembrane  116 - 132 ( 110 - 134)
   INTEGRAL    Likelihood = -4.57   Transmembrane   74 -  90 (  73 -  97)
   INTEGRAL    Likelihood = -2.66   Transmembrane   51 -  67 (  50 -  68)
   INTEGRAL    Likelihood = -2.60   Transmembrane    8 -  24 (   7 -  27)
   INTEGRAL    Likelihood = -1.28   Transmembrane  344 - 360 ( 344 - 360)
   INTEGRAL    Likelihood = -0.22   Transmembrane   30 -  46 (  30 -  46)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:CAC10175 GB:AJ278302 histidine kinase [Streptococcus pneumoniae]
 Identities = 136/449 (30%), Positives = 234/449 (51%), Gaps = 26/449 (5%)

Query: 8   FLLLSIIVYYMTKIYIFSFLSDITLP---VWKQLTI-LALALFFNQFPYLS-----PLLI 58
           ++LL  +V +  KI IF  +  I+L    ++K     + LA+ F    Y+         +
Sbjct: 5   WILLYTLVTHGLKIVIFFKVDGISLTFERIFKAFLFKILLAVVFGMLGYMVGNVYLSYFM 64

Query: 59  DPL----LFLVVLRQETKQLFSLKALFLAVAPSVLVDLLSRFMGTIVIPYLFLSSGIYLG 114
           +PL    L  ++LR+  K+L     LF  + P +LV+L  R +   V+P  FL  G
Sbjct: 65  EPLYGIGLSFLLLRELPKKLL----LFYGLFPMILVNLFYRGVSYFVLP--FLGQGQVYD 118

Query: 115 HIIFDLLAYLLIFPSFAIINYMIGKDYKMIC-QSGYSKRSHNFYQTLLMFVLVYYVDIFV 173
              F  L  ++IF  F  + ++   DY     + G   +      T + +++  Y  +
Sbjct: 119 DYSFIWLC-IIIFNFFISLAFLKWLDYDFTSLRKGILDKDFQKSLTQINWIMGAYYLVIQ 177

Query: 174 ILGFTDPFLHFHHSLFVPTPYKLLFLMFILLLVYLLSYFNHSSKEYLKNELRREQQAYMT 233
            L + +    +   +   T   L+ + ++L  + ++   +   K+ L   L +EQ
Sbjct: 178 NLSYFE----YEQGIQSTTVRHLILVFYLLFFMGIIKKLDTYLKDKLHERLNQEQDLRYR 233

Query: 234 NLEIYGKHLEKLYRDVRAFQSDYLSRIERLGQAIKSESITQIQDIYAQTVHEANDYWDDK 293
            +E Y +H+E+LY++VR+F+ DY + +  L   I+ E + QI++IY   + ++++   D
Sbjct: 234 EMERYSRHIEELYKEVRSFRHDYTNLLTSLRLGIEEEDMEQIKEIYDSVLKDSSEKLQDN 293

Query: 294 HYNISKLRKINISSIKSLLSAKIISAEKSGIDLNVEVPDNIKETYIPELDLLLLMSIFCD 353
            Y++ +L  +   ++KSLL+ K I A    I  NVEVP+ I+   + LD L ++SI CD
Sbjct: 294 KYDLGRLVNVRDRALKSLLAGKFIKARDKNIVFNVEVPEEIQVEGVSLLDFLTVVSILCD 353

Query: 354 NAIEAALEAQQPHMSIAYFLLGDYQMFVVTNTTKKK-VDINKIFEEGYSSKGSERGIGLS 412
           NAIEA++EA QPH+SIA+F  G  + F++ N+ K++ +DI++IF  G SSKG ERG+GL
Sbjct: 354 NAIEASVEACQPHVSIAFFKNGAQETFIIENSIKEEGIDISEIFSFGASSKGEERGVGLY 413

Query: 413 NAQRILKKYPYLSLRTKSFDKEFSQTLTM 441
              +I++ +P  SL T   D  F Q LT+
Sbjct: 414 TVMKIVESHPNTSLNTTCQDHVFRQVLTV 442
```

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

Example 2438

A DNA sequence (GASx242R) was identified in *S.pyogenes* <SEQ ID 7375> which encodes the amino acid sequence <SEQ ID 7376>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4165(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2439

A DNA sequence (GASx243) was identified in *S.pyogenes* <SEQ ID 7377> which encodes the amino acid sequence <SEQ ID 7378>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.09   Transmembrane  188 - 204 ( 182 - 208)
   INTEGRAL    Likelihood = -7.17   Transmembrane   52 -  68 (  47 -  69)
   INTEGRAL    Likelihood = -4.73   Transmembrane  119 - 135 ( 114 - 142)
   INTEGRAL    Likelihood = -4.62   Transmembrane   83 -  99 (  77 - 107)
   INTEGRAL    Likelihood = -1.86   Transmembrane  328 - 344 ( 328 - 345)
   INTEGRAL    Likelihood = -1.65   Transmembrane    7 -  23 (   6 -  23)
   INTEGRAL    Likelihood = -0.22   Transmembrane   35 -  51 (  35 -  51)

----- 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>
```

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

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

```
>GP:CAC10175 GB:AJ278302 histidine kinase [Streptococcus pneumoniae]
 Identities = 123/438 (28%), Positives = 229/438 (52%), Gaps = 49/438 (11%)

Query: 20  VIFAKVSAIKLSWKRVS-------IIGISFVIANMIFDKVIL---IDQLFFIIVSLL--- 66
           VIF KV  I L+++R+         ++ + F +   +   V L   ++ L+ I +S L
Sbjct: 19  VIFFKVDGISLTFERIFKAFLFKILLAVVFGMLGYMVGNVYLSYFMEPLYGIGLSFLLLR 78

Query: 67  SAPKKKLFEHMFNGFFTILIVELLFRVIGSFFLPAVLGFSIGQINNNLKLLELCYLFVLP 126
             PKK L   +F G F +++V L +R +  F LP +     GQ+ ++   + LC + +
Sbjct: 79  ELPKKLL---LFYGLFPMILVNLFYRGVSYFVLPFL---GQGQVYDDYSFIWLC-IIIFN 131

Query: 127 IFYLFSYIFSIDL---SLIRFISEDKMKKWVFWMNTAMFSYYFFAHFLVTVQSGFLALYF 183
            F  +++  +D    SL + I +   +K +  +N  M +YY     L          YF
Sbjct: 132 FFISLAFLKWLDYDFTSLRKGILDKDFQKSLTQINWIMGAYYLVIQNLS---------YF 182

Query: 184 QY---------RSILVFIYLAIFIWVIVKLDRFAKDQLSQKLTQAQNERIAYLENYNQSI 234
           +Y         R +++  YL  F+ +I KLD + KD+L ++L Q Q+ R   +E Y++ I
Sbjct: 183 EYEQGIQSTTVRHLILVFYLLFFMGIIKKLDTYLKDKLHERLNQEQDLRYREMERYSRHI 242

Query: 235 EQLYREIRTVKHDSENILISLKDSIDSGDIDLITRVYDTVIQQSATSMMRTNYEISSLDN 294
           E+LY+E+R+ +HD  N+L SL+  I+  D++ I  +YD+V++ S+  +    Y++  L N
Sbjct: 243 EELYKEVRSFRHDYTNLLTSLRLGIEEEDMEQIKEIYDSVLKDSSEKLQDNKYDLGRLVN 302

Query: 295 IKEAVIRSIMNSKLLEAQYLGIELYIEIPDVIDHLPIKLIDLIVLFTGLVDNAIETAKGS 354
           +++  ++S++  K ++A+   I   +E+P+ I    + L+D + + + L DNAIE +  +
Sbjct: 303 VRDRALKSLLAGKFIKARDKNIVFNVEVPEEIQVEGVSLLDFLTVVSILCDNAIEASVEA 362

Query: 355 RRPFLSIAYFKQDNKQLFIIENSTKTNRVDIAKRFDAQQQNSAH--------FLTVLDSY 406
            +P +SIA+FK   ++ FIIENS K   +DI++ F     +             + +++S+
Sbjct: 363 CQPHVSIAFFKNGAQETFIIENSIKEEGIDISEIFSFGASSKGEERGVGLYTVMKIVESH 422

Query: 407 PQITLSTKSDHYRLRQLL 424
           P  +L+T    +  RQ+L
Sbjct: 423 PNTSLNTTCQDHVFRQVL 440
```

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

Example 2440

A DNA sequence (GASx248) was identified in *S.pyogenes* <SEQ ID 7379> which encodes the amino acid sequence <SEQ ID 7380>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5665(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2441

A DNA sequence (GASx255) was identified in *S.pyogenes* <SEQ ID 7381> which encodes the amino acid sequence <SEQ ID 7382>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1437(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2442

A DNA sequence (GASx270R) was identified in *S.pyogenes* <SEQ ID 7383> which encodes the amino acid sequence <SEQ ID 7384>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.89   Transmembrane   20 -  36 (  17 -  36)

----- Final Results -----
             bacterial membrane --- Certainty=0.3357(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2443

A DNA sequence (GASx272) was identified in *S.pyogenes* <SEQ ID 7385> which encodes the amino acid sequence <SEQ ID 7386>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2488(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
>GP:CAB11887 GB:Z99104 ribosomal protein S7 (BS7) [Bacillus subtilis]
 Identities = 117/156 (75%), Positives = 139/156 (89%)

Query: 1   MSRKNQAPKREVLPDPLYNSKIVTRLINRVMLDGKRGTAATIVYDAFNAIKEATGNDALE 60
            M RK    KR+VLPDP+YNSK+V+RLIN++M+DGK+G   TI+Y +F+ IKE TGNDA+E
Sbjct: 1   MPRKGPVAKRDVLPDPIYNSKLVSRLINKMMIDGKKGKPQTILYKSFDIIKERTGNDAME 60

Query: 61  VFETAMDNIMPVLEVRARRVGGSNYQVPVEVRPERRTTLGLRWLVNASRARGEHTMKDRL 120
            VFE A+ NIMPVLEV+ARRVGG+NYQVPVEVRPERRTTLGLRWLVN +R RGE TM++RL
Sbjct: 61  VFEQALKNIMPVLEVKARRVGGANYQVPVEVRPERRTTLGLRWLVNYARLRGEKTMEERL 120

Query: 121 AKEIMDAANNTGASVKKREDTHKMAEANRAFAHFRW 156
            A EI+DAANNTGA+VKKREDTHKMAEAN+AFAH+RW
Sbjct: 121 ANEILDAANNTGAAVKKREDTHKMAEANKAFAHYRW 156
```

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

Example 2444

A DNA sequence (GASx274) was identified in *S.pyogenes* <SEQ ID 7387> which encodes the amino acid sequence <SEQ ID 7388>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> 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 sequence was also identified in GAS <SEQ ID 9095> which encodes the amino acid sequence <SEQ ID 9096>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty= 0.291(Affirmative) < succ>
             bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
              bacterial outside --- Certainty= 0.000(Not Clear) < succ>
```

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

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

Example 2445

A DNA sequence (GASx275R) was identified in *S.pyogenes* <SEQ ID 7389> which encodes the amino acid sequence <SEQ ID 7390>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5664(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2446

A DNA sequence (GASx283) was identified in *S.pyogenes* <SEQ ID 7391> which encodes the amino acid sequence <SEQ ID 7392>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0724(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2447

A DNA sequence (GASx298) was identified in *S.pyogenes* <SEQ ID 7393> which encodes the amino acid sequence <SEQ ID 7394>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2840(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2448

A DNA sequence (GASx300) was identified in *S.pyogenes* <SEQ ID 7395> which encodes the amino acid sequence <SEQ ID 7396>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.91   Transmembrane    4 -  20 (   4 -  20)

----- 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>
```

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

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

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

Example 2449

A DNA sequence (GASx301) was identified in *S.pyogenes* <SEQ ID 7397> which encodes the amino acid sequence <SEQ ID 7398>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4884(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2450

A repeated DNA sequence (GASx302) was identified in *S.pyogenes* <SEQ ID 7399> which encodes the amino acid sequence <SEQ ID 7400>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

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

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

```
bacterial cytoplasm --- Certainty=0.2581(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2451

A DNA sequence (GASx316) was identified in *S.pyogenes* <SEQ ID 7401> which encodes the amino acid sequence <SEQ ID 7402>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.80   Transmembrane   23 -  39 (  22 -  39)

----- 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>
```

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

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

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

Example 2452

A DNA sequence (GASx323R) was identified in *S.pyogenes* <SEQ ID 7403> which encodes the amino acid sequence <SEQ ID 7404>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.0005(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 2453

A DNA sequence (GASx334) was identified in *S.pyogenes* <SEQ ID 7405> which encodes the amino acid sequence <SEQ ID 7406>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
```

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

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

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

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2454

A DNA sequence (GASx336) was identified in *S.pyogenes* <SEQ ID 7407> which encodes the amino acid sequence <SEQ ID 7408>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3379(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2455

A DNA sequence (GASx361R) was identified in *S.pyogenes* <SEQ ID 7409> which encodes the amino acid sequence <SEQ ID 7410>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2807(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2456

A DNA sequence (GASx387) was identified in *S.pyogenes* <SEQ ID 7411> which encodes the amino acid sequence <SEQ ID 7412>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2740(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2457

A DNA sequence (GASx389) was identified in *S.pyogenes* <SEQ ID 7413> which encodes the amino acid sequence <SEQ ID 7414>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0744(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2458

A DNA sequence (GASx392) was identified in *S.pyogenes* <SEQ ID 7415> which encodes the amino acid sequence <SEQ ID 7416>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2162(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2459

A DNA sequence (GASx393R) was identified in *S.pyogenes* <SEQ ID 7417> which encodes the amino acid sequence <SEQ ID 7418>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2520(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2460

A DNA sequence (GASx395) was identified in *S.pyogenes* <SEQ ID 7419> which encodes the amino acid sequence <SEQ ID 7420>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2590(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2461

A DNA sequence (GASx396) was identified in *S.pyogenes* <SEQ ID 7421> which encodes the amino acid sequence <SEQ ID 7422>. Analysis of this protein sequence reveals the following:

```
Possible site: 41

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13373 GB:Z99111 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 23/88 (26%), Positives = 52/88 (58%)
```

```
Query: 4  KQERIGLVVYLYYNRDARKLSKFGDLYYHSKRSRYLIIYINKNDLDTKLEEMRRLKCVKD 63
          +  R G+VVYL+  + ++ L KFG+++Y SKR +Y+++Y + + ++  ++++     VK
Sbjct: 2  ENRRQGMVVYLHSLKQSKMLRKFGNVHYVSKRLKYVVLYCDMDQIEKTMDKIASYSFVKK 61

Query: 64 IRPSAFDDIDRQFVGNLHRDETNNHQKG 91
          + PS    +  +F   L + +  +++ G
Sbjct: 62 VEPSYKPFLKLEFESKLDKAKEYDYKIG 89
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2462

A DNA sequence (GASx400) was identified in *S.pyogenes* <SEQ ID 7423> which encodes the amino acid sequence <SEQ ID 7424>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2010(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2463

A DNA sequence (GASx401) was identified in *S.pyogenes* <SEQ ID 7425> which encodes the amino acid sequence <SEQ ID 7426>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1176(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2464

A DNA sequence (GASx402) was identified in *S.pyogenes* <SEQ ID 7427> which encodes the amino acid sequence <SEQ ID 7428>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2938(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2465

A DNA sequence (GASx403R) was identified in *S.pyogenes* <SEQ ID 7429> which encodes the amino acid sequence <SEQ ID 7430>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2466

A DNA sequence (GASx406) was identified in *S.pyogenes* <SEQ ID 7431> which encodes the amino acid sequence <SEQ ID 7432>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.26   Transmembrane   15 -  31 (   4 -  36)
   INTEGRAL    Likelihood = -6.64   Transmembrane   96 - 112 (  94 - 115)

----- Final Results -----
               bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2467

A DNA sequence (GASx408R) was identified in *S.pyogenes* <SEQ ID 7433> which encodes the amino acid sequence <SEQ ID 7434>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.23   Transmembrane   17 -  33 (  15 -  34)
   INTEGRAL    Likelihood = -0.85   Transmembrane   38 -  54 (  38 -  54)

----- Final Results -----
               bacterial membrane --- Certainty=0.1893(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2468

A DNA sequence (GASx412) was identified in *S.pyogenes* <SEQ ID 7435> which encodes the amino acid sequence <SEQ ID 7436>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.53   Transmembrane    5 -  21 (   4 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.3612(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2469

A DNA sequence (GASx413) was identified in *S.pyogenes* <SEQ ID 7437> which encodes the amino acid sequence <SEQ ID 7438>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3422(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA68903 GB:Y07622 lactate oxidase [Streptococcus iniae]
 Identities = 328/392 (83%), Positives = 359/392 (90%), Gaps = 4/392 (1%)

Query: 3   MAQKTVITEETTDFVMDFKTSSAEGNVDFINVFDLEKMAQQVIPKGAFGYIASGAGDTFT 62
           M  K+ +   TT   ++FKTSSAEG+VDF+NVFDLEKMAQ+VIPKGAFGYIASGAGDTFT
Sbjct: 1   MENKSEMINATT---IEFKTSSAEGSVDFVNVFDLEKMAQKVIPKGAFGYIASGAGDTFT 57

Query: 63  LHENIRSFNHKLIVPHSLKGVENPSTEITFDGDYLTSPLILAPVAAHKLANEQGEVASAK 122
           LHENIRSFNHKLI PH LKGVENPSTEITF GD L SP+ILAPVAAHKLANEQGE+ASAK
Sbjct: 58  LHENIRSFNHKLI-PHGLKGVENPSTEITFIGDKLASPIILAPVAAHKLANEQGEIASAK 116

Query: 123 GLKEFGSIYTTSSYSTTDLPEISAALGGTPHWFQFYYSKDDGINRNIMDRVKAQGCKAIV 182
           G+KEFG+IYTTSSYSTTDLPEIS  LG +PHWFQFYYSKDDGINR+IMDR+KA+G K+IV
Sbjct: 117 GVKEFGTIYTTSSYSTTDLPEISQTLGDSPHWFQFYYSKDDGINRHIMDRLKAEGVKSIV 176

Query: 183 LTADATVGGNREVDRRNGFVFPVGMPIVQEYLPDGAGKTMDYVYKSAKQALTSKDIEYIA 242
           LT DATVGGNREVD+RNGFVFPVGMPIVQEYLP+GAGKTMDYVYK+ KQAL+ KD+EYIA
Sbjct: 177 LTVDATVGGNREVDKRNGFVFPVGMPIVQEYLPNGAGKTMDYVYKATKQALSPKDVEYIA 236

Query: 243 TYSGLPVYVKGPQCAEDTLRALDAGASGIWVTNHGGRQLDGGPAAFDSLQEVAEAVDQKV 302
            YSGLPVYVKGPQCAED  RAL+AGASGIWVTNHGGRQLDGGPAAFDSLQEVAE+VD++V
Sbjct: 237 QYSGLPVYVKGPQCAEDAFRALEAGASGIWVTNHGGRQLDGGPAAFDSLQEVAESVDRRV 296

Query: 303 PIVFDSGIRRGQHIFKALASGADLVALGRPAIYGLAMGGSIGTRQVFEKLNDELKMVMQL 362
           PIVFDSG+RRGQH+FKALASGADLVALGRP IYGLAMGGS+GTRQVFEK+NDELKMVMQL
Sbjct: 297 PIVFDSGVRRGQHVFKALASGADLVALGRPVIYGLAMGGSVGTRQVFEKINDELKMVMQL 356

Query: 363 AGTQTIQDVKAFNLRHNPYDSSIPFDQNALRL 394
           AGTQTI DVK F LRHNPYDSSIPF     ++
Sbjct: 357 AGTQTIDDVKHFKLRHNPYDSSIPFSPKCFKI 388
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2470

A DNA sequence (GASx414) was identified in *S.pyogenes* <SEQ ID 7439> which encodes the amino acid sequence <SEQ ID 7440>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0682(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2471

A DNA sequence (GASx417R) was identified in *S.pyogenes* <SEQ ID 7441> which encodes the amino acid sequence <SEQ ID 7442>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1765(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2472

A DNA sequence (GASx418) was identified in *S.pyogenes* <SEQ ID 7443> which encodes the amino acid sequence <SEQ ID 7444>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2532(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2473

A DNA sequence (GASx419) was identified in *S.pyogenes* <SEQ ID 7445> which encodes the amino acid sequence <SEQ ID 7446>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3082(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2474

A DNA sequence (GASx423) was identified in *S.pyogenes* <SEQ ID 7447> which encodes the amino acid sequence <SEQ ID 7448>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
```

```
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.18   Transmembrane   14 -  30 (  13 -  31)

----- Final Results -----
               bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2475

A DNA sequence (GASx427R) was identified in *S.pyogenes* <SEQ ID 7449> which encodes the amino acid sequence <SEQ ID 7450>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.17   Transmembrane   13 -  29 (  10 -  29)

----- 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>
```

A related sequence was also identified in GAS <SEQ ID 9105> which encodes the amino acid sequence <SEQ ID 9106>. Analysis of this protein sequence reveals the following:

```
     Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.17   Transmembrane   8 -  24

----- Final Results -----
               bacterial membrane --- Certainty=0.1470(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA26616 GB:M63917 epidermal cell differentiation inhibitor
          [Staphylococcus aureus]
 Identities = 58/195 (29%), Positives = 106/195 (53%), Gaps = 13/195 (6%)

Query: 67  RWGKGLI----YPRAEQEAMAAYTCQQAGPINTSLDKAKGELSQLTPELRDQVAQLDAAT 122
           +WG  LI    Y   ++ A+  YT + +  IN  L  A G++++L    +D+V +LD++
Sbjct: 49  KWGNKLIKQAKYSSDDKIALYEYT-KDSSKINGPLRLAGGDINKLDSTTQDKVRRLDSSI 107

Query: 123 HRLVIPWNIVVYRYVYETFLRDI-GVSHADLTSYYR--NHQFDPHILCKIK--LGTR-YT 176
            +   P ++ VYR +    +L  I G ++ DL     +  N Q+D +++ K+    + +R Y
Sbjct: 108 SKSTTPESVYVYRLLNLDYLTSIVGFTNEDLYKLQQTNNGQYDENLVRKLNNVMNSRIYR 167

Query: 177 KHSFMSTTALKNGAMTHRPVEVRICVKKGAKAAFV--EPYSAVPSEVELLFPRGCQLEVV 234
           +  + ST  +   A+  RP+E+R+  KG KAA++  +  +A   + E+L PRG +  V
Sbjct: 168 EDGYSSTQLVSGAAVGGRPIELRLELPKGTKAAYLNSKDLTAYYGQQEVLLPRGTEYAVG 227

Query: 235 GAYVSQDQKKLHIEA 249
              +S D+KK+ I A
```

```
Sbjct: 228 SVELSNDKKKIIITA 242
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2476

A DNA sequence (GASx428) was identified in *S.pyogenes* <SEQ ID 7451> which encodes the amino acid sequence <SEQ ID 7452>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3817(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2477

A DNA sequence (GASx429) was identified in *S.pyogenes* <SEQ ID 7453> which encodes the amino acid sequence <SEQ ID 7454>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
              bacterial outside --- Certainty=0.3000(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2478

A DNA sequence (GASx431) was identified in *S.pyogenes* <SEQ ID 7455> which encodes the amino acid sequence <SEQ ID 7456>. Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.60   Transmembrane   68 -  84 (  66 -  90)
   INTEGRAL    Likelihood = -6.85   Transmembrane   22 -  38 (  16 -  42)
   INTEGRAL    Likelihood = -3.29   Transmembrane   44 -  60 (  43 -  61)

----- Final Results -----
            bacterial membrane --- Certainty=0.4439(Affirmative) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2479

A DNA sequence (GASx432R) was identified in *S.pyogenes* <SEQ ID 7457> which encodes the amino acid sequence <SEQ ID 7458>. Analysis of this protein sequence reveals the following:

```
Possible site:: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2480

A DNA sequence (GASx434) was identified in *S.pyogenes* <SEQ ID 7459> which encodes the amino acid sequence <SEQ ID 7460>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2481

A DNA sequence (GASx435R) was identified in *S.pyogenes* <SEQ ID 7461> which encodes the amino acid sequence <SEQ ID 7462>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.50   Transmembrane    4 -  20 (   3 -  21)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.1999(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB59092 GB:M97157 pyrogenic exotoxin C [Streptococcus pyogenes]
 Identities = 110/229 (48%), Positives = 150/229 (65%), Gaps = 4/229 (1%)

Query: 4   IIKTIILVIIIFHGYGS--VKSDSE-NIKDVKLQLNYAYEIIPVDYTNCNIDYLTTHDFY 60
           IIK + ++ +I     S  +KSDS+ +I +VK  L YAY I P DY +C +++ TTH
Sbjct: 6   IIKIVFIITVILISTISPIIKSDSKKDISNVKSDLLYAYTITPYDYKDCRVNFSTTHTLN 65

Query: 61  IDISSYKKKNFSVDSEVESYITTKFTKNQKVNIFGLPYIFTRYDVYYIYGGVTPSVNSNS 120
           ID   Y+ K++ + SE+    + KF ++  V++FGL YI   +   YIYGG+TP+ N N
Sbjct: 66  IDTQKYRGKDYYISSEMSYEASQKFKRDDHVDVFGLFYILNSHTGEYIYGGITPAQN-NK 124

Query: 121 ENSKIVGNLLIDGVQQKTLINPIKIDKPIFTIQEFDFKIRQYLMQTYKIYDPNSPYIKGQ 180
            N K++GNL I G  Q+ L N I ++K I T QE DFKIR+YLM  YKIYD  SPY+ G+
Sbjct: 125 VNHKLLGNLFISGESQQNLNNKIILEKDIVTFQEIDFKIRKYLMDNYKIYDATSPYVSGR 184

Query: 181 LEIAINGNKHESFNLYDATSSSTRSDIFKKYKDNKTINMKDFSHFDIYL 229
           +EI     KHE  +L+D+ +  TRSDIF KYKDN+ INMK+FSHFDIYL
Sbjct: 185 IEIGTKDGKHEQIDLFDSPNEGTRSDIFAKYKDNRIINMKNFSHFDIYL 233
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2482

A DNA sequence (GASx436R) was identified in *S.pyogenes* <SEQ ID 7463> which encodes the amino acid sequence <SEQ ID 7464>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2483

A DNA sequence (GASx446) was identified in *S.pyogenes* <SEQ ID 7465> which encodes the amino acid sequence <SEQ ID 7466>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
```

```
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2484

A DNA sequence (GASx449) was identified in *S.pyogenes* <SEQ ID 7467> which encodes the amino acid sequence <SEQ ID 7468>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.82   Transmembrane    3 -  19 (   1 -  20)

----- Final Results -----
               bacterial membrane --- Certainty=0.2529(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2485

A DNA sequence (GASx450R) was identified in *S.pyogenes* <SEQ ID 7469> which encodes the amino acid sequence <SEQ ID 7470>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.44   Transmembrane   21 -  37 (  19 -  37)

----- Final Results -----
               bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2486

A DNA sequence (GASx457R) was identified in *S.pyogenes* <SEQ ID 7471> which encodes the amino acid sequence <SEQ ID 7472>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
```

```
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-15.34   Transmembrane   64 -  80 (  57 -  86)
   INTEGRAL    Likelihood =-13.43   Transmembrane   97 - 113 (  91 - 116)
   INTEGRAL    Likelihood = -5.57   Transmembrane   38 -  54 (  32 -  56)

----- Final Results -----
               bacterial membrane --- Certainty=0.7135(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2487

A DNA sequence (GASx476R) was identified in *S.pyogenes* <SEQ ID 7473> which encodes the amino acid sequence <SEQ ID 7474>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3013(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2488

A DNA sequence (GASx477) was identified in *S.pyogenes* <SEQ ID 7475> which encodes the amino acid sequence <SEQ ID 7476>. Analysis of this protein sequence reveals the following:

```
Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1022(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC03521 GB:AJ276410 BlpJ protein [Streptococcus pneumoniae]
 Identities = 47/77 (61%), Positives = 59/77 (76%)

Query: 1  MIKFAEEIQKEELFHIIGGYSATDCKNHLIGGITSGAIAGGVGAGMATLGVGGVAGAFAG 60
          M+   E +  E L  + GGYS+TDC+N LI G+T+G I GG GAG+ATLGV G+AGAF G
Sbjct: 5  MLSQLEVMDTEMLAKVEGGYSSTDCQNALITGVTTGIITGGTGAGLATLGVAGLAGAFVG 64
```

```
Query: 61 AHVGAIAGGLTCVGGML 77
          AH+GAI GGLTC+GGM+
Sbjct: 65 AHIGAIGGGLTCLGGMV 81
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2489

A DNA sequence (GASx478) was identified in *S.pyogenes* <SEQ ID 7477> which encodes the amino acid sequence <SEQ ID 7478>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.07   Transmembrane   42 -  58 (  41 -  58)
   INTEGRAL    Likelihood = -1.59   Transmembrane   22 -  38 (  22 -  38)

----- Final Results -----
               bacterial membrane --- Certainty=0.1829(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC03520 GB:AJ276410 BlpI protein [Streptococcus pneumoniae]
 Identities = 35/56 (62%), Positives = 44/56 (78%)

Query: 1  MDNFLELQFEELVNISGGKGNIGSAIGGCLGGMLIAAAGGPITGGAAAFVCVASGI 56
          M+ F  +  EEL  +SGG+GN+GSAIGGC+G +L+AAA GPITGGAA  +CV SGI
Sbjct: 6  MEQFSVMDNEELEIVSGGRGNLGSAIGGCIGAVLLAAATGPITGGAATLICVGSGI 61
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2490

A DNA sequence (GASx482) was identified in *S.pyogenes* <SEQ ID 7479> which encodes the amino acid sequence <SEQ ID 7480>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.43   Transmembrane   61 -  77 (  61 -  79)

----- 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>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC03524 GB:AJ276410 BlpM protein [Streptococcus pneumoniae]
 Identities = 22/52 (42%), Positives = 30/52 (57%)

Query: 29 MEIKKLETFHQMTIEKLAKVEGGKNNWQANVSGVIAAGSAGAAIGFPVCGVA 80
          M+ K +E FH+M I  L+ +EGGKNNWQ NV     A   G  +G  +C  +
Sbjct: 1  MDTKIMEQFHEMDITMLSSIEGGKNNWQTNVLEGGGAAFGGWGLGTAICAAS 52
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2491

A DNA sequence (GASx483) was identified in *S.pyogenes* <SEQ ID 7481> which encodes the amino acid sequence <SEQ ID 7482>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1832(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2492

A DNA sequence (GASx484) was identified in *S.pyogenes* <SEQ ID 7483> which encodes the amino acid sequence <SEQ ID 7484>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2493

A DNA sequence (GASx485) was identified in *S.pyogenes* <SEQ ID 7485> which encodes the amino acid sequence <SEQ ID 7486>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1037(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2494

A DNA sequence (GASx487) was identified in *S.pyogenes* <SEQ ID 7487> which encodes the amino acid sequence <SEQ ID 7488>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1086(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2495

A DNA sequence (GASx488) was identified in *S.pyogenes* <SEQ ID 7489> which encodes the amino acid sequence <SEQ ID 7490>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2176(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2496

A DNA sequence (GASx489R) was identified in *S.pyogenes* <SEQ ID 7491> which encodes the amino acid sequence <SEQ ID 7492>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2497

A DNA sequence (GASx490) was identified in *S.pyogenes* <SEQ ID 7493> which encodes the amino acid sequence <SEQ ID 7494>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2547(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2498

A DNA sequence (GASx491R) was identified in *S.pyogenes* <SEQ ID 7495> which encodes the amino acid sequence <SEQ ID 7496>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.24   Transmembrane    6 -  22 (   3 -  28)

----- Final Results -----
               bacterial membrane --- Certainty=0.5097(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2499

A DNA sequence (GASx492) was identified in *S.pyogenes* <SEQ ID 7497> which encodes the amino acid sequence <SEQ ID 7498>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2500

A DNA sequence (GASx493) was identified in *S.pyogenes* <SEQ ID 7499> which encodes the amino acid sequence <SEQ ID 7500>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.69   Transmembrane   21 -  37 (  21 -  37)

----- Final Results -----
              bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2501

A DNA sequence (GASx495R) was identified in *S.pyogenes* <SEQ ID 7501> which encodes the amino acid sequence <SEQ ID 7502>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2891(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2502

A DNA sequence (GASx499R) was identified in *S.pyogenes* <SEQ ID 7503> which encodes the amino acid sequence <SEQ ID 7504>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have an uncleavable N-term signal seq
```

```
   INTEGRAL    Likelihood = -2.50   Transmembrane    3 -  19 (   1 -  20)

----- Final Results -----
               bacterial membrane --- Certainty=0.1999(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2503

A DNA sequence (GASx500) was identified in *S.pyogenes* <SEQ ID 7505> which encodes the amino acid sequence <SEQ ID 7506>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC77220 GB:AE000497 orf, hypothetical protein [Escherichia
          coli]
 Identities = 262/480 (54%), Positives = 338/480 (69%), Gaps = 5/480 (1%)

Query: 18  GMLNRHGLIAGATGTGKTVTLKVLAEQLSLAGVPVFLADIKGDLSNLTKAGEVTDKLAAR 77
           GM NRHGLI GATGTGKTVTL+ LAE LS  GVPVF+AD+KGDL+ + +AG V++KL AR
Sbjct: 20  GMANRHGLITGATGTGKTVTLQKLAESLSEIGVPVFMADVKGDLTGVAQAGTVSEKLLAR 79

Query: 78  LATIGVSDYQPQAFPVRMWDVFGQNGQPLRTTISELGPMMLSRLLNLNDTQTGVLNIVFK 137
           L  IGV+D+QP A PV +WD+FG+ G P+R T+S+LGP++L+RLLNLND Q+GVLNI+F+
Sbjct: 80  LKNIGVNDWQPHANPVVVWDIFGEKGHPVRATVSDLGPLLLARLLNLNDVQSGVLNIIFR 139

Query: 138 IADEKGWLLIDLKDLQAILKEVGDHASDYSSHYGNIAKQSIGAIQRSLLTLEQEGAHQFF 197
           IAD++G LL+D KDL+AI + +GD+A  + + YGNI+  S+GAIQR LL+LEQ+GA  FF
Sbjct: 140 IADDQGLLLLDFKDLRAITQYIGDNAKSFQNQYGNISSASVGAIQRGLLSLEQQGAAHFF 199

Query: 198 GEPALDVADLMQLDVASGYGAINILSATKLFQSPTLYTTFLLWLLSELYKLLPEVGDLDK 257
           GEP LD+ D M+ D A+G G INILSA KL+Q P LY   LLW+LSELY+ LPE GDL+K
Sbjct: 200 GEPMLDIKDWMRTD-ANGKGVINILSAEKLYQMPKLYAASLLWMLSELYEQLPEAGDLEK 258

Query: 258 PKMVFFFDEAHLLFKDAPKVFLEKVEQIVRLIRSKGVGIFFVTQNPLDLPETVLAQLGNR 317
           PK+VFFFDEAHLLF DAP+V L+K+EQ++RLIRSKGVG++FV+QNP D+P+ VL QLGNR
Sbjct: 259 PKLVFFFDEAHLLFNDAPQVLLDKIEQVIRLIRSKGVGVWFVSQNPSDIPDNVLGQLGNR 318

Query: 318 IQHAFRAYTPKEQKAVRVAADTFRQNPDLDVARVITELEVGEALISVLNDKGQPSIVERA 377
           +QHA RA+TPK+QKAV+ AA T R NP  D  + I EL  GEALIS L+ KG PS+VERA
Sbjct: 319 VQHALRAFTPKDQKAVKAAAQTMRANPAFDTEKAIQELGTGEALISFLDAKGSPSVVERA 378

Query: 378 YIMPPKSSFAVLSEIESQQLVQSSPFASKYSQSIDRESAYEKLAAKVLEDNRLAQEAIAT 437
           ++ P S    ++E E   L+  SP   KY   +DRESAYE L  K  + +   Q
Sbjct: 379 MVIAPCSRMGPVTEDERNGLINHSPVYGKYEDEVDRESAYEML-QKGFQASTEQQNNPPA 437

Query: 438 AQREKEAKEAIKAQAATKKANRRSVGRSHKTVVEKATDAFISTTVRTIGRELVRGLLGSL 497
```

```
            +E    + I                  K   +        +  R +  ++VRG+LGSL
Sbjct: 438 KGKEVAVDDGILGGLKDILFGTTGPRGGKK---DGVVQTMAKSAARQVTNQIVRGMLGSL 494
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2504

A DNA sequence (GASx502) was identified in *S.pyogenes* <SEQ ID 7507> which encodes the amino acid sequence <SEQ ID 7508>. Analysis of this protein sequence reveals the following:

```
Possible site: 49

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.59   Transmembrane   59 -  75 (  52 -  77)
   INTEGRAL    Likelihood = -9.34   Transmembrane    4 -  20 (   1 -  24)

----- 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>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15368 GB:Z99121 yvaL [Bacillus subtilis]
 Identities = 28/72 (38%), Positives = 44/72 (60%), Gaps = 2/72 (2%)

Query: 1  MYNLLLTILLVLSGLLEIAIFMQPQKNPSSNVFDSSGSEALFERTKARGFEAFMQRFTAV 60
          M+ +L+T+L+++S  L I + +Q  K+   +   S G+E LF + KARG +  + R T V
Sbjct: 1  MHAVLITLLVIVSIALIIVVLLQSSKSAGLSGAISGGAEQLFGKQKARGLDLILHRITVV 60

Query: 61 L--VFFWLAIAL 70
          L  +FF L IAL
Sbjct: 61 LAVLFFVLTIAL 72
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2505

A DNA sequence (GASx505) was identified in *S.pyogenes* <SEQ ID 7509> which encodes the amino acid sequence <SEQ ID 7510>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.44   Transmembrane  140 - 156 ( 138 - 156)

----- Final Results -----
               bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF09704 GB:AE001874 glutamine cyclotransferase [Deinococcus radiodurans]
 Identities = 81/229 (35%), Positives = 128/229 (55%), Gaps = 10/229 (4%)

Query: 16  YSYDSNLYTQGLEQLNNNHILLSAGRYGFSKVGVYDL--TQEIFSEKIAFP-DTVFAEGL 72
```

```
          Y +D   +TQGL+ L   H L S G+ G S + V +L   + ++S  +A      F EG
Sbjct: 54  YPHDRAAFTQGLQYLGGGHYLESTGQVGESDLRVSELRGAKVLWSTPLAQALPQAFGEGS 113

Query: 73  TVVEDYFWLLTYKEGVAYKFDKATCNCLGAYPFEGDGWGLAYDKENQCLWMTSGNAFLQK 132
           T +    + LT+++GVA  +D  T    G + ++G+GWGL  D ++  L M++G + L
Sbjct: 114 TQLGSTVYQLTWQDGVALTYDARTFKETGRHRYQGEGWGLTSDGKS--LIMSNGTSTLVW 171

Query: 133 RDPKDFALLDTVLVAIESVPISMLNELEYVDGYLYANIWQTNTIVKLQPDSGKVVATYDI 192
           RDPK FA   +V V  +  P+  LNELEYV G +YAN+W T+ I ++ P +GKV+   D+
Sbjct: 172 RDPKTFAAQRSVQVTDQGQPVRNLNELEYVQGSVYANVWLTDRIARIHPQTGKVLTWIDV 231

Query: 193 SPLLKALNLDKSHYPDL----NVLNGIAHLDQQ-RFLITGKLYPLMLEV 236
           S L + ++   +          +V NGIA + ++   L+TGK +P + EV
Sbjct: 232 SDLTREVSAAATKQGQALTFDDVPNGIAFIPERGTLLLTGKRWPTLFEV 280
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2506

A DNA sequence (GASx506R) was identified in *S.pyogenes* <SEQ ID 7511> which encodes the amino acid sequence <SEQ ID 7512>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2800(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2507

A DNA sequence (GASx507R) was identified in *S.pyogenes* <SEQ ID 7513> which encodes the amino acid sequence <SEQ ID 7514>. Analysis of this protein sequence reveals the following:

```
Possible site: 53

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.51   Transmembrane  103 - 119 (  97 - 124)
   INTEGRAL    Likelihood = -9.13   Transmembrane  126 - 142 ( 122 - 145)
   INTEGRAL    Likelihood = -8.65   Transmembrane  290 - 306 ( 286 - 307)
   INTEGRAL    Likelihood = -7.17   Transmembrane  200 - 216 ( 198 - 228)
   INTEGRAL    Likelihood = -7.06   Transmembrane   58 -  74 (  54 -  82)
   INTEGRAL    Likelihood = -3.19   Transmembrane  223 - 239 ( 220 - 242)
   INTEGRAL    Likelihood = -2.81   Transmembrane  244 - 260 ( 244 - 261)
   INTEGRAL    Likelihood = -2.71   Transmembrane  174 - 190 ( 169 - 191)

----- Final Results -----
            bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB56669 GB:AL121596 putative membrane protein [Streptomyces
          coelicolor A3(2)]
 Identities = 119/322 (36%), Positives = 182/322 (55%), Gaps = 24/322 (7%)

Query: 9   LETIYILIGLQLFHTAYCTFKDKTNPVYFGTALFWGLLGVTFV-----------GGAFL 56
           +E +Y LIGL     A     D++NP  + +A FWGLLG+TF             GG  L
Sbjct: 4   VEWLYWLIGLVFVVMAVQMAMDRSNPKRWTSAAFWGLLGLTFPYGTGVANATAGNGGWTL 63

Query: 57  PNKVIGFIVIVLALLTLFKQVRIGTLPAFNEQKAEESAHRIGNWIFLPVMLMAMISLLLA 116
           P + +G  V+ L +L  F  +  G       ++ E +A R+GN IF+P + + +++++ A
Sbjct: 64  PAEPLGVAVLALIVLAGFNFLGKGVPVTTTGEQREAAAARLGNKIFVPALTIPLVAIVCA 123

Query: 117 LILPDFSKSAIGIAGILA---------TIAILIITKQKPSALLAENNRMNQQVSTSGILP 167
            +L +      G A +L           + +L+  ++K S  +     M + + ++ +LP
Sbjct: 124 SVLDESGLFETGKATLLGLGLGCVAALVVGMLVTGEKKLSVPIHSGRSMLEAMGSALLLP 183

Query: 168 QLLGALGAIFAAAGVGDVIASLIREIVPADSRFFGVLAYVLGMVIFTMIMGNAFAAFTVI 227
           QLL  LG+IFAAAGVGD +  ++ +++P DS++F VLAY +GM +FT+IMGNAFAAF V+
Sbjct: 184 QLLAVLGSIFAAAGVGDQVGDIMNKVLPDDSKYFAVLAYCVGMFLFTVIMGNAFAAFPVM 243

Query: 228 TTGIGVPFVFAL--GADPIIAGALAMTAGFCGTLLTPMAANFNALPVALMEIKDRNAVIK 285
           T  IG P +     G +P +  A+ M AGF GTL TPMAANFN +P  L+E+KD+   IK
Sbjct: 244 TAAIGWPVLIQQMHGNEPAVL-AIGMLAGFAGTLCTPMAANFNIVPATLLELKDQYGPIK 302

Query: 286 KQAPIALVLIISHIALMYLLAY 307
            Q P  + L+     +M L A+
Sbjct: 303 AQLPTGIALLGCCTVIMALFAF 324
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2508

A DNA sequence (GASx508R) was identified in *S.pyogenes* <SEQ ID 7515> which encodes the amino acid sequence <SEQ ID 7516>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.15   Transmembrane  212 - 228 ( 208 - 235)
   INTEGRAL    Likelihood = -8.81   Transmembrane   23 -  39 (  17 -  64)
   INTEGRAL    Likelihood = -7.43   Transmembrane   45 -  61 (  40 -  64)
   INTEGRAL    Likelihood = -1.49   Transmembrane  114 - 130 ( 114 - 130)
   INTEGRAL    Likelihood = -1.49   Transmembrane    3 -  19 (   3 -  20)
   INTEGRAL    Likelihood = -1.49   Transmembrane   76 -  92 (  76 -  92)

----- Final Results -----
              bacterial membrane --- Certainty=0.5861(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB56670 GB:AL121596 possible integral membrane protein
          [Streptomyces coelicolor A3(2)]
 Identities = 77/220 (35%), Positives = 138/220 (62%), Gaps = 2/220 (0%)

Query: 23  IKLIGIVIIVLGFILKCDAIATVVVAGLVTALVSGISFIDFLDILGKEFTNQRLLTIFFI 82
           I L+G+V+++LGF+ + + +  V VAG+VT L+  ++ ++ L   G+ F + R +T++ I
Sbjct: 2   IVLLGVVVVILGFVTRRNPVLVVGVAGIVTGLLGKMNPLEVLAAFGRSFADSRSVTVYAI 61

Query: 83  TLPLIGLSETYGLKHRATQLIQRVQALTVGRLLTLYLIIRELAGLFSIR-LGGHPQFVRP 141
            LP+IGL E YGL+ +A  LI R+  L+ GR LT+YL++++R++   F +  +GG  Q VRP
Sbjct: 62  VLPVIGLLERYGLREQARHLIGRLGKLSAGRFLTVYLLVRQVTAAFGLNSIGGPAQTVRP 121
```

```
Query: 142 LIQPMGEAAAKANIGEELTDAEKDDIKAMAAANENFGNFFAQNTFVGAGGVLLIAGTLEQ 201
           L+ PM EAAA+ + G +L D  ++ +++ +A+ +  G FF ++ F+  G +LLI G +
Sbjct: 122 LVAPMAEAAAAERSTGAKLPDKLREKVRSYSASADTVGVFFGEDCFIAIGSILLITGFVNS 181

Query: 202 LGY-DGNQAKIAFSSILIAIISIIIVAIYNYLFEKKMERQ 240
             + D     ++A  +I +A+ + +I       L +K++ER+
Sbjct: 182 TYHQDIEPTQLALWAIPLAVCAFLIHGARLLLMDKQLERE 221
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2509

A DNA sequence (GASx520) was identified in *S.pyogenes* <SEQ ID 7517> which encodes the amino acid sequence <SEQ ID 7518>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2652(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2510

A DNA sequence (GASx522R) was identified in *S.pyogenes* <SEQ ID 7519> which encodes the amino acid sequence <SEQ ID 7520>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2511

A DNA sequence (GASx523) was identified in *S.pyogenes* <SEQ ID 7521> which encodes the amino acid sequence <SEQ ID 7522>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2133(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2512

A DNA sequence (GASx525) was identified in *S.pyogenes* <SEQ ID 7523> which encodes the amino acid sequence <SEQ ID 7524>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2364(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2513

A DNA sequence (GASx535) was identified in *S.pyogenes* <SEQ ID 7525> which encodes the amino acid sequence <SEQ ID 7526>. Analysis of this protein sequence reveals the following:

```
Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4223(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2514

A DNA sequence (GASx536) was identified in *S.pyogenes* <SEQ ID 7527> which encodes the amino acid sequence <SEQ ID 7528>. Analysis of this protein sequence reveals the following:

```
Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1102(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB85515 GB:AE000874 conserved protein [Methanobacterium
          thermoautotrophicum]
 Identities = 82/236 (34%), Positives = 132/236 (55%), Gaps = 11/236 (4%)

Query: 9   MNLSIFGLKNIPYLKEGDSIEKLIEESIKTSEFFIEDNDVLCIASKVVSIAEGQVMSLNE 68
           M +S+ G++ +P +  GD I  LI  ++       + D D++ IA  +VS AEG ++SL E
Sbjct: 1   MGISLIGVEGMPLVGAGDDIAYLIISALNEGGEDLLDGDIIVIAETIVSKAEGNIISLEE 60

Query: 69  IQVSDVAKEIHRNIPRKDPRIIEIMLNLVNRDLSRLDIKKNYIGCRLENGLKLTSGGIDR 128
           I+ S  A +I      KDP ++E +L          + +  ++I     +G    + GID
Sbjct: 61  IKPSPEALDIAERTG-KDPSLVEAILG---ESSEIIRVGHDFIVSETRHGFVCANAGIDE 116

Query: 129 KSVDEVFL--LPNNPDASAKRISEYLKKSLGKNVAVVITDSDGREDKRGATQVAIGIYGI 186
            +VD+     LP +PD SA++I   L+++ G+ +AV+I+D+ GR  + GA  VA+G+ G+
Sbjct: 117 SNVDDGLATPLPRDPDGSAEKILRTLQEATGRELAVIISDTQGRPFREGAVGVAVGVAGL 176

Query: 187 HPL--RKTEVIDSQGETIKFQEETLCDMIAACAGLVMGQRGTGIPAVLIRGLDYKW 240
            P+  RK E  D  G +++     + D +AA A LVMGQ   G+PAV+IRG  Y W
Sbjct: 177 SPIWDRKGE-RDLYGRSLETTRVAVADELAAAASLVMGQADEGVPAVIIRG--YPW 229
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2515

A DNA sequence (GASx537) was identified in *S.pyogenes* <SEQ ID 7529> which encodes the amino acid sequence <SEQ ID 7530>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.12   Transmembrane  174 - 190 ( 174 - 190)

----- Final Results -----
             bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2516

A DNA sequence (GASx538) was identified in *S.pyogenes* <SEQ ID 7531> which encodes the amino acid sequence <SEQ ID 7532>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3852(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB99212 GB:U67562 conserved hypothetical protein [Methanococcus
          jannaschii]
 Identities = 129/387 (33%), Positives = 208/387 (53%), Gaps = 44/387 (11%)

Query: 18  EVVERKGLGHPDTLADGIAEQIEIDYSLYCLDKFGVIPHHNFDKIIIRGGHSVQDFGGSD 77
           E+VERKGLGHPD++ DGIAE +        ++KFG I HHN D++ + GGH+   FGG
Sbjct: 20  EIVERKGLGHPDSICDGIAESVSRALCKMYMEKFGTILHHNTDQVELVGGHAYPKFGGGV 79

Query: 78  FIEPIKIIFLGRASKKCFNS------SIPLFKIQKKAATKYLNRILPNLDVENYVEFETL 131
            + PI I+  GRA+ +  +        +P+     KAA +YL ++L N+DV+  V  +
Sbjct: 80  MVSPIYILLSGRATMEILDKEKNEVIKLPVGTTAVKAAKEYLKKVLRNVDVDKDVIID-- 137

Query: 132 TSDFTTKTNWFSPEAIEDLP-EYLDVPKANDTATMISYWPLTISEELALMIEGYFYKLD- 189
                +    S + ++    +  +VP ANDT+  + Y PL+ +E L L  E +    +
Sbjct: 138 -----CRIGQGSMDLVDVFERQKNEVPLANDTSFGVGYAPLSTTERLVLETERFLNSDEL 192

Query: 190 KNELPTPRFTKMGGDIKVMVVRNDLEYSIRINFPLISKFFNNDIESQLYVDKHVEKIKKY 249
           KNE+P      +G DIKVM +R   + ++ I   ++ ++  N IE    V   +EK++K
Sbjct: 193 KNEIPA-----VGEDIKVMGLREGKKITLTIAMAVVDRYVKN-IEEYKEV---IEKVRKK 243

Query: 250 IEQKYKNIS--FSIDYH----------YYLTTTGSCIDFGEEGAVGRGNKTHGIISSFR 296
           +E   K I+  + ++ H            YLT TG+  + G++G+VGRGN+ +G+I+ FR
Sbjct: 244 VEDLAKKIADGYEVEIHINTADDYERESVYLTVTGTSAEMGDDGSVGRGNRVNGLITPFR 303

Query: 297 PNTMEAPAGKNCTYFVGKVWGFLSDTIAKEIYEAFNT-PCQIIMQLNIGSKLYRPTHLFI 355
           P +MEA +GKN    VGK++  L++ IA +I +      C + +   IG  +  P  L I
Sbjct: 304 PMSMEAASGKNPVNHVGKIYNILANLIANDIAKLEGVKECYVRILSQIGKPINEPKALDI 363

Query: 356 Q--TEESVD----QERVLEIVNRHLNN 376
           +  TE+S D    + +  EI N+ L+N
Sbjct: 364 EIITEDSYDIKDIEPKAKEIANKWLDN 390
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2517

A DNA sequence (GASx539) was identified in *S.pyogenes* <SEQ ID 7533> which encodes the amino acid sequence <SEQ ID 7534>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1436(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2518

A DNA sequence (GASx540) was identified in *S.pyogenes* <SEQ ID 7535> which encodes the amino acid sequence <SEQ ID 7536>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3956(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD36304 GB:AE001779 conserved hypothetical protein [Thermotoga maritima]
 Identities = 105/353 (29%), Positives = 173/353 (48%), Gaps = 46/353 (13%)

Query: 3   VIGIPTLNEADNISRLVKQIDEYAVNL-GKEIIIINSDSKSTDGTPQIFLETKTYNT-KV 60
           V+GIP+ N A+ IS + +   +  V+    + +I+NSD  S DGT + F+ET T+   K
Sbjct: 106 VVGIPSYNNAETISHVARTAAQGIVDFFDGDGMIVNSDGGSADGTRERFMETDTFGLPKE 165

Query: 61  SIVSEA-KGKGYNVRNIFEYAINHVPNFSGLILIDGDVVSMKKMWLEKMFIAIESGN-DL 118
           S V E   GKG  +R I E+A+    +   ++ +D D+ S+K  W+E++   + G  D
Sbjct: 166 SFVYEGLPGKGSAMRAIMEFALKQ--DAEAVVFLDADLRSVKPWWVERLAGPVLKGEADY 223

Query: 119 IIPNYARKSFEGNATNHFIYPMLVKIFKRDMPYQCISGDFGFSRGLIKDLTLKCN--WHK 176
           + P Y R  F+G  TN+  +PM   ++ + + Q I GDFG  R L++    K    W+
Sbjct: 224 VTPFYLRHRFDGTITNNVCFPMTAVLYGKKVR-QPIGGDFGVGRKLLEIYLGKPKEIWNT 282

Query: 177 YTLGYGIDIFLTLTAILKSYKIKEIDLQSKIH--KKSFEKIEKIFLEVSQSFFETINDNS 234
              +GIDI++T TAI +S ++ +  L +K+H  K   + ++ +FL+V  + FE +
Sbjct: 283 DVARFGIDIWMTTTAINESGRVVQAALGTKVHDVKDPGKHLKGMFLQVVGTLFELV---- 338

Query: 235 LNQDKLRLNINFESHSRQFIKSSDI----------LSSNDIENLKLRALFLLQEEKQY 282
                I +E+  ++  K  D+              S DI NLK  A   L+E +
Sbjct: 339 ---------ITYENVWKEIWKIEDVPIYGETPQEEVPSMSIDIGNLKKLARETLEEVEYI 389

Query: 283 LHG-LSEVEWDGI--LSNTINNIYRYSSEEHSL-------YLLPLYLLRVYNY 325
            G LSEV+  G   LS+ ++ +YR + +           LLP Y  R   +
Sbjct: 390 DRGILSEVKESGTLSLSSWVDTLYRSAVQYRKTRDKKVVENLLPFYFARTARF 442
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2519

A DNA sequence (GASx542) was identified in *S.pyogenes* <SEQ ID 7537> which encodes the amino acid sequence <SEQ ID 7538>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.31   Transmembrane    3 -  19 (   1 -  21)

----- Final Results -----
            bacterial membrane --- Certainty=0.3123(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB07427 GB:AP001519 nucleotide sugar dehydrogenase [Bacillus halodurans]
 Identities = 184/388 (47%), Positives = 274/388 (70%), Gaps = 3/388 (0%)

Query: 1   MKITVVGIGYVGLSIGLLLAKEHDVTFFDIDNKKIDLINKRQSPLKEAAINKLLC-KAKN 59
           M IT+ G GYVGLS  +LLA+ +DV  +DI  +K+D+IN R+SP+ +  I + L  K  N
Sbjct: 1   MNITIAGTGYVGLSNAVLLAQHNDVIAYDIVQEKVDMINNRKSPIVDREIEEFLATKELN 60

Query: 60  INATSSEELAYKDATFIILSLPTNL--KFNKLDTSIIEISVSNILKINKKATIVIKSTVP 117
           + AT+ +E A+KDA F+++S PTN   + N  DTS +E  +S++L IN  A +VIKST+P
Sbjct: 61  LTATTDKEKAFKDAQFVVISTPTNYDPEKNYFDTSSVEAVISDVLSINPNAVMVIKSTIP 120

Query: 118 IGFTEYLRNRFHYNDIIFSPEFLREGSTIHDQLYPSRTIVGNESRNSQLFLDILTDISVE 177
           +G+T  +  RF+  +IIFSPEFLREGS ++D L+PSR +VG  ++ +++F  +L   +++
Sbjct: 121 VGYTREVNERFNTKNIIFSPEFLREGSALYDNLHPSRIVVGERTQRAKIFAALLVQGAIK 180

Query: 178 KDSPSLLVGSSEAEAIKLFSNAYLAQKIAFFNELDTFAEMQNLDSKKIIEAMGYDQRIGN 237
           ++   L   S+EAEAIKLF+N YLA ++AFFNELD++AE++ LD+K+II+ +G D RIG
Sbjct: 181 ENIDVLFTDSTEAEAIKLFANTYLAMRVAFFNELDSYAELKGLDAKQIIDGVGLDPRIGT 240

Query: 238 SHNNPSFGFGGYCLPKDIKQLEYHFKEIPAPIITSISESNLLRKIHIAKMILNSSAKTIG 297
            +NNPSFG+GGYCLPKD KQL  +F+++P  II +I ++N  RK H+A MIL    K +G
Sbjct: 241 HYNNPSFGYGGYCLPKDTKQLLANFEDVPNNIIGAIVDANDTRKDHVANMILKREPKVVG 300

Query: 298 IYRINSKKDSDNCRESSTIDVAKLLKSSGKDVIIFEPLINQKKFLGCPLSNDFNEFIKYS 357
           IYR+  K  SDN R+S+ +DV   L ++G +V+++EP ++  +F G  +  DF EF K S
Sbjct: 301 IYRLTMKTGSDNFRQSAILDVMTRLNNAGAEVVVYEPALDATEFDGSKVIEDFAEFKKMS 360

Query: 358 DIIVANRIDDALRKCNSKVFTRDIFQYD 385
           D+IVANR+ D L++   KV+TRD++  D
Sbjct: 361 DVIVANRLSDDLKEVAEKVYTRDLYTRD 388
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2520

A DNA sequence (GASx544R) was identified in *S.pyogenes* <SEQ ID 7539> which encodes the amino acid sequence <SEQ ID 7540>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.06   Transmembrane   46 -  62 (  46 -  62)

----- 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>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2521

A DNA sequence (GASx545R) was identified in *S.pyogenes* <SEQ ID 7541> which encodes the amino acid sequence <SEQ ID 7542>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.49   Transmembrane  186 - 202 ( 186 - 203)

----- Final Results -----
               bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2522

A DNA sequence (GASx546R) was identified in *S.pyogenes* <SEQ ID 7543> which encodes the amino acid sequence <SEQ ID 7544>. Analysis of this protein sequence reveals the following:

```
Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2422(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2523

A DNA sequence (GASx547R) was identified in *S.pyogenes* <SEQ ID 7545> which encodes the amino acid sequence <SEQ ID 7546>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1612(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2524

A DNA sequence (GASx548) was identified in *S.pyogenes* <SEQ ID 7547> which encodes the amino acid sequence <SEQ ID 7548>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5156(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2525

A DNA sequence (GASx549R) was identified in *S.pyogenes* <SEQ ID 7549> which encodes the amino acid sequence <SEQ ID 7550>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
              bacterial outside --- Certainty=0.3000(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2526

A DNA sequence (GASx552) was identified in *S.pyogenes* <SEQ ID 7551> which encodes the amino acid sequence <SEQ ID 7552>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.59   Transmembrane   83 -  99 (  83 -  99)

----- Final Results -----
             bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2527

A DNA sequence (GASx553) was identified in *S.pyogenes* <SEQ ID 7553> which encodes the amino acid sequence <SEQ ID 7554>. Analysis of this protein sequence reveals the following:

```
Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2781(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2528

A DNA sequence (GASx554) was identified in *S.pyogenes* <SEQ ID 7555> which encodes the amino acid sequence <SEQ ID 7556>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2792(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2529

A DNA sequence (GASx555) was identified in *S.pyogenes* <SEQ ID 7557> which encodes the amino acid sequence <SEQ ID 7558>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.00   Transmembrane   49 -  65 (  49 -  65)

----- Final Results -----
              bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA36631 GB:AB016282 ORF25 [bacteriophage phi-105]
 Identities = 43/118 (36%), Positives = 69/118 (58%), Gaps = 2/118 (1%)

Query: 3   LLDLIGRKRARDKPQNSYEGQDFSYLFG--RTTSGENVDEFKTMQTTAVYACVRVLAEAV 60
            LL+ +  KR+               +FG  +T SGE V E  ++     ++ACV VL++ +
Sbjct: 2   LLERMFEKRSGSSDHEDGFNNILLNMFGGRKTASGERVSESNSLVQPDIFACVNVLSDDI 61

Query: 61  ASLPIHIYERTENGKEKKLDHPLYFLLHDEPNPEMSSFIFRETIMSHLLIWGNAYVQI 118
            A LPIH Y+RT+ G E+K +H      ++  PNP M++F +++ +M+H+L WGNAY  I
Sbjct: 62  AKLPIHTYKRTDGGIERKPEHKSAHAVYARPNPYMTAFTWKKLMMTHVLTWGNAYSYI 119
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2530

A DNA sequence (GASx556) was identified in *S.pyogenes* <SEQ ID 7559> which encodes the amino acid sequence <SEQ ID 7560>. Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2055(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2531

A DNA sequence (GASx557) was identified in *S.pyogenes* <SEQ ID 7561> which encodes the amino acid sequence <SEQ ID 7562>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1696(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2532

A DNA sequence (GASx559) was identified in *S.pyogenes* <SEQ ID 7563> which encodes the amino acid sequence <SEQ ID 7564>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1556(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15798 GB:Z99123 alternate gene name: ipa-83d [Bacillus subtilis]
 Identities = 70/263 (26%), Positives = 121/263 (45%), Gaps = 25/263 (9%)

Query: 68  KTIEQIKELK--YSIDAVACWDEALTHIADDISKELGLNPISSLDSQSFRFKDRMRMVCE 125
           + +EQI ++    +  DA+   +E          + LGL       +++ R K++MR
Sbjct: 87  EVVEQIVKVAEMFGADAITTNNELFIAPMAKACERLGLRGAGVQAAENARDKNKMRDAFN 146

Query: 126 AGGLKMPKYKIINQFSDTNKIINW-KYPLIVKPTSFLASIGVKKVYNFSELQQAVSQMLN 184
             G+K  K K +    D    +     PLI+KPT   +SIGV  + +    +   +++ +
Sbjct: 147 KAGVKSIKNKRVTTLEDFRAALEEIGTPLILKPTYLASSIGVTLITDTETAEDEFNRVND 206

Query: 185 VKFPVYIASGVYELGELYNLEPRVLVEEFIDGE-----------EY-SLESVVRNGIYTP 232
               + +   V          E   + EEF+ GE           +Y S+E ++ +G Y P
Sbjct: 207 YLKSINVPKAV-------TFEAPFIAEEFLQGEYGDWYQTEGYSDYISIEGIMADGEYFP 259

Query: 233 LGITKKIVDEKLFMDEIGHIFPSNLNKEEKSRVYSWAEKLHQILQLNHITTHTEFRIGRN 292
           + I  K    ++   E  HI PS L++E K ++   A+K ++ L L +  THTE ++ +N
Sbjct: 260 IAIHDKT--PQIGFTETSHITPSILDEEAKKKIVEAAKKANEGLGLQNCATHTEIKLMKN 317

Query: 293 GDIILIEIGARIGG-DCIPNLMK 314
            +  LIE  AR  G + IPN+ K
Sbjct: 318 REPGLIESAARFAGWNMIPNIKK 340
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2533

A DNA sequence (GASx561) was identified in *S.pyogenes* <SEQ ID 7565> which encodes the amino acid sequence <SEQ ID 7566>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2602(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2534

A DNA sequence (GASx562) was identified in *S.pyogenes* <SEQ ID 7567> which encodes the amino acid sequence <SEQ ID 7568>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD06696 GB:AE001539 HISTIDYL-TRNA SYNTHETASE [Helicobacter
          pylori J99]
 Identities = 75/309 (24%), Positives = 129/309 (41%), Gaps = 35/309 (11%)

Query: 11  KGYRRQFNQILLGAWGIESAYVDAEIIVATWRGLQRFKGIKVE--FIQLSNKNIFDVLEK 68
           KG  R+F Q      G ES   DAEII      L   K + +E   + ++++ I + + +
Sbjct: 115 KGRYREFTQCDFDFIGSESLVCDAEIIQVIIASL---KALDLEDFCVSINHRKILNGICE 171

Query: 69  DLSKKLRFEDISIEAILGKYLCNNDIEIIKCLYEKDKINMELLISLISKISNKLVKQEFI 128
                E + I   L K   N   E +K   + D   ++ L+ ++     N L   EF
Sbjct: 172 YFGIAQVNEVLRIVDKLEKIGLNGVEEELKKECDLDSNTIKDLLEMVQIKQNDLSHAEFF 231

Query: 129 -KVLVLYEYVKNFLP----VDCIYFSLS------NLY--------GTGHYSSMNYKIFIR 169
            K+  L +Y +N        ++ +Y  L        NLY         G G+Y+ + Y+  +
Sbjct: 232 EKIAYLKDYNENLKKGIQDLERLYQLLGDLQISQNLYKIDFSIARGLGYYTGIVYETTLN 291

Query: 170 TKSGDIFDIADGGRIDDMVSKFNKVNVLGVCMGIGTTVLSQEI-------EYEIEDRIMI 222
                +  +  GGR D +   F+K N+ GV   IG   L   +       E   + +++I
Sbjct: 292 DMKS-LGSVCSGGRYDHLTKNFSKENLQGVGASIGIDRLIVALSEMQLLDERSTQAKVLI 350

Query: 223 LVEKIDVKIYKNCLELANKLSGYHCSVFEFPYKKIKKFFKHELYSRHHYIIVRLDGSMEY 282
                +   Y N L  + + SG    V+    +KIKK F +  +  H ++ V   G  E+
Sbjct: 351 ACMHEEYFSYANRLAESLRQSGIFSEVYP-EAQKIKKPFSYANHKGHEFVAV--IGEEEF 407

Query: 283 RFSSVALKN 291
           +  +++LKN
Sbjct: 408 KSETLSLKN 416
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2535

A DNA sequence (GASx564) was identified in *S.pyogenes* <SEQ ID 7569> which encodes the amino acid sequence <SEQ ID 7570>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1264(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2536

A DNA sequence (GASx576) was identified in *S.pyogenes* <SEQ ID 7571> which encodes the amino acid sequence <SEQ ID 7572>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2537

A DNA sequence (GASx577R) was identified in *S.pyogenes* <SEQ ID 7573> which encodes the amino acid sequence <SEQ ID 7574>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60   Transmembrane    2 -  18 (   1 -  18)

----- Final Results -----
             bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2538

A DNA sequence (GASx579) was identified in *S.pyogenes* <SEQ ID 7575> which encodes the amino acid sequence <SEQ ID 7576>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3161(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12286 GB:Z99106 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 62/140 (44%), Positives = 88/140 (62%), Gaps = 3/140 (2%)

Query: 3   LTNYVQEVSLADFGKPLHHKAYWNKRLKTTGGRFFPKDGHLDFNPRMLEEHGELIFRKIV 62
            L    +++S   F KP  H+A +N RLKTTGGR+     +++ N + L EHG      I+
Sbjct: 6   LQKLTEDISETYFKKPFRHQALFNDRLKTTGGRYLLTSHNIELNRKYLIEHGREELIGII 65

Query: 63  RHELCHYHLYFEGRGYHHKDRDFKDLLAQVNGLRY---VPTSSKSKTNHHYSCQTCGQVY 119
            +HELCHYHL+ EG+GY H+DRDF+ LL QVN  R+   +   +++K  + Y C TCGQ Y
Sbjct: 66  KHELCHYHLHLEGKGYKHRDRDFRMLLQQVNAPRFCTPLKKKAENKKTYMYICTTCGQQY 125

Query: 120 QRKRRINLAKYVCGNCHGKL 139
            +KR +N  +Y CG C GK+
Sbjct: 126 IKKRAMNPDRYRCGKCRGKI 145
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2539

A DNA sequence (GASx587R) was identified in *S.pyogenes* <SEQ ID 7577> which encodes the amino acid sequence <SEQ ID 7578>. Analysis of this protein sequence reveals the following:

```
Possible site: 53

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.40   Transmembrane   46 -  62 (  39 -  89)
   INTEGRAL    Likelihood = -5.36   Transmembrane   65 -  81 (  63 -  89)

----- Final Results -----
               bacterial membrane --- Certainty=0.5161(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2540

A DNA sequence (GASx590R) was identified in *S.pyogenes* <SEQ ID 7579> which encodes the amino acid sequence <SEQ ID 7580>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2036(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2541

A DNA sequence (GASx592R) was identified in *S.pyogenes* <SEQ ID 7581> which encodes the amino acid sequence <SEQ ID 7582>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -4.62   Transmembrane   25 -  41 (  24 -  43)

----- Final Results -----
               bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2542

A DNA sequence (GASx600) was identified in *S.pyogenes* <SEQ ID 7583> which encodes the amino acid sequence <SEQ ID 7584>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.18   Transmembrane    3 -  19 (   2 -  19)

----- Final Results -----
               bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2543

A DNA sequence (GASx603R) was identified in *S.pyogenes* <SEQ ID 7585> which encodes the amino acid sequence <SEQ ID 7586>. Analysis of this protein sequence reveals the following:

```
Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3027(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA03927 GB:AJ000109 gluthatione peroxidase [Lactococcus lactis]
 Identities = 79/133 (59%), Positives = 103/133 (77%)

Query: 1   VVLVVNTATKCGLTPQYQALQALYDTYHDKGFEVLDFPCNQFLNQAPGDAEEINHFCSLT 60
           VV+VVNTA+KCG TPQ++ L+ LY+TY D+G E+L FPCNQF NQ  G+  EIN FC L
Sbjct: 25  VVIVVNTASKCGFTPQFEGLEKLYETYKDQGLEILGFPCNQFANQDAGENTEINEFCQLN 84

Query: 61  YHTTFPRFAKIKVNGKDADPLFTWLKEEKSGPLGKRIEWNFTKFLIDQNGQVIKRYSSKT 120
           Y  TF  F KIKVNGK+A PL+ +LK+E  G L   I+WNFTKFLID++GQVI+R++ KT
Sbjct: 85  YGVTFTMFQKIKVNGKEAHPLYQFLKKEAKGALSGTIKWNFTKFLIDRDGQVIERFAPKT 144

Query: 121 DPKLIEEDLKALL 133
           +P+ +EE++K LL
Sbjct: 145 EPEEMEEEIKKLL 157
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2544

A DNA sequence (GASx605) was identified in *S.pyogenes* <SEQ ID 7587> which encodes the amino acid sequence <SEQ ID 7588>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3687(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2545

A DNA sequence (GASx608R) was identified in *S.pyogenes* <SEQ ID 7589> which encodes the amino acid sequence <SEQ ID 7590>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1327(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2546

A DNA sequence (GASx616) was identified in *S.pyogenes* <SEQ ID 7591> which encodes the amino acid sequence <SEQ ID 7592>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2547

A DNA sequence (GASx617R) was identified in *S.pyogenes* <SEQ ID 7593> which encodes the amino acid sequence <SEQ ID 7594>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0677(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2548

A DNA sequence (GASx622R) was identified in *S.pyogenes* <SEQ ID 7595> which encodes the amino acid sequence <SEQ ID 7596>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.32   Transmembrane    4 -  20 (   1 -  26)

----- Final Results -----
             bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2549

A DNA sequence (GASx632) was identified in *S.pyogenes* <SEQ ID 7597> which encodes the amino acid sequence <SEQ ID 7598>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.40   Transmembrane   83 -  99 (  82 - 102)
   INTEGRAL    Likelihood = -1.28   Transmembrane  108 - 124 ( 108 - 124)

----- Final Results -----
               bacterial membrane --- Certainty=0.2359(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2550

A DNA sequence (GASx638) was identified in *S.pyogenes* <SEQ ID 7599> which encodes the amino acid sequence <SEQ ID 7600>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.64   Transmembrane   12 -  28 (  12 -  28)

----- 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>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2551

A DNA sequence (GASx652R) was identified in *S.pyogenes* <SEQ ID 7601> which encodes the amino acid sequence <SEQ ID 7602>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2622(Affirmative) < succ>
```

```
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA74610 GB:Y14232 hypothetical protein [Bacteriophage TP901-1]
 Identities = 225/485 (46%), Positives = 308/485 (63%), Gaps = 20/485 (4%)

Query: 2   RKVAIYSRVSTINQAEEGYSIQGQIEALTKYCEAMEWKIYKNYSDAGFSGGKLERPAITE 61
           +KVAIY+RVST NQAEEG+SI  QI+ LTKY EAM W++   Y+DAGFSG KLERPA+
Sbjct: 3   KKVAIYTRVSTTNQAEEGFSIDEQIDRLTKYAEAMGWQVSDTYTDAGFSGAKLERPAMQR 62

Query: 62  LIEDGKNNKFDTILVYKLDRLSRNVKDTLYLVKDVFTANNIHFVSLKENIDTSSAMGNLF 121
           LI D +N  FDT+LVYKLDRLSR+V+DTLYLVKDVFT N I F+SL E+IDTSSAMG+LF
Sbjct: 63  LINDIENKAFDTVLVYKLDRLSRSVRDTLYLVKDVFTKNKIDFISLNESIDTSSAMGSLF 122

Query: 122 LTLLSAIAEFEREQIKERMQFGVMNRAKSGKTTAWKTPPYGYRYNKDEKTLSVNELEAAN 181
           LT+LSAI EFERE IKERM  G + RAKSGK+  W     +GY +N+     L +  L+A
Sbjct: 123 LTILSAINEFERENIKERMTMGKLGRAKSGKSMMWTKTAFGYYHNRKTGILEIVPLQATI 182

Query: 182 VRQMFDMIISGCSIMSITNYARDN-FVGN--TWTHVKVKRILENETYKGLVKYREQTFSG 238
           V Q+F   +SG S+  + +   ++  +G    W++  +++ L+N  Y G +K+++  F G
Sbjct: 183 VEQIFTDYLSGISLTKLRDKLNESGHIGKDIPWSYRTLRQTLDNPVYCGYIKFKDSLFEG 242

Query: 239 DHQAIIDEKTYNKAQIALAHRT----DTKTNTRPFQGKYMLSHIAKCGYCGAPLKVCTGR 294
            H+ II  +TY K Q  L  R     +   N RPFQ KYMLS +A+CGYCGAPLK+  G
Sbjct: 243 MHKPIIPYETYLKVQKELEERQQQTYERNNNPRPFQAKYMLSGMARCGYCGAPLKIVLGH 302

Query: 295 AKNDGTRRQTYVCVNKTESLARRSVNNYNNQKICNTGRYEKKHIEKYVIDVLYKLQHDKE 354
           + DG+R   Y C N+      + +  YN+ K C++G Y+  ++E  VID L   Q + +
Sbjct: 303 KRKDGSRTMKYHCANRFPR-KTKGITVYNDNKKCDSGTYDLSNLENTVIDNLIGFQENND 361

Query: 355 YLKKIKKDDN--IIDITPLKKEIEIIDKKINRLNDLYINDLIDLPKLKKDIEELNHLKDD 412
            L KI   +N  I+D +  KK+I  IDKKI + +DLY+ND I + +LK   + L   K
Sbjct: 362 SLLKIINGNNQPILDTSSFKKQISQIDKKIQKNSDLYLNDFITMDELKDRTDSLQAEK-- 419

Query: 413 YNKAIKLNYLDKKNEDSLGML------MDNLDIRKSSYDVQSRIVKQLIDRVEVTMDNID 466
            K +K    + K  DS  +       + ++ I + SYD + +IV  L+ +V+VT DN+D
Sbjct: 420 --KLLKAKISENKFNDSTDVFELVKTQLGSIPINELSYDNKKKIVNNLVSKVDVTADNVD 477

Query: 467 IIFKF 471
           IIFKF
Sbjct: 478 IIFKF 482
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2552

A DNA sequence (GASx653R) was identified in *S.pyogenes* <SEQ ID 7603> which encodes the amino acid sequence <SEQ ID 7604>. Analysis of this protein sequence reveals the following:

```
Possible site: 48

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.22   Transmembrane   86 - 102 (  86 - 102)

----- Final Results -----
              bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF12707 GB:AF066865 unknown [bacteriophage TPW22]
 Identities = 45/67 (67%), Positives = 53/67 (78%), Gaps = 2/67 (2%)

Query: 57  EKEAVRCPKCKSTNVGFMQQGKKTFSVKKAVAGTLLIG--GTVMGFLGEKGKKQWHCNEC 114
           +K A++CPKCKST+V FMQQGKK FSV KAV G +L G  GT+ GF G+KGKKQWHCN C
Sbjct: 138 DKHAIKCPKCKSTDVVFMQQGKKGFSVGKAVGGAVLTGGIGTLAGFAGKKGKKQWHCNNC 197

Query: 115 SCIFETK 121
             +FETK
Sbjct: 198 GRVFETK 204
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2553

A DNA sequence (GASx655) was identified in *S.pyogenes* <SEQ ID 7605> which encodes the amino acid sequence <SEQ ID 7606>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3956(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB63661 GB:AJ251789 Cro protein [Lactobacillus casei
           bacteriophage A2]
 Identities = 43/76 (56%), Positives = 55/76 (71%)

Query: 26  MTINLKRLKAERIASGMTQCEVAQSMGWKTRTPYAKRENGIVSIGADELAKITLIFGLPI 85
           MT+NLKRL+AERIA GM Q E+A++MGW TR+ YAKRENGI +I A EL K+  I G
Sbjct: 1   MTLNLKRLRAERIAKGMNQDEMAKAMGWHTRSSYAKRENGITTISATELVKMASILGYGT 60

Query: 86  EKIAIFFDKDVPVMER 101
            ++ +FF  +VP  ER
Sbjct: 61  NQLDLFFTNNVPDRER 76
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2554

A DNA sequence (GASx656) was identified in *S.pyogenes* <SEQ ID 7607> which encodes the amino acid sequence <SEQ ID 7608>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4505(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2555

A DNA sequence (GASx657) was identified in *S.pyogenes* <SEQ ID 7609> which encodes the amino acid sequence <SEQ ID 7610>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.6593(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2556

A DNA sequence (GASx658) was identified in *S.pyogenes* <SEQ ID 7611> which encodes the amino acid sequence <SEQ ID 7612>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5244(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2557

A DNA sequence (GASx660) was identified in *S.pyogenes* <SEQ ID 7613> which encodes the amino acid sequence <SEQ ID 7614>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1133(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB99331 GB:U67572 purine NTPase [Methanococcus jannaschii]
 Identities = 71/346 (20%), Positives = 154/346 (43%), Gaps = 52/346 (15%)

Query: 8   MSITINKLEIENVK-----RIKAVKIEPSATGLTIIGGNNNQGKTSVLDAIAWAL--GGN 60
           MS+ + ++ + N K     RIK  K      G+  I G N  GK+S+ +A+ +AL   G+
Sbjct: 1   MSMILKEIRMNNFKSHVNSRIKFEK------GIVAIIGENGSGKSSIFEAVFFALFGAGS 54

Query: 61  KYKPSQAMREGSQ---VPPTLKITMSNGLIVERKGKNASLKVIDPNGQ----------KG 107
            +     + +G +   V     ++  +N  I+            +  NG+          K
Sbjct: 55  NFNYDTIITKGKKSVYVELDFEVNGNNYKIIREYDSGRGGAKLYKNGKPYATTISAVNKA 114

Query: 108 GQQLL----DSFVEELAI---NLPKFMDSTPKEKADVLLEIIGVGDQLAELELKEKEIYN 160
             ++L    + F+  + I    + KF+   P EK + + +++G+ D+  +   K  EI
Sbjct: 115 VNEILGVDRNMFLNSIYIKQGEIAKFLSLKPSEKLETVAKLLGI-DEFEKCYQKMGEIVK 173

Query: 161 QRHAIGVIADQKEKFAKEMTYYPDAPKQLVS-ISELIQQHQAILAKNGE-NAQKR--QNV 216
           +          + E+   E+ Y  +  K+L + +S+L ++++ ++  N + N  K+  +++
Sbjct: 174 E------YEKRLERIEGELNYKENYEKELKNKMSQLEEKNKKLMEINDKLNKIKKEFEDI 227

Query: 217 ERIRYDYNQSILEVDRLRKLLADAEAKTNKLSEDLKIANTD------AMDLHDESTAEIE 270
           E++  ++    L  ++    L + +      +++LKI   D      A +  +   E E
Sbjct: 228 EKLFNEWENKKLLYEKFINKLEERKRALELKNQELKILEYDLNTVVEARETLNRHKDEYE 287

Query: 271 ANIADIDEVNRKVRANFDKDKAE-EDAKQQREQYNILTNDIESIRQ 315
              + +DE+ RK+ +   + K+  ED  +  +Q  I+  DIE +++
Sbjct: 288 KYKSLVDEI-RKIESRLRELKSHYEDYLKLTKQLEIIKGDIEKLKE 332
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2558

A DNA sequence (GASx661) was identified in *S.pyogenes* <SEQ ID 7615> which encodes the amino acid sequence <SEQ ID 7616>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1559(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2559

A DNA sequence (GASx662) was identified in *S.pyogenes* <SEQ ID 7617> which encodes the amino acid sequence <SEQ ID 7618>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3292(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2560

A DNA sequence (GASx663) was identified in *S.pyogenes* <SEQ ID 7619> which encodes the amino acid sequence <SEQ ID 7620>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4867(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2561

A DNA sequence (GASx664) was identified in *S.pyogenes* <SEQ ID 7621> which encodes the amino acid sequence <SEQ ID 7622>. Analysis of this protein sequence reveals the following:

```
Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2141(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2562

A DNA sequence (GASx667) was identified in *S.pyogenes* <SEQ ID 7623> which encodes the amino acid sequence <SEQ ID 7624>. Analysis of this protein sequence reveals the following:

```
Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2614(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF80834 GB:AF165214 Orf78 [Pseudomonas phage D3]
 Identities = 68/200 (34%), Positives = 109/200 (54%), Gaps = 10/200 (5%)

Query: 12  GLRFGSLTVINRNRNNSKGGNARWNCLCDCGNKTVVI-GSKLRSGYTKSCGCARKNDNAK 70
           GLR G + V      ++  G  + W C CDCGN+ ++  G+ +R+  T SCGC+R +
Sbjct: 8   GLRVGKVVV--EAFSHCAGKASHWVCRCDCGNRVIMRRGNLMRNRTTTSCGCSRFSH--- 62

Query: 71  GYSSTRLYRIWKGMMNRCYNHKNDNYKYYGGKGISICDEWLTFINFRTWSLSNGYKESLT 130
           G + T  Y  W  M++RC N  N  Y  Y G+GI++C+ W+TF NF       G +   T
Sbjct: 63  GMTGTPTYSSWSNMIDRCTNPSNKRYVDYQGRGITVCERWMTFANFLA---DMGERPDAT 119

Query: 131 -IDRINPKGNYTPLNCRWVSMKMQQNNKTNNRYLSYLGQEYTIAEFSEKLNVTYWTVINQ 189
            +DRI+    Y   NCRW +   Q NN   N ++ YLG+  T+++++ +L +   T+ ++
Sbjct: 120 SLDRIDNDAGYFKENCRWATALEQMNNTRRNTFVEYLGRRQTVSQWAGQLGIPECTLRSR 179

Query: 190 LKLGWSVERIVEEARMKNDR 209
           L  GWS+E  +++   K  R
Sbjct: 180 LNRGWSIEDAMQKPISKQRR 199
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2563

A DNA sequence (GASx668) was identified in *S.pyogenes* <SEQ ID 7625> which encodes the amino acid sequence <SEQ ID 7626>. Analysis of this protein sequence reveals the following:

```
Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1476(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB75598 GB:AJ271879 putative DNA helicase [uncultured
           eubacterium]
 Identities = 42/168 (25%), Positives = 75/168 (44%), Gaps = 7/168 (4%)

Query: 374 IAGPSKAGKSFALIELSIALAEGQKWLG-WQCEQGKVLYVNLELDRPSALHRFKDVYDAM 432
           +  P  AGKS   ++L+  +A G   LG  +   G V+Y+  E D P+A+H       A
Sbjct: 35  LVSPGGAGKSMLALQLAAQIAGGPDLLGVGELPTGPVIYLPAE-DPPTAIHHRLHALGAH 93
```

```
Query: 433 GLPPANVANIDIWNLRGKTVPMDKLAPKLIRRSLKKNYQA---VIIDPIYKVLTGDENSA 489
                A  D   ++     +  +       +LK+  +    +I+D + +    +EN++
Sbjct: 94  LSAEERQAVADGLLIQPLIGSLPNIMASNWFEALKRAAEGRRLMILDTLRRFHIEEENAS 153

Query: 490 DQMAHFTNQFDKVATELGCSVIYCHHHSKGS--QGGKKSMDRASGSGV 535
             MA    + + +A + GCS+++ HH SKG+   G       + GS V
Sbjct: 154 GPMAQVIGRMEAIAADTGCSIVFLHHASKGATMMGAGDQQQASRGSSV 201
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2564

A DNA sequence (GASx669) was identified in *S.pyogenes* <SEQ ID 7627> which encodes the amino acid sequence <SEQ ID 7628>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2555(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2565

A DNA sequence (GASx670) was identified in *S.pyogenes* <SEQ ID 7629> which encodes the amino acid sequence <SEQ ID 7630>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2921(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF74082 GB:AF212845 ORF129 [Lactococcus lactis bacteriophage
            ul36]
 Identities = 36/108 (33%), Positives = 63/108 (58%), Gaps = 1/108 (0%)

Query: 8   IEFFLPMDKIPTTTHQQKKVTVINGKPHFYEPESLKNARDKFTSLLAQHVPPSKLDGPIR 67
           ++F   +DK+PTT  QQK +  + GK  FY+    KN   K   +  +     + + P++
Sbjct: 1   MKFEFELDKMPTT-QQQKGIKKVKGKLQFYDRRGTKNYSLKAQLMKNKPKECFEKNVPLK 59

Query: 68  LTVKWLFPKIKGSTNGQYKTTKPDTDNLQKLLKDCMTELGFWNDDAQV 115
           L+V + +    +     Q+KT++PD DNL K L+D MT+L +++DD+Q+
Sbjct: 60  LSVTFFYAIKQKKRWWQWKTSRPDLDNLMKNLQDYMTKLRYYSDDSQI 107
```