Example 233

A DNA sequence (GBSx0247) was identified in *S.agalactiae* <SEQ ID 741> which encodes the amino acid sequence <SEQ ID 742>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2901(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 743> which encodes the amino acid sequence <SEQ ID 744>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2536(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 = 57/84 (67%), Positives = 73/84 (86%)

Query: 1  MSYEQEFLKDFEEWLQSQISINQMAMDSAKKVLEEDKDERAADAYIRYESKLDAYRFLQG 60
          MSYE+EFLKDFE+W+++QI +NQ+AM ++++V +ED DERA DA+IRYESKLDAY FL G
Sbjct: 1  MSYEKEFLKDFEDWVKTQIQVNQLAMATSQEVAQEDGDERAKDAFIRYESKLDAYEFLLG 60

Query: 61 KFNNYHNQKSFHDLPDGLFGQRHY 84
          KF+NY N K+FHD+PD LFG RHY
Sbjct: 61 KFDNYKNGKAFHDIPDELFGARHY 84
```

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

Example 234

A DNA sequence (GBSx0248) was identified in *S.agalactiae* <SEQ ID 745> which encodes the amino acid sequence <SEQ ID 746>. This protein is predicted to be methyltransferase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2469(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 747> which encodes the amino acid sequence <SEQ ID 748>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3352(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 = 26/60 (43%), Positives = 37/60 (61%)

Query: 23  LKNERCPHPKLINVLERKLEIILGDQKHILEKDSLISLSPQETHHLRAIENSKFLQIELD 82
           +  E  P  K+I VLE +L   L DQK +L ++SLI++  Q+ HHL A  + K LQ+ LD
Sbjct: 42  ISQETSPRDKVILVLEGQLIFDLEDQKQVLTQESLIAIPAQKVHHLEAKTDCKLLQVLLD 101
```

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

Example 235

A DNA sequence (GBSx0249) was identified in *S.agalactiae* <SEQ ID 749> which encodes the amino acid sequence <SEQ ID 750>. This protein is predicted to be integrase (codV). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3842(Affirmative) < succ>
             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 236

A DNA sequence (GBSx0250) was identified in *S.agalactiae* <SEQ ID 751> which encodes the amino acid sequence <SEQ ID 752>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> May be a lipoprotein

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

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

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

SEQ ID 752 (GBS128) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 5; MW 15kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 32 (lane 4; 2 bands).

The GBS128-GST fusion product was purified (Figure 198, lane 2) and used to immunise mice. The resulting antiserum was used for FACS (Figure 288), which confirmed that the protein is immunoaccessible on GBS bacteria.

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

Example 237

A DNA sequence (GBSx0251) was identified in *S.agalactiae* <SEQ ID 753> which encodes the amino acid sequence <SEQ ID 754>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2940(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 755> which encodes the amino acid sequence <SEQ ID 756>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2518(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 = 30/90 (33%), Positives = 49/90 (54%), Gaps = 10/90 (11%)

Query: 3   TVAVRVDDQLKDDATELFQSLGLDMSTAVKMFLIQSVKTQSIPFEIK--------NKSSV 54
           T+ +RVDD +K  A ++ + LG+ MSTA+ MFL Q + T  IPF++          N   +
Sbjct: 15  TLNLRVDDSVKSAADDILKRLGIPMSTAIDMFLNQIILTGGIPFDVSLPEAPQRVNVDYM 74

Query: 55  SDEEFQNLVETKLKGIRVKASDPESVNAFF 84
           S E+F + + T  +     K  +P+ V  F+
Sbjct: 75  SQEKFYDKLITSFED--AKTCNPQDVGKFY 102
```

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

Example 238

A DNA sequence (GBSx0252) was identified in *S.agalactiae* <SEQ ID 757> which encodes the amino acid sequence <SEQ ID 758>. 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
   INTEGRAL     Likelihood = -2.81   Transmembrane  370 - 386 ( 368 - 388)

----- Final Results -----
             bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9593> which encodes amino acid sequence <SEQ ID 9594> was also identified. A related GBS nucleic acid sequence <SEQ ID 10773> which encodes amino acid sequence <SEQ ID 10774> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 759> which encodes the amino acid sequence <SEQ ID 760>. Analysis of this protein sequence reveals the following:

```
      Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -4.57   Transmembrane  354 - 370 ( 353 - 371)

----- 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>

LPXTG motif: 344-348
```

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

```
 Identities = 64/277 (23%), Positives = 99/277 (35%), Gaps = 31/277 (11%)

Query: 126 SIGNLPDLPKGTTVAFETPVDTATPGDKPAKVVVTYPDGSKDTVDVTVKVVDPRTDADKN 185
           ++ +LP   + TT   E PV         + V       + D+ + T     P   A
Sbjct: 121 AVKDLPASTESTTQPVEAPVQETQASASDSMVTGDSTSVTTDSPEETPSSESPVAPALSE 180

Query: 186 DPAGKDQQVNVGETPKAEDSIGNLPDLPKGTTVAFETPVDTATPGDKPAKVVVTYPDGSK 245
            PA    Q    E P    S    P     T  A ETP + A P   P     +    S+
Sbjct: 181 APA----QPAESEEPSVAASSEETPS--PSTPAAPETPEEPAAPSPSPESEEPSVAAPSE 234

Query: 246 DTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPKGTTVAFETPVDT 305
           +T       P   A  + PA  ++      T  +       P  P  +    +TP
Sbjct: 235 ETPSPET----PEEPAAPSQPAESEESSVAATTSPS-------PSTPAESET--QTPPAV 281

Query: 306 ATPGDKPAKVVVTYPDGSKDTVDVTVKVVDPRTDADK----------NDPAGKDQQVNGK 355
               DKP+       P  S    + TV+    +  +DK          N    +  + +
Sbjct: 282 TKDSDKPSSAAEK-PAASSLVSEQTVQQPTSKRSSDKKEEQEQSYSPNRSLSRQVRAHES 340

Query: 356 GNKLPATGENATPFFNVVALTIMSSVGLLSVSKKKED 392
           G  LP+TGE A P F +   +T+MS  G L V+K++++
Sbjct: 341 GKYLPSTGEKAQPLF-IATMTLMSLFGSLLVTKRQKE 376
```

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

Example 239

A DNA sequence (GBSx0253) was identified in *S.agalactiae* <SEQ ID 761> which encodes the amino acid sequence <SEQ ID 762>. 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 240

A DNA sequence (GBSx0254) was identified in *S.agalactiae* <SEQ ID 763> which encodes the amino acid sequence <SEQ ID 764>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.06   Transmembrane   39 -  55 (  39 -  55)

----- 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 GBS nucleic acid sequence <SEQ ID 9591> which encodes amino acid sequence <SEQ ID 9592> was also identified.

The protein differs significantly from U58333 in several places:

```
Query: 157  TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK----------KAED-----PKWDEGSRDK 201
             T PDG  D V+V++ + +     DK D   K          KAED     P   +G+
Sbjct: 683  TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPKGTTVA 742

Query: 202  VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQVNV 258
                +D   T  D   K   T  D +    +VT   K++ PR  DADKNDPAGKDQQVNV
Sbjct: 743  FETPVDTA-TPGDKPAKVVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVNV 798

Query: 157  TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK----------KAED-----PKWDEGSRDK 201
             T PDG  D V+V++ + +     DK D   K          KAED     P   +G+
Sbjct: 841  TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPKGTTVA 900

Query: 202  VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQVNV 258
                +D   T  D   K   T  D +    +VT   K++ PR  DADKNDPAGKDQQVNV
Sbjct: 901  FETPVDTA-TPGDKPAKVVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVNV 956

Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK----------KAED-----PKWDEGSRDK 201
            T PDG  D V+V++ + +     DK D   K          KAED     P   +G+
Sbjct: 288 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPKGTTVA 347

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQVNV 258
                +D   T  D   K   T  D +    +VT   K++ PR  DADKNDPAGKDQQVNV
Sbjct: 348 FETPVDTA-TPGDKPAKVVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVNV 403

Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK----------KAED-----PKWDEGSRDK 201
            T PDG  D V+V++ + +     DK D   K          KAED     P   +G+
Sbjct: 604 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPKGTTVA 663

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQVNV 258
                +D   T  D   K   T  D +    +VT   K++ PR  DADKNDPAGKDQQVNV
Sbjct: 664 FETPVDTA-TPGDKPAKVVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVNV 719

Query: 157 TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK----------KAED-----PKWDEGSRDK 201
            T PDG  D V+V++ + +     DK D   K          KAED     P   +G+
Sbjct: 446 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPKGTTVA 505

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQVNV 258
                +D   T  D   K   T  D +    +VT   K++ PR  DADKNDPAGKDQQVNV
Sbjct: 506 FETPVDTA-TPGDKPAKVVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVNV 561

Query: 157  TKPDGQVDIVNVSLTIYNSSALRDKIDEVKK----------KAED-----PKWDEGSRDK 201
             T PDG  D V+V++ + +     DK D   K          KAED     P   +G+
```

```
Sbjct: 920  TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLPDLPKGTTVA 979

Query: 202  VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQVNV 258
                +D   T  D   K   T  D +    +VT   K++ PR  DADKNDPAGKDQQVNV
Sbjct: 980  FETPVDTA-TPGDKPAKVVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVNV 1035
```

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

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

Example 241

A DNA sequence (GBSx0255) was identified in *S.agalactiae* <SEQ ID 765> which encodes the amino acid sequence <SEQ ID 766>. This protein is predicted to be ara-C-like activator. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane    8 -  24 (   8 -  25)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9589> which encodes amino acid sequence <SEQ ID 9590> was also identified.

There is homology to SEQ ID 460.

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

Example 242

A DNA sequence (GBSx0256) was identified in *S.agalactiae* <SEQ ID 767> which encodes the amino acid sequence <SEQ ID 768>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1200(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 9587> which encodes amino acid sequence <SEQ ID 9588> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 769> which encodes the amino acid sequence <SEQ ID 770>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0679(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 = 135/176 (76%), Positives = 161/176 (90%)

Query: 1   MSYMVKDRQIQKTKVAIYNAFISLLQENDYSKITVQDVIGLANVGRSTFYSHYESKEVLL 60
           +S M KDRQI+KTK AIY+AFI+LLQ+ +YSKITV+D+I LANVGRSTFY+HYESKE+LL
Sbjct: 1   VSDMTKDRQIKKTKTAIYSAFIALLQKKEYSKITVRDMITLANVGRSTFYAHYESKEMLL 60

Query: 61  KELCEDLFHHLFKQGRDVTFEEYLVHILKHFEQNQDSIATLLLSDDPYFLLRFRSELEHD 120
           KELCE+LFHHLF+Q R+VTFE+YLVHILKHFEQN+DSIATLLLS+DPYFLLRF++ELEHD
Sbjct: 61  KELCEELFHHLFRQKRNVTFEDYLVHILKHFEQNKDSIATLLLSNDPYFLLRFKNELEHD 120

Query: 121 VYPRLREEYITKVDIPEDFLKQFLLSSFIETLKWWLHQRQKMTVEDLLKYYLTMVE 176
           VYP LR +YI K  IPE FLKQF+LSSFIETLKWWLHQRQ+M+  +LLKYYL +++
Sbjct: 121 VYPNLRCKYIDKTTIPEVFLKQFVLSSFIETLKWWLHQRQRMSANELLKYYLELIK 176
```

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

Example 243

A DNA sequence (GBSx0257) was identified in *S.agalactiae* <SEQ ID 771> which encodes the amino acid sequence <SEQ ID 772>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3573(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 244

A DNA sequence (GBSx0258) was identified in *S.agalactiae* <SEQ ID 773> which encodes the amino acid sequence <SEQ ID 774>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.19   Transmembrane  112 - 128 ( 107 - 131)
   INTEGRAL    Likelihood = -8.07   Transmembrane   77 -  93 (  71 -  97)
   INTEGRAL    Likelihood = -6.10   Transmembrane  144 - 160 ( 138 - 165)
   INTEGRAL    Likelihood = -3.03   Transmembrane  165 - 181 ( 164 - 182)

----- 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>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 775> which encodes the amino acid sequence <SEQ ID 776>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.13   Transmembrane  112 - 128 ( 107 - 130)
```

```
   INTEGRAL    Likelihood = -5.89   Transmembrane  144 - 160 ( 138 - 163)
   INTEGRAL    Likelihood = -5.47   Transmembrane    7 -  23 (   6 -  29)
   INTEGRAL    Likelihood = -3.50   Transmembrane   77 -  93 (  74 -  94)
   INTEGRAL    Likelihood = -2.07   Transmembrane  166 - 182 ( 165 - 183)

----- Final Results -----
               bacterial membrane --- Certainty=0.4652(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 = 212/287 (73%), Positives = 245/287 (84%)

Query: 1   MTSNKKVAIAFILNISFSVLEFIFGSLFFSGAILADAVHDFGDAIAIGISATLEKKSKKD 60
           M ++KKV I FILN+SFS++EFIFG+LFFSGAILADAVHDFGDAIAIGISA LE+K+ K
Sbjct: 1   MPASKKVTIIFILNLSFSLIEFIFGTLFFSGAILADAVHDFGDAIAIGISAILERKAVKK 60

Query: 61  EDTIFSLGYKRFSLLGALITSLILISGSILVMIENIPKLWHPTPVNYHGMFILAVIAIII 120
           E   FSLGYKRFSLLGAL T+LILISGS+LVMIE IPKLWHPT VNY GMF+LA+ AIII
Sbjct: 61  ESPNFSLGYKRFSLLGALTTNLILISGSLLVMIETIPKLWHPTIVNYDGMFVLAIFAIII 120

Query: 121 NGLASFILHSGQSKHEEILSLHFLEDILGWLAIIVISLILNWKPLYILDPLLSVAISTFI 180
           NG ASFI+HS Q+K+EEILSLHFLEDILGWLAII++SLIL WKP YILDPLLS+AI++FI
Sbjct: 121 NGFASFIIHSNQTKNEEILSLHFLEDILGWLAIIILSLILKWKPWYILDPLLSIAIASFI 180

Query: 181 LSKALPKLLSTLKLFLDGVPDSIDYAALHDELKGLSQVRSINQLNIWSMDGIDNRAIIHC 240
           LSKALPKL++T  +FLDGVPDSIDY  LH EL  L  + S+NQLN+WSMDGID+RA IHC
Sbjct: 181 LSKALPKLVATANIFLDGVPDSIDYCTLHHELSQLPHIVSVNQLNVWSMDGIDHRATIHC 240

Query: 241 CLNQLISEKDCKRAIRTICQHYKINDVTVEIDYSLREHQNHCKPLKN 287
           CL +  +EK CK++IR ICQ Y IN VTVEID SL EHQ+HC  L +
Sbjct: 241 CLRESTTEKHCKKSIRLICQRYNINSVTVEIDTSLNEHQHHCSSLSS 287
```

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

Example 245

A DNA sequence (GBSx0259) was identified in *S.agalactiae* <SEQ ID 777> which encodes the amino acid sequence <SEQ ID 778>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.22   Transmembrane  221 - 237 ( 221 - 237)

----- 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>
```

There is also homology to SEQ ID 780.

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

Example 246

A DNA sequence (GBSx0260) was identified in *S.agalactiae* <SEQ ID 781> which encodes the amino acid sequence <SEQ ID 782>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.50   Transmembrane    2 -  18 (   1 -  18)
```

```
----- 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>
```

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 247

A DNA sequence (GBSx0261) was identified in *S.agalactiae* <SEQ ID 783> which encodes the amino acid sequence <SEQ ID 784>. This protein is predicted to be dehydrogenase (Zn-dependent). Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.77   Transmembrane  171 - 187 ( 170 - 187)

----- 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:AAG20655 GB:AE005134 alcohol dehydrogenase; Adh2 [Halobacterium
           sp. NRC-1]
 Identities = 169/348 (48%), Positives = 232/348 (66%), Gaps = 9/348 (2%)

Query: 1   MKVATFIEPGKMVITDTPKPVIEQETDAVIKIVRACVCGSDLWWYRGISKRESGSFAGHE 60
           M+ A +  PG++ + + PKP IE   DAVI++    VCGSDLW+YRG S RE+GS  GHE
Sbjct: 1   MRAAVYQGPGEIAVEEVPKPDIESPEDAVIRVTHTAVCGSDLWFYRGDSDREAGSRVGHE 60

Query: 61  AIGIVEEVGTKVTDVSKGDFVIVPFTHGCGQCPSCKAGFDGNCTNHQA---AKNVGYQGQ 117
            +GIVEEVG  VT V+ GD VI PF   CG+C  C+ G   +C   ++     N G QG+
Sbjct: 61  PMGIVEEVGDDVTSVAPGDRVIAPFAISCGECEFCRQGLYTSCVEDESWGSEANGGGQGE 120

Query: 118 YLRYTNANWALVKIPGQPSDYDNETLNSLLTLSDVMATGYHAAATAEVKEGDTVVVMGDG 177
           Y++   A+  LV++P + +D D + L SLL L+DVM TG+HAA +A V EGDT VV+GDG
Sbjct: 121 YVKCPFADGTLVRVPDRYAD-DEDVLESLLPLTDVMGTGHHAAVSAGVGEGDTAVVVGDG 179

Query: 178 AVGLCGVIAAKMLGANRIIAMSRHKDRQELALTFGATDIVEERGDEAVKRVLDLTNQAGA 237
           AVGLCGV+AA+ LGA RIIAM  H+DR ELA  FGATD +  RGD+A++R  DLT+  GA
Sbjct: 180 AVGLCGVLAAQRLGAERIIAMGHHEDRLELAAEFGATDTISARGDDAIERARDLTH-GGA 238

Query: 238 DAVLECVGTEQSVDTATQIARPGAVIGRVGIP---QNPDMNTNNLFWKNIGLRGGIASVT 294
           + V+ECVG   ++D+A  IARPG  +G VG+P   ++  ++   +F  NI +RGG+A V
Sbjct: 239 NHVMECVGAASAMDSAIAIARPGGTVGYVGVPYGVEDGGLDVFTMFSDNITIRGGVAPVR 298

Query: 295 TFDKSVLLDAVLTHKINPGLVFTKSFVLDDIQKAYEAMDKRDAIKSLV 342
            F K  LL  V T K+ P LVFTK+ VLDDIQ+AYEAMD+RDAIK LV
Sbjct: 299 AYAEELMADVLQGTLDPSPIFTKTVDLDGVPEGYAAMDDREAIKVLV 345
```

There is also homology to SEQ ID 786.

A related sequence was also identified in GAS <SEQ ID 9145> which encodes the amino acid sequence <SEQ ID 9146>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.41   Transmembrane  170 - 186

----- Final Results -----
               bacterial membrane --- Certainty=0.3166(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 = 121/353 (34%), Positives = 182/353 (51%), Gaps = 16/353 (4%)

Query: 1   MKVATFIEPGKMVITDTPKPVIEQETDAVIKIVRACVCGSDLWWYRG-ISKRESGSFAGH 59
           MK AT++  G + + D PKPVI + TDA++++V+  +CG+DL    G +   + G+  GH
Sbjct: 15  MKAATYLSTGNLQLIDKPKPVIIKPTDAIVQLVKTTICGTDLHILGGDVPACKEGTILGH 74

Query: 60  EAIGIVEEVGTKVTDVSKGDFVIVPFTHGCGQCPSCKAGFDGNCTNHQAAKN---VGYQG 116
           E IGIV+EVG  VT+   GD VI+     C  C  CK G   +C +          G Q
Sbjct: 75  EGIGIVKEVGDAVTNFKIGDKVIISCVTSCHTCYYCKRGLSSHCQDGGWILGHLINGTQA 134

Query: 117 QYLRYTNANWALVKIPGQPSDYDNETLNSLLTLSDVMATGYH-AAATAEVKEGDTVVVMG 175
           +Y+   +A+ +L   P    D       +L+ LSD++ T Y      + VK GD V ++G
Sbjct: 135 EYVHIPHADGSLYHAPDTIDD------EALVMLSDILPTSYEIGVLPSHVKPGDNVCIVG 188

Query: 176 DGAVGLCGVIAAKMLGANRIIAMSRHKDRQELALTFGATDIVEERGDEAVKRVL-DLTNQ 234
            G VGL  ++  +      II +   ++R E A TFGAT  +     E VK ++ D+TN
Sbjct: 189 AGPVGLAALLTVQFFSPANIIMVDLSQNRLEAAKTFGATHTICSGSSEEVKAIIDDITNG 248

Query: 235 AGADAVLECVGTEQSVDTATQIARPGAVIGRVGIPQNP-DMNTNNLFWKNIGLRGGIASV 293
            G D  +ECVG   + D   +I   G  I  VG+   P D N + L+ KNI L  G+ +
Sbjct: 249 RGVDISMECVGYPATFDICQKIISVGGHIANVGVHGKPVDFNLDELWIKNITLNTGLVNA 308

Query: 294 TTFDKSVLLDAVLTHKINPGLVFTKSFVLDDIQKAYEAMDKRDAIKSL-VIVD 345
            T +  +LL+ + T KI+   + T  F L +++KAYE      A  +L VI+D
Sbjct: 309 NTTE--MLLNVLKTGKIDATRLITHHFKLSEVEKAYETFKHAGANNALKVIID 359
```

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

Example 248

A DNA sequence (GBSx0262) was identified in *S.agalactiae* <SEQ ID 787> which encodes the amino acid sequence <SEQ ID 788>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2169(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:AAD36075 GB:AE001762 hypothetical protein [Thermotoga maritima]
 Identities = 55/128 (42%), Positives = 72/128 (55%), Gaps = 8/128 (6%)

Query: 8   IFPKGEKNPYGEFFIGQSYLAALAKSPDG--NVSVGNVTFEAGCRNNWHVHLDGYQILLV 65
           IF +G K    +FF G  ++  L    +G  N  V +V FE G R +WH H  G QIL+V
Sbjct: 5   IFERGSKGS-SDFFTGNVWVKMLVTDENGVFNTQVYDVVFEPGARTHWHSHPGG-QILIV 62

Query: 66  TEGSGWYQEEGKEAVSLKPGDVIVTDKGVRHWHGAKKDSEFAHIAITA----GKSEFYEA 121
           T G G+YQE GK A  LK GDV+     V HWHGA  D E  HI I+     G +E+  +
Sbjct: 63  TRGKGFYQERGKPARILKKGDVVEIPPNVVHWHGAAPDEELVHIGISTQVHLGPAEWLGS 122
```

```
Query: 122 VSDEEYSR 129
           V++EEY +
Sbjct: 123 VTEEEYRK 130
```

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 249

A DNA sequence (GBSx0263) was identified in *S.agalactiae* <SEQ ID 789> which encodes the amino acid sequence <SEQ ID 790>. This protein is predicted to be gamma-carboxymuconolactone decarboxylase. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4089(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:CAA20070 GB:AL031155 3-oxoadipate enol-lactone
           hydrolase/4-carboxymuconolactone decarboxylase
           [Streptomyces coelicolor A3(2)]
 Identities = 33/93 (35%), Positives = 59/93 (62%), Gaps = 1/93 (1%)

Query: 11  QLEEFAPEFARYNDDILFGEVWAKEDHLTDKTRSIITISALISGGNLEQLEHHLQFAKQN 70
           Q +EF+ +F  +      +GE+W +   L  ++RS +T++AL++GG+L++L  HL+ A +N
Sbjct: 349 QADEFSGDFQEFLTRYAWGEIWDRPG-LDRRSRSCVTLTALVAGGHLDELAPHLRAALRN 407

Query: 71  GVTKEEIADIITHLAFYVGWPKAWSAFNKAKEI 103
           G+T  EI +++   A Y G P A  AF  A+++
Sbjct: 408 GLTPGEIKEVLLQAAVYCGVPAANGAFRVAQQV 440
```

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 250

A DNA sequence (GBSx0265) was identified in *S.agalactiae* <SEQ ID 791> which encodes the amino acid sequence <SEQ ID 792>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5529(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 251

A DNA sequence (GBSx0266) was identified in *S.agalactiae* <SEQ ID 793> which encodes the amino acid sequence <SEQ ID 794>. This protein is predicted to be probable transcriptional regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> 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 9585> which encodes amino acid sequence <SEQ ID 9586> was also identified.

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

```
>GP:AAG08263 GB:AE004901 probable transcriptional regulator
           [Pseudomonas aeruginosa]
 Identities = 36/148 (24%), Positives = 68/148 (45%), Gaps = 22/148 (14%)

Query: 5   QIVEKPAMILAG----------------VTLENVKSNQEGIQQAIGICKTQPDFRFD 45
           +IVE+PA  + G                  +  E+  + +   + + GIC  QP+  F
Sbjct: 123 RIVERPAFSVVGMEYFGSAPGDTIGQLWERFIPREHEIAGKHDPEVSYGICAQQPNGEFH 182

Query: 46  YSATYQVETSVQAPKGLEIIRIPSATYAVISVKGPMPSSLQETWRKIIQGFFQENNLKPA 105
           Y A ++V+     P+G+   ++P+  YAV + KG  P  + E+++ I      E  L+P
Sbjct: 183 YVAGFEVQEGWPVPEGMVRFQVPAQKYAVFTHKGTAP-QIAESFQAIYSHLLAERGLEPK 241

Query: 106 NSPNLEIYSSQH--PQDTDYQMEIWLAI 131
             + E Y  +   P D + Q+++++ I
Sbjct: 242 AGVDFEYYDQRFRGPLDPNSQVDLYIPI 269
```

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

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

Example 252

A DNA sequence (GBSx0267) was identified in *S.agalactiae* <SEQ ID 795> which encodes the amino acid sequence <SEQ ID 796>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.0887(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:AAB84919 GB:AE000825 conserved protein [Methanothermobacter
          thermoautotrophicus]
 Identities = 42/130 (32%), Positives = 71/130 (54%), Gaps = 3/130 (2%)
```

```
Query: 1   MITQEMKEIINSQLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG 60
           M+T EM + I  +L  VAT D +G PN+ P     R  D++T +  +N   +T  N+ +N
Sbjct: 1   MMTPEMMDAIEKELVFVATADEEGTPNVVPIGFARPLDERTILIADNYMKKTIRNLHENP 60

Query: 61  KIEIAFVDRERLLGYRFVGTAEIQTEGTYYEAAKKWAEGRMG--VPKAVGIIHVERIFNL 118
           +I +       R   Y+F GT EI    G Y++    +WA+  M     PK+  ++ VE I+++
Sbjct: 61  RIAL-IPQNARECPYQFKGTVEIFKSGKYFDMVVEWAQNVMTELEPKSAILMTVEEIYSV 119

Query: 119 QSGANAGKEI 128
           + G  AG+++
Sbjct: 120 KPGPEAGEKV 129
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 797> which encodes the amino acid sequence <SEQ ID 798>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0789(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 = 123/128 (96%), Positives = 127/128 (99%)

Query: 1   MITQEMKEIINSQLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG 60
           MITQEMK++IN+QLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG
Sbjct: 1   MITQEMKDLINNQLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG 60

Query: 61  KIEIAFVDRERLLGYRFVGTAEIQTEGTYYEAAKKWAEGRMGVPKAVGIIHVERIFNLQS 120
           KIEIAFVDRERLLGYRFVGTAEIQTEG YYEAAKKWA+GRMGVPKAVGIIHVERIFNLQS
Sbjct: 61  KIEIAFVDRERLLGYRFVGTAEIQTEGAYYEAAKKWAQGRMGVPKAVGIIHVERIFNLQS 120

Query: 121 GANAGKEI 128
           GANAGKEI
Sbjct: 121 GANAGKEI 128
```

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

Example 253

A DNA sequence (GBSx0268) was identified in *S.agalactiae* <SEQ ID 799> which encodes the amino acid sequence <SEQ ID 800>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -5.47   Transmembrane 1028 -1044 (1027 -1048)

----- Final Results -----
               bacterial membrane --- Certainty=0.3187(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
!GB:AF054892 surface antigen BspA [Bacteroides forsy...
!GB:AF054892 surface antigen BspA [Bacteroides forsy...
!GB:AF054892 surface antigen BspA [Bacteroides forsy...
!GB:AF054892 surface antigen BspA [Bacteroides forsy...
!GB:AF054892 surface antigen BspA [Bacteroides forsy...

>GP:AAC82625 GB:AF054892 surface antigen BspA [Bacteroides
```

```
          forsythus]
 Identities = 143/566 (25%), Positives = 243/566 (42%), Gaps = 52/566 (9%)

Query: 95  VPKAKPEVTQEASNSSNDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVG----F 150
           +P +   + + A    +    + +P   TA  +  L   T      +    T +G    F
Sbjct: 120 IPNSVTTIGEWAFKGCSGLKSITLPNSLTAIGQSALSGCTGLTSITIPNSVTTIGEWAFF 179

Query: 151 SKSGINKLSQTSHLVLPSHAA--DGTQLTQVASFAFTPDKKTAIAEYTSRLGENGKPSRL 208
             SG+  ++  + L     +A      LT +      PD  T I E   + G +G  S
Sbjct: 180 GCSGLTSITFPNSLTAIGESAFYGCGALTSIT----LPDALTTIGESAFK-GCSGLKSIT 234

Query: 209 DIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQDAFVDNKNIAEVNLPESLETISDYAFA 268
             +    I E   ++   LT +T+P+   +IG+ AF     +  +  P SL TI + AF
Sbjct: 235 FPNSLTTIGESAFYDCGALTSITLPDALTTIGRSAFYGCSGLKSITFPNSLTTIGESAFY 294

Query: 269 HM-SLKQVKLPDNLKVIGELAFFDNQIGGKLYLPRHLIKLAERAFKSNRIQTVEFLGSKL 327
           +  SL   + +P+++   IG  AF+       + LP  L  + ERAF +  + T    + +
Sbjct: 295 NCGSLTSITIPNSVTTIGRSAFYGCSGLKSITLPDGLTTIEERAFYNCGVLTSITIPNSV 354

Query: 328 KVIGEASFQD-NNLRNVMLPDGLEKIESEAFTGNPGDEHYNNQVVLRTRTGQNPHQLATE 386
             IGE++F    + L+++ LPDGL  IE  AF          N   L + T  N      E
Sbjct: 355 ATIGESAFYGCSGLKSITLPDGLTTIEWGAFY---------NCGALTSITIPNSVSTIGE 405

Query: 387 NTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVTGFS---NKGLQKVRRNKNLEIPKQH 443
           + +      +L   T   D     ++ D  +++ +++G        G + V   K  ++ K+
Sbjct: 406 SAFYGCG-ALKDVTVAWDTPIDIQRD-VFRELTLSGIRLHVPAGKKTVYEAK--DVWKEF 461

Query: 444 NGITITEIGDNAFRNVDFQSKTLRKYDLEEIKLPSTIRKIGAFAFQSNNLKSFEASEDLE 503
           N +    + G    + N D  +KTL   +       P T + +   FA  ++ L
Sbjct: 462 NIVEDDDFGGLQW-NYDAATKTLTITN----PTPDTPKPMPNFATPNDQLW--------- 507

Query: 504 EIKEGAFMNNRIGTLDLKDKLIKIGDAAFH-INHIYAIVLPESVQEIGRSAFRQNGALHL 562
              GAF    I  + + D +  +GD AF   + + +I LP+SV  IG+SAF      L
Sbjct: 508 ----GAFQKE-IQKITIGDGVTSVGDFAFSGCDALKSITLPKSVTTIGQSAFSGCWDLRS 562

Query: 563 MFIGNKVKTIGEMAFLSNKLESVNLSEQKQLKTIEVQAFS-DNALSEVVLPPNLQTIREE 621
           + + + V TIGE AF  + LE  +++  K +  I  + F    +L+ + LP  L  I ++
Sbjct: 563 LTLPDGVNTIGEKAFY-DCLELTSITIPKSVTAIGQETFHYCVSLTSLTLPDALTAIGKK 621

Query: 622 AF-KRNHLKEVKGSSTLSQITFNAFD 646
           AF   N L  V    +++ I  NAFD
Sbjct: 622 AFYSCNALTSVTFPKSITTIGENAFD 647
 Identities = 109/407 (26%), Positives = 175/407 (42%), Gaps = 48/407 (11%)

Query: 222 FNAYQLTKLTIPNGYKSIGQDAFVDNKNIAEVNLPESLETISDYAFAHMS-LKQVKLPDN 280
           F+   LT +T+PN   +IG  AF     + + +P S+ TI ++AF   S LK + LP++
Sbjct: 87  FSDCALTSVTLPNSLTAIGDHAFKGCSGLTSITIPNSVTTIGEWAFKGCSGLKSITLPNS 146

Query: 281 LKVIGELAFFDNQIGGKLYLPRHLIKLAERAFKSNRIQTVEFLGSKLKVIGEASFQD-NN 339
           L  IG+ A       + +P  +  + E AF      T     + L  IGE++F    + N
Sbjct: 147 LTAIGQSALSGCTGLTSITIPNSVTTIGEWAFFGCSGLTSITFPNSLTAIGESAFYGCGA 206

Query: 340 LRNVMLPDGLEKIESEAFTGNPGDEHYNNQVVLRTRTGQNPHQLATENTYVNPDKSLWRA 399
           L ++ LPD L  I   AF G  G         L++ T  N      E+ + +
Sbjct: 207 LTSITLPDALTTIGESAFKGCSG---------LKSITFPNSLTTIGESAFYDCGALTSIT 257

Query: 400 TPDMDYTKWLEEDFTYQKNSVTGFSNKGLQKVRRNKNLEIPKQHNGITITEIGDNAFRNV 459
             PD              ++T          K++  P      ++T IG++AF N
Sbjct: 258 LPD---------------ALTTIGRSAFYGCSGLKSITFPN-----SLTTIGESAFYNC 296

Query: 460 DFQSKTLRKYDLEEIKLPSTIRKIGAFAFQS-NNLKSFEASEDLEEIKEGAFMNNRIGT- 517
                L  I +P+++   IG  AF   + LKS     + L  I+E AF N  + T
Sbjct: 297 G---------SLTSITIPNSVTTIGRSAFYGCSGLKSITLPDGLTTIEERAFYNCGVLTS 347

Query: 518 LDLKDKLIKIGDAAFH-INHIYAIVLPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMA 576
           + + + +  IG++AF+  + + +I LP+ +  I   AF   GAL  + I N V TIGE A
Sbjct: 348 ITIPNSVATIGESAFYGCSGLKSITLPDGLTTIEWGAFYNCGALTSITIPNSVSTIGESA 407
```

```
Query: 577 FLS-NKLESVNLSEQKQLKTIEVQAFSDNALSEVVL--PPNLQTIRE 620
           F     L+ V ++    +  I+   F +  LS + L  P   +T+ E
Sbjct: 408 FYGCGALKDVTVAWDTPI-DIQRDVFRELTLSGIRLHVPAGKKTVYE 453
 Identities = 111/465 (23%), Positives = 185/465 (38%), Gaps = 56/465 (12%)

Query: 141 VTRGDTLVGFSKSGINKLSQTSHLVLPSHAADGTQLTQVASFAF----------TPDKKT 190
           +T  D L    +S     S    +  P+       LT +   AF           PD  T
Sbjct: 210 ITLPDALTTIGESAFKGCSGLKSITFPN------SLTTIGESAFYDCGALTSITLPDALT 263

Query: 191 AIAEYTSRLGENGKPSRLDIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQDAFVDNKNI 250
            I   ++  G +G  S    +    I E   +N   LT +TIPN   +IG+ AF     +
Sbjct: 264 TIGR-SAFYGCSGLKSITFPNSLTTIGESAFYNCGSLTSITIPNSVTTIGRSAFYGCSGL 322

Query: 251 AEVNLPESLETISDYAFAHMS-LKQVKLPDNLKVIGELAFFDNQIGGKLYLPRHLIKLAE 309
             + LP+ L TI + AF +   L  + +P+++  IGE AF+      + LP  L  +
Sbjct: 323 KSITLPDGLTTIEERAFYNCGVLTSITIPNSVATIGESAFYGCSGLKSITLPDGLTTIEW 382

Query: 310 RAFKSNRIQTVEFLGSKLKVIGEASFQD-NNLRNVMLP-DGLEKIESEAF-----TGNPG 362
            AF +    T  + + +  IGE++F     L++V + D    I+ + F      +G
Sbjct: 383 GAFYNCGALTSITIPNSVSTIGESAFYGCGALKDVTVAWDTPIDIQRDVFRELTLSGIRL 442

Query: 363 DEHYNNQVVLRTRTGQNPHQLATEN-------TYVNPDKSLWRATPDMDYTKWLEEDFTY 415
             + V   +      +  ++        Y    K+L    P  D  K +  +F
Sbjct: 443 HVPAGKKTVYEAKDVWKEFNIVEDDDFGGLQWNYDAATKTLTITNPTPDTPKPM-PNFAT 501

Query: 416 QKNSVTGFSNKGLQKVRRNKNLEIPKQHNGITITEIGDNAFRNVDFQSKTLRKYDLEEIK 475
             + + G   K +QK+              G  +T +GD AF   D          L+ I
Sbjct: 502 PNDQLWGAFQKEIQKIT-----------IGDGVTSVGDFAFSGCD---------ALKSIT 541

Query: 476 LPSTIRKIGAFAFQSN-NLKSFEASEDLEEIKEGAFMN-NRIGTLDLKDKLIKIGDAAFH 533
           LP ++  IG  AF    +L+S    + +  I E AF +   + ++ +   IG   FH
Sbjct: 542 LPKSVTTIGQSAFSGCWDLRSLTLPDGVNTIGEKAFYDCLELTSITIPKSVTAIGQETFH 601

Query: 534 -INHIYAIVLPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMAF 577
              + ++ LP+++  IG+ AF    AL  +    + TIGE AF
Sbjct: 602 YCVSLTSLTLPDALTAIGKKAFYSCNALTSVTFPKSITTIGENAF 646
 Identities = 98/351 (27%), Positives = 152/351 (42%), Gaps = 53/351 (15%)

Query: 315 NRIQTVEFLGSKLKVIGEASFQDNNLRNVMLPDGLEKIESEAFTGNPGDEHYNNQVVLRT 374
           ++IQTV  +G  +  +G  +F D  L +V LP+ L  I   AF G  G          L +
Sbjct: 68  SKIQTVT-IGDGVTSVGNNAFSDCALTSVTLPNSLTAIGDHAFKGCSG---------LTS 117

Query: 375 RTGQNPHQLATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVTGFSNKGLQKVRRN 434
            T   P+ + T    +     S  ++                  NS+T       L
Sbjct: 118 IT--IPNSVTTIGEWAFKGCSGLKSIT-------------LPNSLTAIGQSALSGCTGL 161

Query: 435 KNLEIPKQHNGITITEIGDNAF------RNVDFQSKTLRKYD--------LEEIKLPSTI 480
            ++ IP      ++T IG+ AF       ++ F +      +         L  I LP  +
Sbjct: 162 TSITIPN-----SVTTIGEWAFFGCSGLTSITFPNSLTAIGESAFYGCGALTSITLPDAL 216

Query: 481 RKIGAFAFQS-NNLKSFEASEDLEEIKEGAFMN-NRIGTLDLKDKLIKIGDAAFH-INHI 537
             IG  AF+  + LKS      L  I E AF +   + ++ L D L  IG  +AF+  + +
Sbjct: 217 TTIGESAFKGCSGLKSITFPNSLTTIGESAFYDCGALTSITLPDALTTIGRSAFYGCSGL 276

Query: 538 YAIVLPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMAFLS-NKLESVNLSEQKQLKTI 596
           +I  P S+  IG SAF   G+L  + I N V TIG  AF   + L+S+ L +   L TI
Sbjct: 277 KSITFPNSLTTIGESAFYNCGSLTSITIPNSVTTIGRSAFYGCSGLKSITLPD--GLTTI 334

Query: 597 EVQAFSD-NALSEVVLPPNLQTIREEAFKR-NHLKEVKGSSTLSQITFNAF 645
           E  +AF +   L+ + +P ++ TI E AF   + LK +     L+ I + AF
Sbjct: 335 EWGAFYNCGALTSITIPNSVATIGESAFYGCSGLKSITLPDGLTTIEWGAF 385
 Identities = 78/282 (27%), Positives = 123/282 (42%), Gaps = 46/282 (16%)

Query: 111 NDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLPS-- 168
           N+AS  E+P      SK +T            VT GD +      +  +  TS + LP+
Sbjct: 56  NNAS--EIPWHSLQSKIQT-----------VTIGDGVTSVGNNAFSDCALTS-VTLPNSL 101

Query: 169 -----HAADG----------TQLTQVASFAFT----------PDKKTAIAEYTSRLGENG 203
```

```
                HA  G            +T +  +AF           P+  TAI + ++  G  G
Sbjct: 102 TAIGDHAFKGCSGLTSITIPNSVTTIGEWAFKGCSGLKSITLPNSLTAIGQ-SALSGCTG 160

Query: 204 KPSRLDIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQDAFVDNKNIAEVNLPESLETIS 263
             S    +    I E   F    LT +T PN   +IG+ AF     +  + LP++L TI
Sbjct: 161 LTSITIPNSVTTIGEWAFFGCSGLTSITFPNSLTAIGESAFYGCGALTSITLPDALTTIG 220

Query: 264 DYAFAHMS-LKQVKLPDNLKVIGELAFFDNQIGGKLYLPRHLIKLAERAFKS-NRIQTVE 321
            + AF   S LK +  P++L  IGE AF+D      + LP  L  +   AF   + ++++
Sbjct: 221 ESAFKGCSGLKSITFPNSLTTIGESAFYDCGALTSITLPDALTTIGRSAFYGCSGLKSIT 280

Query: 322 FLGSKLKVIGEASFQD-NNLRNVMLPDGLEKIESEAFTGNPG 362
            F  S L  IGE++F +  +L ++ +P+ +  I   AF G  G
Sbjct: 281 FPNS-LTTIGESAFYNCGSLTSITIPNSVTTIGRSAFYGCSG 321
 Identities = 43/144 (29%), Positives = 70/144 (47%), Gaps = 4/144 (2%)

Query: 220 EIFNAYQ--LTKLTIPNGYKSIGQDAFVDNKNIAEVNLPESLETISDYAFAHM-SLKQVK 276
            +++ A+Q  + K+TI +G  S+G  AF     +  + LP+S+ TI   AF+    L+ +
Sbjct: 505 QLWGAFQKEIQKITIGDGVTSVGDFAFSGCDALKSITLPKSVTTIGQSAFSGCWDLRSLT 564

Query: 277 LPDNLKVIGELAFFDNQIGGKLYLPRHLIKLAERAFKSNRIQTVEFLGSKLKVIGEASFQ 336
            LPD +  IGE AF+D      + +P+ +  + +  F      T   L   L  IG+ +F
Sbjct: 565 LPDGVNTIGEKAFYDCLELTSITIPKSVTAIGQETFHYCVSLTSLTLPDALTAIGKKAFY 624

Query: 337 D-NNLRNVMLPDGLEKIESEAFTG 359
              N L +V  P  +  I   AF G
Sbjct: 625 SCNALTSVTFPKSITTIGENAFDG 648
 Identities = 43/134 (32%), Positives = 66/134 (49%), Gaps = 12/134 (8%)

Query: 511 MNNRIGTLDLKDKLIKIGDAAFHINHIYAIVLPESVQEIGRSAFRQNGALHLMFIGNKVK 570
            + ++I T+ + D +  +G+ AF     + ++ LP S+  IG  AF+    L  + I N V
Sbjct: 66  LQSKIQTVTIGDGVTSVGNNAFSDCALTSVTLPNSLTAIGDHAFKGCSGLTSITIPNSVT 125

Query: 571 TIGEMAFLS-NKLESVNLSEQKQLKTIEVQAFSD-NALSEVVLPPNLQTIREEAFKRNHL 628
            TIGE AF   + L+S+ L     L  I   A S    L+ + +P ++ TI E AF
Sbjct: 126 TIGEWAFKGCSGLKSITL--PNSLTAIGQSALSGCTGLTSITIPNSVTTIGEWAF----- 178

Query: 629 KEVKGSSTLSQITF 642
               G S L+ ITF
Sbjct: 179 ---FGCSGLTSITF 189
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 801> which encodes the amino acid sequence <SEQ ID 802>. Analysis of this protein sequence reveals the following:

```
     Possible site: 21
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -2.44   Transmembrane  984 -1000 ( 984 -1001)

----- 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>

LPXTG motif: 975-979
```

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

```
 Identities = 751/1050 (71%), Positives = 861/1050 (81%), Gaps = 45/1050 (4%)

Query: 3    KKHLKTLALALTTVSVVTYSQEVYGLEREESVKQEQTQSA-SEDDWFEEDNERKTNVSKE 61
            KKHLKT+AL LTTVSVVT++QEV+ L +E   +KQ Q  S+ S  D+ E    + K  +++
Sbjct: 2    KKHLKTVALTLTTVSVVTHNQEVFSLVKEPILKQTQASSSISGADYAESSGKSKLKINET 61

Query: 62   NSTVDETVSDLFSDGNSNNSSSKTESVVSDPKQVPKAKPEVTQEASNSSNDASKVEVPKQ 121
            +  VD+TV+DLFSD  +     K         +Q  KA  E T E+        S++E  K+
Sbjct: 62   SGPVDDTVTDLFSDKRTTPEKIKDNLAKGPREQELKAVTENT-ESEKQITSGSQLEQSKE 120
```

```
Query: 122  DTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLPSHAADGTQLTQVAS 181
              +  K     TS WE  DF+T+G+TLVG SKSG+ KLSQT HLVLPS AADGTQL QVAS
Sbjct: 121  SLSLNKTVPSTSNWEICDFITKGNTLVGLSKSGVEKLSQTDHLVLPSQAADGTQLIQVAS 180

Query: 182  FAFTPDKKTAIAEYTSRLGENGKPSRLDIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQ 241
            FAFTPDKKTAIAEYTSR GENG+ S+LD+D KEII+EGE+FN+Y L K+TIP GYK IGQ
Sbjct: 181  FAFTPDKKTAIAEYTSRAGENGEISQLDVDGKEIINEGEVFNSYLLKKVTIPTGYKHIGQ 240

Query: 242  DAFVDNKNIAEVNLPESLETISDYAFAHMSLKQVKLPDNLKVIGELAFFDNQIGGKLYLP 301
            DAFVDNKNIAEVNLPESLETISDYAFAH++LKQ+ LPDNLK IGELAFFDNQI GKL LP
Sbjct: 241  DAFVDNKNIAEVNLPESLETISDYAFAHLALKQIDLPDNLKAIGELAFFDNQITGKLSLP 300

Query: 302  RHLIKLAERAFKSNRIQTVEFLGSKLKVIGEASFQDNNLRNVMLPDGLEKIESEAFTGNP 361
            R L++LAERAFKSN I+T+EF G+ LKVIGEASFQDN+L  +MLPDGLEKIESEAFTGNP
Sbjct: 301  RQLMRLAERAFKSNHIKTIEFRGNSLKVIGEASFQDNDLSQLMLPDGLEKIESEAFTGNP 360

Query: 362  GDEHYNNQVVLRTRTGQNPHQLATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVT 421
            GD+HYNN+VVL T++G+NP  LATENTYVNPDKSLW+ +P++DYTKWLEEDFTYQKNSVT
Sbjct: 361  GDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420

Query: 422  GFSNKGLQKVRRNKNLEIPKQHNGITITEIGDNAFRNVDFQSKTLRKYDLEEIKLPSTIR 481
            GFSNKGLQKV+RNKNLEIPKQHNG+TITEIGDNAFRNVDFQ+KTLRKYDLEE+KLPSTIR
Sbjct: 421  GFSNKGLQKVKRNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 480

Query: 482  KIGAFAFQSNNLKSFEASEDLEEIKEGAFMNNRIGTLDLKDKLIKIGDAAFHINHIYAIV 541
            KIGAFAFQSNNLKSFEAS+DLEEIKEGAFMNNRI TL+LKDKL+ IGDAAFHINHIYAIV
Sbjct: 481  KIGAFAFQSNNLKSFEASDDLEEIKEGAFMNNRIETLELKDKLVTIGDAAFHINHIYAIV 540

Query: 542  LPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMAFLSNKLESVNLSEQKQLKTIEVQAF 601
            LPESVQEIGRSAFRQNGA +L+F+G+KVKT+GEMAFLSN+LE ++LSEQKQL  I VQAF
Sbjct: 541  LPESVQEIGRSAFRQNGANNLIFMGSKVKTLGEMAFLSNRLEHLDLSEQKQLTEIPVQAF 600

Query: 602  SDNALSEVVLPPNLQTIREEAFKRNHLKEVKGSSTLSQITFNAFDQNDGDKRFGKKVVVR 661
            SDNAL EV+LP +L+TIREEAFK+NHLK+++ +S LS I FNA D NDGD++F  KVVV+
Sbjct: 601  SDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSHIAFNALDDNDGDEQFDNKVVVK 660

Query: 662  THNNSHMLADGERFIIDPDKLSSTMVDLEKVLKIIEGLDYSTLRQTTQTQFREMTTAGKA 721
            TH+NS+ LADGE FI+DPDKLSST+VDLEK+LK+IEGLDYSTLRQTTQTQFR+MTTAGKA
Sbjct: 661  THHNSYALADGEHFIVDPDKLSSTIVDLEKILKLIEGLDYSTLRQTTQTQFRDMTTAGKA 720

Query: 722  LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIAKAEKALVTKKATKNGHLLERSINKAVL 781
            LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIAKAEKALVTKKATKNG LLERSINKAVL
Sbjct: 721  LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIAKAEKALVTKKATKNGQLLERSINKAVL 780

Query: 782  AYNNSAIKKANVKRLEKELDLLTDLVEGKGPLAQATMVQGVYLLKTPLPLPEYYIGLNVY 841
            AYNNSAIKKANVKRLEKELDLLT LVEGKGPLAQATMVQGVYLLKTPLPLPEYYIGLNVY
Sbjct: 781  AYNNSAIKKANVKRLEKELDLLTGLVEGKGPLAQATMVQGVYLLKTPLPLPEYYIGLNVY 840

Query: 842  FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDNEGYHTLAVATLADYEGLYIKDILN 901
            FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDNEGYH LAVATLADYEGL IK ILN
Sbjct: 841  FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDNEGYHALAVATLADYEGLDIKTILN 900

Query: 902  SSLDKIKAIRQIPLAKYHRLGIFQAIRNAAAEADRLLPKTPKGYLNEVPNYRKKQVEKNL 961
            S L ++ +IRQ+P A YHR GIFQAI+NAAAEA++LLPK
Sbjct: 901  SKLSQLTSIRQVPTAAYHRAGIFQAIQNAAAEAEQLLPK---------------------- 939

Query: 962  KPVDYKTPIFNKALPNEKVDGDRAAKGHNINAETNNSVAVTPIRSEQQLHKSQSDVNLPQ 1021
                                  ++++   + N++          ++S  + ++ +    LP+
Sbjct: 940  ----------------PGTHSEKSSSSESANSKDRG------LQSNPKTNRGRHSAILPR 977

Query: 1022 TSSKNNFIYEILGYVSLCLLFLVTAGKKGK 1051
            T SK +F+Y ILGY S+ LL L+TA KK K
Sbjct: 978  TGSKGSFVYGILGYTSVALLSLITAIKKKK 1007
```

SEQ ID 800 (GBS97) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 12; MW 113.4kDa).

GBS97-His was purified as shown in Figure 193, lane 6.

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

Example 254

A DNA sequence (GBSx0269) was identified in *S.agalactiae* <SEQ ID 803> which encodes the amino acid sequence <SEQ ID 804>. This protein is predicted to be ribonucleoside-diphosphate reductase alpha chain (nrdE). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4274(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:AAB96160 GB:AE000050 ribonucleoside-diphosphate reductase alpha
           chain-MPN324(new), 513(Himmelreich et al., 1996)
           [Mycoplasma pneumoniae]
 Identities = 476/725 (65%), Positives = 586/725 (80%), Gaps = 20/725 (2%)

Query: 2   TQSD--AYLSLNAKTRFRDRTGNYHFTSDKEAVEQYMIEHVEPNTMVFTSLIEKLDYLVS 59
           TQ D  +Y+SLNA T+          F  D  AVE Y+ EHV+P T VF S  E+LD+LV
Sbjct: 12  TQEDLESYISLNAYTKVYG-----DFKMDLHAVEAYIQEHVKPKTKVFHSTKERLDFLVK 66

Query: 60  NNYYESDLLKQYNLEFICQIFEHAYAKKFAFLNFMGALKFYNAYALKTEDNRYYLEHYED 119
           N+YY+ +++  Y+ E   +I   AYA +F + NFMGA KFYNAYALKT D ++YLE+YED
Sbjct: 67  NDYYDENIINMYSFEQFEEITRKAYAYRFRYANFMGAFKFYNAYALKTFDGKWYLENYED 126

Query: 120 RVVMNALFLAAGDEKAAYDLVDDMLANRFQPATPTFLNAGKKRRGEYISCYLLRIEDNME 179
           RVVMN LFLA G+   A  L+  ++ NRFQPATPTFLNAG+K+RGE++SCYLLRIEDNME
Sbjct: 127 RVVMNVLFLANGNYNKALKLLKQIITNRFQPATPTFLNAGRKKRGEFVSCYLLRIEDNME 186

Query: 180 SISRAISTSLQLSKRGGGVALCLTNLREFGAPIKGIKNQATGIVPVMKLLEDSFSYANQL 239
           SI RAI+T+LQLSKR GGVAL LTN+RE GAPIK I+NQ++GI+P+MKLLEDSFSYANQL
Sbjct: 187 SIGRAITTTLQLSKRDGGVALLLTNIRESGAPIKKIENQSSGIIPIMKLLEDSFSYANQL 246

Query: 240 GQRQGAGAVYLHAHHPEVLTFLDTKRENADEKIRIKSLSLGLVIPDITFELAKANKDMAL 299
           GQRQGAGAVYLHAHHP+V+ FLDTKRENADEKIRIKSLSLGLVIPDITF LAK N++MAL
Sbjct: 247 GQRQGAGAVYLHAHHPDVMQFLDTKRENADEKIRIKSLSLGLVIPDITFTLAKNNEEMAL 306

Query: 300 FSPYDIERVYGKPMSDISITEEYETLLANADIRKTFISARKLFQTIAELHFESGYPYILF 359
           FSPYD+   YGKP+SDIS+TE Y  LLAN  I+KTFI+ARK FQT+AELHFESGYPYILF
Sbjct: 307 FSPYDVYEEYGKPLSDISVTEMYYELLANQRIKKTFINARKFFQTVAELHFESGYPYILF 366

Query: 360 EDTVNAKNPHKKEGRIVMSNLCSEIAQVNTASQFSEDLTFTKVGHDVCCNLGSINIARAM 419
           +DTVN +N H     RIVMSNLCSEI Q +T S+F  DL F KVG+D+ CNLGS+NIA+AM
Sbjct: 367 DDTVNRRNAH--PNRIVMSNLCSEIVQPSTPSEFHHDLAFKKVGNDISCNLGSLNIAKAM 424

Query: 420 DQAADFEKLIANSIRALDRVSRTSDLDSAPSIKKGNAANHAVGLGAMNLHGFLATNHIYY 479
           +  +F +L+  +I +LD VSR S+L++APSI+KGN+ NHA+GLGAMNLHGFLATN IYY
Sbjct: 425 ESGPEFSELVKLAIESLDLVSRVSNLETAPSIQKGNSENHALGLGAMNLHGFLATNQIYY 484

Query: 480 DSQEAIDFTDCFFYAMAYYAFKASNHLAKEKGTFEGFSESSYADGSYFYQY--TEQNF-E 536
           +S EAIDFT+ FFY +AY+AFKAS+ LA EKG F+ F  + +ADGSYF +Y   E +F
Sbjct: 485 NSPEAIDFTNIFFYTVAYHAFKASSELALEKGKFKNFENTKFADGSYFDKYIKVEPDFWT 544

Query: 537 PKTQRVKNLLAEYGLTLPSQEDWRKLVQSIKEIGLANAHLLAVAPTGSISYLSSCTPSLQ 596
           PKT+RVK L  +Y + +P++E+W++L  +I++ GLAN+HLLA+APTGSISYLSSCTPSLQ
Sbjct: 545 PKTERVKALFQKYQVEIPTRENWKELALNIQKNGLANSHLLAIAPTGSISYLSSCTPSLQ 604
```

```
Query: 597 PVVSPVEVRKEGALGRVYVPAYKIDADNYVYYKKGAYEVGSEAIINIAAAAQKHIDQAIS 656
           PVVSPVEVRKEG LGR+YVPAY+++ D+Y +YK GAYE+G E IINIAAAAQ+H+DQAIS
Sbjct: 605 PVVSPVEVRKEGRLGRIYVPAYQLNKDSYPFYKDGAYELGPEPIINIAAAAQQHVDQAIS 664

Query: 657 LTLFMTDQATTRDLNKAYIQAFKQKCASIYYVRVRQDILEGSESYDDMLDDFTSSDLEDC 716
           LTLFMTD+ATTRDLNKAYI AFK+ C+SIYYVRVRQ++LE SE +        +  ++ C
Sbjct: 665 LTLFMTDKATTRDLNKAYIYAFKKGCSSIYYVRVRQEVLEDSEDH--------TIQMQQC 716

Query: 717 QSCMI 721
           ++C+I
Sbjct: 717 EACVI 721
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 805> which encodes the amino acid sequence <SEQ ID 806>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1843(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAC82625 GB:AF054892 surface antigen BspA [Bacteroides forsythus]
 Identities = 124/451 (27%), Positives = 202/451 (44%), Gaps = 65/451 (14%)

Query: 221 FNSYLLKKVTIPTGYKHIGQDAFVDNKNIAEVNLPESLETISDYAFAHLA-LKQIDLPDN 279

           F+   L  VT+P     IG  AF     +  + +P S+ TI ++AF   + LK I LP++
Sbjct: 87  FSDCALTSVTLPNSLTAIGDHAFKGCSGLTSITIPNSVTTIGEWAFKGCSGLKSITLPNS 146

Query: 280 LKAIGELAFFDNQITGKLSLPRQLMRLAERA-FKSNHIKTIEFRGNSLKVIGEASFQD-N 337
           L AIG+ A         +++P  +  + E A F  + + +I F  NSL  IGE++F
Sbjct: 147 LTAIGQSALSGCTGLTSITIPNSVTTIGEWAFFGCSGLTSITF-PNSLTAIGESAFYGCG 205

Query: 338 DLSQLMLPDGLEKIESEAFTGNPGDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQ 397
            L+ + LPD L  I   AF G  G            KS   P+ L T       +S +
Sbjct: 206 ALTSITLPDALTTIGESAFKGCSG----------LKSITFPNSLTTIG------ESAFY 248

Query: 398 ESPEIDYTKWLEEDFTYQKNSVTGFSNKGLQKVKRNKNLEIPKQHNGVTITEIGDNAFRN 457
           +   +      +   T  +++  G S  GL      K++  P      ++T IG++AF N
Sbjct: 249 DCGALTSITLPDALTTIGRSAFYGCS--GL------KSITFPN-----SLTTIGESAFYN 295

Query: 458 VDFQNKTLRKYDLEEVKLPSTIRKIGAFAFQS-NNLKSFEASDDLEEIKEGAFMNNRIET 516
                    L  + +P+++  IG  AF   + LKS    D L  I+E AF N  + T
Sbjct: 296 CG---------SLTSITIPNSVTTIGRSAFYGCSGLKSITLPDGLTTIEERAFYNCGVLT 346

Query: 517 -LELKDKLVTIGDAAFH-INHIYAIVLPESVQEIGRSAFRQNGANNLIFMGSKVKTLGEM 574
            + + + + TIG++AF+  + + +I LP+ +  I   AF   GA   I + + V T+GE
Sbjct: 347 SITIPNSVATIGESAFYGCSGLKSITLPDGLTTIEWGAFYNCGALTSITIPNSVSTIGES 406

Query: 575 AFLS-NRLEHLDLSEQKQLTEIPVQAFSDNALKEVLL--PASLKTIREEAFKKNHLKQLE 631
           AF     L+ + ++    + +I    F +  L + L  PA  KT+ E    K+  K+
Sbjct: 407 AFYGCGALKDVTVAWDTPI-DIQRDVFRELTLSGIRLHVPAGKKTVYE---AKDVWKE-- 460

Query: 632 VASALSHIAFNALDDND-GDEQFDNKVVVKT 661
                  FN ++D+D G  Q++      KT
Sbjct: 461 ---------FNIVEDDDFGGLQWNYDAATKT 482
```

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

```
 Identities = 534/726 (73%), Positives = 614/726 (84%), Gaps = 5/726 (0%)

Query: 1   MTQSDA-YLSLNAKTRFRDRTGNYHFTSDKEAVEQYMIEHVEPNIMVFTSLIEKLDYLVS 59
```

```
           M+Q++A YLSLNA TRF+   G+YHF SDKEAV +Y+ EHV PN M F SL +KL YL++
Sbjct: 1   MSQTNASYLSLNALTRFKKPDGSYHFDSDKEAVRRYLEEHVSPNQMAFNSLEDKLAYLIN 60

Query: 60  NNYYESDLLKQYNLEFICQIFEHAYAKKFAFLNFMGALKFYNAYALKTEDNRYYLEHYED 119
             YYE  +   Y  + I + F +AY + + FLN MGA+KFY +YALKT D + YLE +ED
Sbjct: 61  EGYYEQAIFDAYPNDLIKEAFHYAYQQGYRFLNLMGAMKFYQSYALKTLDGKQYLETFED 120

Query: 120 RVVMNALFLAAGDEKAAYDLVDDMLANRFQPATPTFLNAGKKRRGEYISCYLLRIEDNME 179
           R VMNALFLA GD+   +D++D +L  RFQPATPTFLNAGKKRRGEYISCYLLR+EDNME
Sbjct: 121 RAVMNALFLADGDQTFVFDVIDAILHRRFQPATPTFLNAGKKRRGEYISCYLLRVEDNME 180

Query: 180 SISRAISTSLQLSKRGGGVALCLTNLREFGAPIKGIKNQATGIVPVMKLLEDSFSYANQL 239
           SISRAISTSLQLSKRGGGVALCLTNLRE GAPIKGI+NQATGIVPVMKLLEDSFSYANQL
Sbjct: 181 SISRAISTSLQLSKRGGGVALCLTNLREIGAPIKGIENQATGIVPVMKLLEDSFSYANQL 240

Query: 240 GQRQGAGAVYLHAHHPEVLTFLDTKRENADEKIRIKSLSLGLVIPDITFELAKANKDMAL 299
           GQRQGAGAVYLHAHHPEVLTFLDTKRENADEKIRIKSL+LGLVIPDITF+LAK NKDMAL
Sbjct: 241 GQRQGAGAVYLHAHHPEVLTFLDTKRENADEKIRIKSLALGLVIPDITFQLAKENKDMAL 300

Query: 300 FSPYDIERVYGKPMSDISITEEYETLLANADIRKTFISARKLFQTIAELHFESGYPYILF 359
           FSPYDI+R YGK MSDISITEEY+ LLAN  I+KT+ISARK FQ IAELHFESGYPY+LF
Sbjct: 301 FSPYDIKRAYGKDMSDISITEEYDKLLANPAIKKTYISARKFFQLIAELHFESGYPYLLF 360

Query: 360 EDTVNAKNPHKKEGRIVMSNLCSEIAQVNTASQFSEDLTFTKVGHDVCCNLGSINIARAM 419
           +DTVN +NPH K+GRIVMSNLCSEIAQV+T S F EDL+F  +G D+CCNLGSINIA+AM
Sbjct: 361 DDTVNKRNPHAKKGRIVMSNLCSEIAQVSTPSTFKEDLSFETIGEDICCNLGSINIAQAM 420

Query: 420 DQAADFEKLIANSIRALDRVSRTSDLDSAPSIKKGNAANHAVGLGAMNLHGFLATNHIYY 479
             A  FE+LI  SIRALDRVSR SDL+ APS++ GNAANHAVGLGAMNLHGFLATNHIYY
Sbjct: 421 ADAPHFEQLITTSIRALDRVSRVSDLNCAPSVETGNAANHAVGLGAMNLHGFLATNHIYY 480

Query: 480 DSQEAIDFTDCFFYAMAYYAFKASNHLAKEKGTFEGFSESSYADGSYFYQYTEQNFEPKT 539
           D++EA+DFTD FF+AMAYYAFKAS  LAKEKG F GFS S+Y+DG+YF +Y +++ +P+T
Sbjct: 481 DTKEAVDFTDLFFHAMAYYAFKASCQLAKEKGAFAGFSLSTYSDGTYFAKYLQEDAKPQT 540

Query: 540 QRVKNLLAEYGLTLPSQEDWRKLVQSIKEIGLANAHLLAVAPTGSISYLSSCTPSLQPVV 599
            +V  LL +YG TLP+  DW+ LV  IK+ GLANAHLLAVAPTGSISYLSSCTPSLQPVV
Sbjct: 541 AKVATLLQDYGFTLPTVADWQALVADIKQFGLANAHLLAVAPTGSISYLSSCTPSLQPVV 600

Query: 600 SPVEVRKEGALGRVYVPAYKIDADNYVYYKKGAYEVGSEAIINIAAAAQKHIDQAISLTL 659
           +PVEVRKEG+LGR+YVPAY+ID  NY YY++GAYEVG +AII++ AAAQKH+DQAISLTL
Sbjct: 601 APVEVRKEGSLGRIYVPAYQIDQANYAYYERGAYEVGPKAIIDVVAAAQKHVDQAISLTL 660

Query: 660 FMTDQATTRDLNKAYIQAFKQKCASIYYVRVRQDILEGSESYDD----MLDDFTSSDLED 715
           FMTDQATTRDLN++YIQAFKQ CASIYYVRVRQD+L GSE YD+          +     +
Sbjct: 661 FMTDQATTRDLNRSYIQAFKQNCASIYYVRVRQDVLAGSEQYDEDSLVTAPGASDETTTE 720

Query: 716 CQSCMI 721
           CQSCMI
Sbjct: 721 CQSCMI 726
```

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

Example 255

A DNA sequence (GBSx0270) was identified in *S.agalactiae* <SEQ ID 807> which encodes the amino acid sequence <SEQ ID 808>. This protein is predicted to be nrdI protein (nrdI). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2952(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:AAC71451 GB:U39702 nrdI protein (nrdI) [Mycoplasma genitalium]
 Identities = 77/127 (60%), Positives = 104/127 (81%), Gaps = 1/127 (0%)

Query: 7   VVYFSSKSNNTHRFVQKLACSNQRIPSD-GSSILVTEDYILIVPTYAGGGDDTKGAVPKQ 65
           +VYFSS SNNTHRF++KL   ++RIP D   SI V+ +Y+LI PTY+GGG+  +GAVPKQ
Sbjct: 22  IVYFSSISNNTHRFIEKLGFQHKRIPVDITQSITVSNEYVLICPTYSGGGNQVEGAVPKQ 81

Query: 66  VVQFLNVRQNREHCQGVISSGNTNFGDTYAIAGPIIARKLNVPLLHQFELLGTQEDVTRV 125
           V+QFLN + NRE C+GVI+SGNTNFGDT+ +AG +I++KLNVPLL+QFELLGT+ DV +
Sbjct: 82  VIQFLNNKHNRELCRGVIASGNTNFGDTFCLAGTVISKKLNVPLLYQFELLGTKNDVEQT 141

Query: 126 KELLCQF 132
           ++++  F
Sbjct: 142 QKIIANF 148
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 809> which encodes the amino acid sequence <SEQ ID 810>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0089(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 = 84/125 (67%), Positives = 100/125 (79%)

Query: 7   VVYFSSKSNNTHRFVQKLACSNQRIPSDGSSILVTEDYILIVPTYAGGGDDTKGAVPKQV 66
           +VYFSSKSNNTHRFVQKL    QRIP D   + V+  Y+LIVPTYA GG D KGAV KQV
Sbjct: 6   IVYFSSKSNNTHRFVQKLGLPAQRIPVDNRPLEVSTHYLLIVPTYAAGGSDAKGAVSKQV 65

Query: 67  VQFLNVRQNREHCQGVISSGNTNFGDTYAIAGPIIARKLNVPLLHQFELLGTQEDVTRVK 126
           ++FLN   NR+HC+GVISSGNTNFGDT+A+AGPII++KL VPLLHQFELLGT  DV +V+
Sbjct: 66  IRFLNNPNNRKHCKGVISSGNTNFGDTFALAGPIISQKLQVPLLHQFELLGTATDVKKVQ 125

Query: 127 ELLCQ 131
            +  +
Sbjct: 126 AIFAR 130
```

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

Example 256

A DNA sequence (GBSx0271) was identified in *S.agalactiae* <SEQ ID 811> which encodes the amino acid sequence <SEQ ID 812>. This protein is predicted to be ribonucleoside-diphosphate reductase beta chain (nrdF). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3889(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:AAB96162 GB:AE000050 ribonucleoside-diphosphate reductase beta
          chain [Mycoplasma pneumoniae]
 Identities = 261/335 (77%), Positives = 301/335 (88%)

Query: 2   QSYYDRSQSPLDYALSEKAFPMRSVNWNKLNDDKDLEVWNRVTQNFWLPEKIPVSNDLNS 61
           + Y+  S SPL+YA  +     +RSVNWN ++D+KDLEVWNR+TQNFWLPEKIPVSND+ S
Sbjct: 5   KKYFLESVSPLEYAQKKFQGNLRSVNWNLVDDEKDLEVWNRITQNFWLPEKIPVSNDIPS 64

Query: 62  WRTLDADWQQLITRTFTGLTLLDSVQATVGDIAQIKHSQTDHEQVIYANFAFMVAIHARS 121
           W+ L  +WQ LIT+TFTGLTLLD++QAT+GDI QI ++ TDHEQVIYANFAFMV +HARS
Sbjct: 65  WKQLSKEWQDLITKTFTGLTLLDTIQATIGDIKQIDYALTDHEQVIYANFAFMVGVHARS 124

Query: 122 YGTIFSTLCTSQQIEEAHEWVVDTESLQARSRILIPFYTGDDPLKSKVAAAMMPGFLLYG 181
           YGTIFSTLCTS+QI EAHEWVV TESLQ R++ LIP+YTG DPLKSKVAAA+MPGFLLYG
Sbjct: 125 YGTIFSTLCTSEQITEAHEWVVKTESLQKRAKALIPYYTGKDPLKSKVAAALMPGFLLYG 184

Query: 182 GFYLPFYLSARGKLPNTSDIIRLILRDKVIHNYYSGYKYQQKVAKLSVEKQAEMKTFVFD 241
           GFYLPFYLS+R +LPNTSDIIRLILRDKVIHNYYSGYK+Q+KV K+S EKQAEMK FVFD
Sbjct: 185 GFYLPFYLSSRKQLPNTSDIIRLILRDKVIHNYYSGYKFQRKVEKMSKEKQAEMKRFVFD 244

Query: 242 LLYQLIDLEKAYLYELYDGFDLAEDAIRFSIYNAGKFLQNLGYDSPFTEEETRISPEVFA 301
           L+Y+LI+LEKAYL ELY+GF + EDAI+FSIYNAGKFLQNLGYDSPFTEEETRI PE+FA
Sbjct: 245 LMYELIELEKAYLKELYEGFGIVEDAIKFSIYNAGKFLQNLGYDSPFTEEETRIKPEIFA 304

Query: 302 QLSARADENHDFFSGNGSSYIMGITEETLDEDWEF 336
           QLSARADENHDFFSGNGSSY+MGI+EET D+DW+F
Sbjct: 305 QLSARADENHDFFSGNGSSYVMGISEETEDKDWDF 339
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 813> which encodes the amino acid sequence <SEQ ID 814>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3779(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 = 292/335 (87%), Positives = 318/335 (94%)

Query: 2   QSYYDRSQSPLDYALSEKAFPMRSVNWNKLNDDKDLEVWNRVTQNFWLPEKIPVSNDLNS 61
           Q YY+RSQSP++YALSE     +RS+NWN LNDDKDLEVWNRVTQNFWLPEK+PVSNDLNS
Sbjct: 3   QHYYERSQSPIEYALSETQKQLRSINWNYLNDDKDLEVWNRVTQNFWLPEKVPVSNDLNS 62

Query: 62  WRTLDADWQQLITRTFTGLTLLDSVQATVGDIAQIKHSQTDHEQVIYANFAFMVAIHARS 121
           WR+L  DWQQLITRT+TGLTLLD+VQATVGD+AQI+HSQTDHEQVIY NFAFMV IHARS
Sbjct: 63  WRSLGEDWQQLITRTYTGLTLLDTVQATVGDVAQIQHSQTDHEQVIYTNFAFMVGIHARS 122

Query: 122 YGTIFSTLCTSQQIEEAHEWVVDTESLQARSRILIPFYTGDDPLKSKVAAAMMPGFLLYG 181
           YGTIFSTLC+S+QIEEAHEWVV T+SLQ R+R+LIP+YTGDDPLKSKVAAAMMPGFLLYG
Sbjct: 123 YGTIFSTLCSSEQIEEAHEWVVSTQSLQDRARVLIPYYTGDDPLKSKVAAAMMPGFLLYG 182

Query: 182 GFYLPFYLSARGKLPNTSDIIRLILRDKVIHNYYSGYKYQQKVAKLSVEKQAEMKTFVFD 241
           GFYLPFYLSARGK+PNTSDIIRLILRDKVIHNYYSGYKYQQKVA+LS EKQAEMK FVFD
Sbjct: 183 GFYLPFYLSARGKMPNTSDIIRLILRDKVIHNYYSGYKYQQKVARLSPEKQAEMKAFVFD 242

Query: 242 LLYQLIDLEKAYLYELYDGFDLAEDAIRFSIYNAGKFLQNLGYDSPFTEEETRISPEVFA 301
           LLY+LIDLEKAYL ELY GFDLAEDAIRFS+YNAGKFLQNLGY+SPFT+EETR+SPEVFA
Sbjct: 243 LLYELIDLEKAYLRELYAGFDLAEDAIRFSLYNAGKFLQNLGYESPFTDEETRVSPEVFA 302

Query: 302 QLSARADENHDFFSGNGSSYIMGITEETLDEDWEF 336
           QLSARADENHDFFSGNGSSY+MGITEET D+DWEF
Sbjct: 303 QLSARADENHDFFSGNGSSYVMGITEETTDDDWEF 337
```

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

Example 257

A DNA sequence (GBSx0272) was identified in *S.agalactiae* <SEQ ID 815> which encodes the amino acid sequence <SEQ ID 816>. This protein is predicted to be rhamnosyltransferase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1741(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 9583> which encodes amino acid sequence <SEQ ID 9584> was also identified.

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

```
>GP:BAA32090 GB:AB010970 rhamnosyltransferase [Streptococcus mutans]
 Identities = 104/309 (33%), Positives = 173/309 (55%), Gaps = 21/309 (6%)

Query: 11  QINICLATYNGQKYLRQQLDSIIQQGYTDWICLIRDDGSTDDTVAIIKEYVNRDSRFIFI 70
           ++NI ++TYNGQ+++ QQ+ SI +Q + +W  LIRDDGS+D T  II ++   D+R  FI
Sbjct: 2   KVNILMSTYNGQEFIAQQIQSIQKQTFENWNLLIRDDGSSDGTPKIIADFAKSDARIRFI 61

Query: 71  NSNDDRKLGSHRSFYELVNYKKADFYVFSDQDDVWKENRLERYLEEAEKFNQELPLLVYS 130
           N++     G  ++FY L+ Y+KAD+Y FSDQDDVW   +LE  L   EK N ++PL+VY+
Sbjct: 62  NADKRENFGVIKNFYTLLKYEKADYYFFSDQDDVWLPQKLELTLASVEKENNQIPLMVYT 121

Query: 131 NWTSVDEKLTVL-------KEHNPATVIQEQIAFNQINGMVIMMNHELAKLWE--YRQIG 181
           + T VD  L VL       + H+  T + E++  N + G  +M+NH LAK W+  Y  +
Sbjct: 122 DLTVVDRDLQVLHDSMIKTQSHHANTSLLEELTENTVTGGTMMVNHCLAKQWKQCYDDLI 181

Query: 182 AHDSYVGTLAYAVGNVAYISDSTVLWRRQ----VGAES----LNNYGRQYG-VATFWQMI 232
            HD Y+  LA ++G + Y+ ++T L+R+     +GA +    L N+ R +  V  +W ++
Sbjct: 182 MHDWYLALLAASLGKLIYLDETTELYRQHESNVLGARTWSKRLKNWLRPHRLVKKYWWLV 241

Query: 233 NTSFDRASLIFAQVSDKMSLERKLFFSRFIELKNANLMRRIYLLSKLKLRRKSLKETVAM 292
            +S  +AS +   +   +    K     ++ L + + + RI  L +    +     T
Sbjct: 242 TSSQQQASHL---LELDLPAANKAIIRAYVTLLDQSFLNRIKWLKQYGFAKNRAFHTFVF 298

Query: 293 TILLLTGYG 301
             L++T +G
Sbjct: 299 KTLIITKFG 307
```

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

Example 258

A DNA sequence (GBSx0273) was identified in *S.agalactiae* <SEQ ID 819> which encodes the amino acid sequence <SEQ ID 820>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.19   Transmembrane 1213 -1229 (1211 -1230)

----- Final Results -----
               bacterial membrane --- Certainty=0.2678(Affirmative) < succ>
```

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

A related GBS nucleic acid sequence <SEQ ID 9581> which encodes amino acid sequence <SEQ ID 9582> was also identified.

There is also homology to SEQ ID 822.

A related GBS gene <SEQ ID 8525> and protein <SEQ ID 8526> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
SRCFLG: 0
McG: Length of UR:   3
     Peak Value of UR:   2.28
     Net Charge of CR: 4
McG: Discrim Score:      1.29
GvH: Signal Score (-7.5): 2.84
     Possible site: 30
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 31
ALOM program   count: 0 value:   1.16 threshold:  0.0
   PERIPHERAL  Likelihood =  1.16     344
 modified ALOM score:  -0.73

*** 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>

LPXTG motif: 1197-1201
```

SEQ ID 8526 (GBS147) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 4; MW 132kDa).

The GBS147-His fusion product was purified (Figure 200, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 286), 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 259

A DNA sequence (GBSx0274) was identified in *S.agalactiae* <SEQ ID 823> which encodes the amino acid sequence <SEQ ID 824>. This protein is predicted to be Acetyltransferase (GNAT) family. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2781(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:AAG03505 GB:AE004449 conserved hypothetical protein [Pseudomonas aeruginosa]
```

```
 Identities = 66/143 (46%), Positives = 94/143 (65%), Gaps = 5/143 (3%)

Query: 2   WNVKTFDNLTTHELFQIYKLRVSVFVVEQDCPYQEVDDEDLI--CLHGMNWVDGQLAAYY 59
           W  K   +LT  EL+ + +LR  VFVVEQ CPYQEVD  DL+    H M W DGQL AY
Sbjct: 5   WTCKHHADLTLKELYALLQLRTEVFVVEQKCPYQEVDGLDLVGDTHHLMAWRDGQLLAYL 64

Query: 60  RLIP---EDDKVHLGRVIVNPDFRKKGLGNQLVEYAIKFSEANYPNKPIYAQAQAYLQDF 116
           RL+     + +V +GRV+ +   R +GLG+QL+E A++ +E  + + P+Y  AQA+LQ +
Sbjct: 65  RLLDPVRHEGQVVIGRVVSSSAARGQGLGHQLMERALQAAERLWLDTPVYLSAQAHLQAY 124

Query: 117 YQSFGFQPVSDIYLEDNIPHLDM 139
           Y  +GF  V+++YLED+IPH+ M
Sbjct: 125 YGRYGFVAVTEVYLEDDIPHIGM 147
```

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 260

A DNA sequence (GBSx0275) was identified in *S.agalactiae* <SEQ ID 825> which encodes the amino acid sequence <SEQ ID 826>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2010(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 261

A DNA sequence (GBSx0276) was identified in *S.agalactiae* <SEQ ID 827> which encodes the amino acid sequence <SEQ ID 828>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2935(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:CAB12631 GB:Z99108 similar to RNA methyltransferase [Bacillus subtilis]
 Identities = 217/448 (48%), Positives = 298/448 (66%), Gaps = 4/448 (0%)

Query: 7   QRIPLKIKRMGINGEGIGFYKKTLIFVPGALKGEEVFCQISSVRRNFAEAKLLKINKKSK 66
           Q  PL IKR+GINGEG+G++KK ++FVPGAL GEEV  Q + V+  F+E ++ KI K S+
Sbjct: 16  QTFPLTIKRLGINGEGVGYFKKKVVFVPGALPGEEVVVQATKVQPKFSEGRIKKIRKASE 75

Query: 67  NRVEPPCSIYKECGGCQIMHLQYDKQLEFKTDVIRQALMKFKPEGYENYEIRKTIGMSEP 126
           +RV PPC +Y++CGGCQ+ HL Y +QL  K D++ Q+L +       EN EI++TIGM  P
```

```
Sbjct: 76  HRVAPPCPVYEQCGGCQLQHLAYSQQLREKRDIVIQSLERHTKFKVENMEIKETIGMDNP 135

Query: 127 EHYRAKLQFQV-RSFGGNVKAGLYAQGTHRLIDIKDCLVQDSLTQEMINRVAELLGKYKL 185
           +YR K QFQ+ RS  G++ AGLY   +H ++ IKDC+VQ   T +    V  +L  + +
Sbjct: 136 WNYRNKSQFQIGRSQSGSIIAGLYGLDSHDIVPIKDCIVQHPATNKTTGIVRRILEDFNV 195

Query: 186 PIYNERKIAG-VRTVMIRRAQASGEVQLIFITSKRL--DFDDVVIELVREFPELKTVAVN 242
            +YNERK  G VRT++ R    +GEVQ++ +T+K      +++V  + +  PE+K++  N
Sbjct: 196 SVYNERKRKGDVRTIVTRVGFETGEVQVVLVTAKETLPHKEEIVKAIQKRLPEVKSIIQN 255

Query: 243 INASKTSDIYGQITEVIWGQESINEEVLDYGFSLSPRAFYQLNPKQTQILYSEAVKALDV 302
           +N +KTS I+G+ T+ + G+  I E + D  F LS RAF+QLNP+QT  LY E  KA  +
Sbjct: 256 VNGAKTSVIFGEKTKQLAGKTVIQEVLGDVSFELSARAFFQLNPEQTVKLYDEVKKAAQL 315

Query: 303 KEDDDLIDAYCGVGTIGLAFAGKVKSVRGMDIIPEAIQDAKENALYMGFTNTHYEAGKAE 362
             + ++DAYCGVGTIG+  A   K VRGMD+I E+I DAK+NA   G  N  Y  G AE
Sbjct: 316 TGKEKVVDAYCGVGTIGMWVADGAKEVRGMDVIKESIDDAKKNAKKHGMANATYVTGTAE 375

Query: 363 DIIPRWYSEGFRANALIVDPPRTGLDDKLLNTILKMPPEKMVYVSCNTSTLARDLVTLTK 422
            +P+W  EGFR + +IVDPPRTG D   L+TI K+ P++ VYVSCN STLA+DL TL+K
Sbjct: 376 HWLPKWTKEGFRPDVVIVDPPRTGCDSTFLDTIKKVKPKRFVYVSCNPSTLAKDLQTLSK 435

Query: 423 VYHVHYIQSVDMFPHTARTEAVVKLQRK 450
           Y V YIQ VDMFP TA  EAV +L  K
Sbjct: 436 DYRVDYIQPVDMFPQTAHVEAVARLVLK 463
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 829> which encodes the amino acid sequence <SEQ ID 830>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2980(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 = 327/450 (72%), Positives = 397/450 (87%)

Query: 1   MNVVLKQRIPLKIKRMGINGEGIGFYKKTLIFVPGALKGEEVFCQISSVRRNFAEAKLLK 60
           M V +KQ+IPLKIKRMGINGEGIGFY+KTL+FVPGALKGE++FCQI++V+RNFAEAKLL
Sbjct: 1   MVVKVKQKIPLKIKRMGINGEGIGFYQKTLVFVPGALKGEDIFCQITAVKRNFAEAKLLT 60

Query: 61  INKKSKNRVEPPCSIYKECGGCQIMHLQYDKQLEFKTDVIRQALMKFKPEGYENYEIRKT 120
           +NK SKNRV+P CS+Y+ CGGCQIMHL Y KQL+FK DVIRQAL KFKP GYE +EIR T
Sbjct: 61  VNKASKNRVKPACSVYETCGGCQIMHLAYPKQLDFKDDVIRQALKKFKPTGYEQFEIRPT 120

Query: 121 IGMSEPEHYRAKLQFQVRSFGGNVKAGLYAQGTHRLIDIKDCLVQDSLTQEMINRVAELL 180
           +GM +P+HYRAKLQFQ+RSFGG VKAGL++QG+HRL+ I +CLVQD LTQ++IN++ +L+
Sbjct: 121 LGMKKPDHYRAKLQFQLRSFGGTVKAGLFSQGSHRLVPIDNCLVQDQLTQDIINKITQLV 180

Query: 181 GKYKLPIYNERKIAGVRTVMIRRAQASGEVQLIFITSKRLDFDDVVIELVREFPELKTVA 240
            KYKLPIYNERKIAG+RT+M+R+AQAS +VQ+I ++SK +   + + EL + FP++KTVA
Sbjct: 181 DKYKLPIYNERKIAGIRTIMVRKAQASDQVQIIVVSSKEVRLANFIGELTKAFPQVKTVA 240

Query: 241 VNINASKTSDIYGQITEVIWGQESINEEVLDYGFSLSPRAFYQLNPKQTQILYSEAVKAL 300
           +N N SK+S+IYG  TE++WGQE+I+EEVLDYGF+LSPRAFYQLNP+QT++LY E VKAL
Sbjct: 241 LNSNRSKSSEIYGDETEILWGQEAIHEEVLDYGFALSPRAFYQLNPQQTEVLYGEVVKAL 300

Query: 301 DVKEDDDLIDAYCGVGTIGLAFAGKVKSVRGMDIIPEAIQDAKENALYMGFTNTHYEAGK 360
           DV   D +IDAYCGVG+IG AFAGKVKSVRGMDIIPEAI+DA++NA  MGF N +YEAGK
Sbjct: 301 DVGSKDHIIDAYCGVGSIGFAFAGKVKSVRGMDIIPEAIEDAQKNAKAMGFDNAYYEAGK 360

Query: 361 AEDIIPRWYSEGFRANALIVDPPRTGLDDKLLNTILKMPPEKMVYVSCNTSTLARDLVTL 420
           AEDII +WY +G+RA+A+IVDPPRTGLDDKLL TIL   P++MVYVSCNTSTLARDLV L
```

```
Sbjct: 361 AEDIISKWYKQGYRADAVIVDPPRTGLDDKLLKTILHYQPKQMVYVSCNTSTLARDLVQL 420

Query: 421 TKVYHVHYIQSVDMFPHTARTEAVVKLQRK 450
           TKVY VHYIQSVDMFPHTARTEAVVKLQ++
Sbjct: 421 TKVYDVHYIQSVDMFPHTARTEAVVKLQKR 450
```

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

Example 262

A DNA sequence (GBSx0277) was identified in *S.agalactiae* <SEQ ID 831> which encodes the amino acid sequence <SEQ ID 832>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3505(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:BAB04643 GB:AP001510 unknown conserved protein in B. subtilis
        [Bacillus halodurans]
 Identities = 74/263 (28%), Positives = 141/263 (53%), Gaps = 9/263 (3%)

Query: 3   ITKIEKKKR---LYTLEL-DNTENLY---ITEDTIVHFMLSKGMIINAEKLENIKKFAQL 55
           IT+IE +KR    Y + +  N +++Y   + E  ++   L KG+ I+AE+++ I    ++
Sbjct: 4   ITRIEVQKRNNERYNIFIHQNGQDVYAFSVDEQVLIKQGLRKGLDIDAEQMKQILYEDEV 63

Query: 56  SYGKNLGLYYISFKQRTEKEVIKYLQQHDIDSKIIPQIIDNLKSENWINDKNYVQSFIQQ 115
              NL L+Y+S++ R+  EV  YL++ D +  II  ++  L  +  ++D  + ++FIQ
Sbjct: 64  QKTFNLALHYLSYRMRSVHEVRTYLKKKDREEPIIEHVLHRLTEQRLLDDHAFAEAFIQT 123

Query: 116 NLNTGDKGPYVIKQKLLQKGIKSKIIESELQAINFQDLASKISQKLYKKYQNKLPLKAL- 174
              T  KGP  +KQ+L +KG+  K IE  L   ++++   ++   L K+         +L
Sbjct: 124 KRATTSKGPLKLKQELAEKGVSEKTIEGALTTFSYEEQVEQVKAWLEKQKGRTFKGSSLA 183

Query: 175 -KDKLMQSLTTKGFDYQIVHTVIQNLEIEKDQELEEDLIYKELDKQYQKLSKKHDQYELK 233
             K KL + L  KG+   ++     ++ I++++E E + +    +K  +K + K   +EL+
Sbjct: 184 WKQKLSRQLLAKGYTSPVIEEAFADVPIKQEEEEEWEALKAFGEKAMRKYAGKKTGWELQ 243

Query: 234 QRIINALMRKGYQYEDIKSALRE 256
           Q++  AL RKG+  E I+  L +
Sbjct: 244 QKVKQALYRKGFSLEMIERYLND 266
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 833> which encodes the amino acid sequence <SEQ ID 834>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2388(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 = 146/258 (56%), Positives = 190/258 (73%)

Query: 1   MKITKIEKKKRLYTLELDNTENLYITEDTIVHFMLSKGMIINAEKLENIKKFAQLSYGKN 60
           MKITKIEKKKRLY +ELDN E+LY+TEDTIV FMLSK  +++ ++LE++K FAQLSYGKN
```

```
Sbjct: 1   MKITKIEKKKRLYLIELDNDESLYVTEDTIVRFMLSKDKVLDNDQLEDMKHFAQLSYGKN 60

Query: 61  LGLYYISFKQRTEKEVIKYLQQHDIDSKIIPQIIDNLKSENWINDKNYVQSFIQQNLNTG 120
           L LY++SF+QR+ K+V  YL++H+I+  II  II  L+ E WI+D     ++I+QN  G
Sbjct: 61  LALYFLSFQQRSNKQVADYLRKHEIEEHIIADIITQLQEEQWIDDTKLADTYIRQNQLNG 120

Query: 121 DKGPYVIKQKLLQKGIKSKIIESELQAINFQDLASKISQKLYKKYQNKLPLKALKDKLMQ 180
           DKGP V+KQKLLQKGI S  I+  L   +F  LA K+SQKL+ KYQ KLP KALKDK+ Q
Sbjct: 121 DKGPQVLKQKLLQKGIASHDIDPILSQTDFSQLAQKVSQKLFDKYQEKLPPKALKDKITQ 180

Query: 181 SLTTKGFDYQIVHTVIQNLEIEKDQELEEDLIYKELDKQYQKLSKKHDQYELKQRIINAL 240
           +L TKGF Y +    + +L  ++D +  EDL+ KELDKQY+KLS+K+D Y LKQ++  AL
Sbjct: 181 ALLTKGFSYDLAKHSLNHLNFDQDNQEIEDLLDKELDKQYRKLSRKYDGYTLKQKLYQAL 240

Query: 241 MRKGYQYEDIKSALREYL 258
            RKGY  +DI   LR YL
Sbjct: 241 YRKGYNSDDINCKLRNYL 258
```

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

Example 263

A DNA sequence (GBSx0278) was identified in *S.agalactiae* <SEQ ID 835> which encodes the amino acid sequence <SEQ ID 836>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3912(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:BAB04659 GB:AP001510 unknown conserved protein in B. subtilis
          [Bacillus halodurans]
 Identities = 96/175 (54%), Positives = 122/175 (68%)

Query: 1   MRLPKEGDFITIQSYKHDGSLHRTWRDTMVLKTTENALIGVNDHTLVTENDGRRWVTREP 60
           M  PK G  I IQSYKH+GS+HR W +T+VLK T   +IG ND  LV E+DGR W TREP
Sbjct: 1   MNFPKVGSKIQIQSYKHNGSIHRIWEETIVLKGTSKVVIGGNDRILVKESDGRHWRTREP 60

Query: 61  AIVYFHKKYWFNIIAMIRETGVSYYCNLASPYILDPEALKYIDYDLDVKVFADGEKRLLD 120
           AI YF  + WFN I MIR  G+ +YCNL +P+  D EALKYIDYDLD+KVF D   +LLD
Sbjct: 61  AICYFDSEQWFNTIGMIRADGIYFYCNLGTPFTWDEEALKYIDYDLDIKVFPDMTFKLLD 120

Query: 121 VDEYEQHKAQMNYPTDIDYILKENVKILVEWINENKGPFSSSYINIWYKRYLELK 175
            DEY  H+  M YP +ID IL+ +V  LV WI++ KGPF+  ++  WY+R+L+ +
Sbjct: 121 EDEYAMHRKMMKYPPEIDRILQRSVDELVSWIHQRKGPFAPQFVESWYERFLQYR 175
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 837> which encodes the amino acid sequence <SEQ ID 838>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3912(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 = 155/177 (87%), Positives = 165/177 (92%)

Query: 1   MRLPKEGDFITIQSYKHDGSLHRTWRDTMVLKTTENALIGVNDHTLVTENDGRRWVTREP 60
           M+LPKEGDFITIQSYKHDGSLHRTWRDTMVLKTTENALIGVNDHTLVTE+DGRRWVTREP
Sbjct: 1   MKLPKEGDFITIQSYKHDGSLHRTWRDTMVLKTTENALIGVNDHTLVTESDGRRWVTREP 60

Query: 61  AIVYFHKKYWFNIIAMIRETGVSYYCNLASPYILDPEALKYIDYDLDVKVFADGEKRLLD 120
           AIVYFHKKYWFNIIAMIR+ GVSYYCNLASPY++D EALKYIDYDLDVKVFADGEKRLLD
Sbjct: 61  AIVYFHKKYWFNIIAMIRDNGVSYYCNLASPYMMDTEALKYIDYDLDVKVFADGEKRLLD 120

Query: 121 VDEYEQHKAQMNYPTDIDYILKENVKILVEWINENKGPFSSSYINIWYKRYLELKKR 177
           VDEYE HK +M Y  D+D+ILKENVKILV+WIN  KGPFS +YI IWYKRYLELK R
Sbjct: 121 VDEYEIHKKEMQYSADMDFILKENVKILVDWINHEKGPFSKAYITIWYKRYLELKNR 177
```

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

Example 264

A DNA sequence (GBSx0288) was identified in *S.agalactiae* <SEQ ID 839> which encodes the amino acid sequence <SEQ ID 840>. This protein is predicted to be jag protein. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1666(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:BAB07782 GB:AP001520 spoIIIJ-associated protein [Bacillus halodurans]
 Identities = 54/198 (27%), Positives = 98/198 (49%), Gaps = 6/198 (3%)

Query: 100 DVVEEYIEEVDETLEKEDVSQPELPKIDDKNVVTTSEAIEKIDLLPNIEVAAAQVTKYVE 159
           + VE+ I E+  T E+   +  E PK     ++ +  A+ ++ + P+      +   ++E
Sbjct: 13  EAVEQAIIELGTTRERITYTVVEEPKSGLFGILGSKPAVIEVVVKPD---PVDRAKAFLE 69

Query: 160 NIIYEMDLDA--TIETTTSKRQINLQIETPEAGRIIGYHGKVLKSLQLLAQNYLHDRFSK 217
            ++ EMD++   TIE   +    N+  E  + G +IG  G+ L SLQ L     +    +
Sbjct: 70  ELLQEMDMEVEVTIEKDPATVLFNISGEQ-DLGTLIGKRGQTLDSLQYLVNLVANKEEGE 128

Query: 218 SFSVSINVHDYVEHRTETLIDFSKKIARRVLETNEPYHMDPMSNSERKTVHKTIATIEGV 277
             + ++  +Y   R E L+  ++++A + L T  P  ++PMS  ERK +H  +  +  V
Sbjct: 129 FIRIKLDAENYRARRKEALVQLAERLASKALRTKRPVSLEPMSAHERKIIHTALQELGDV 188

Query: 278 ESYSEGNDPNRFVVVTKK 295
           E+YSEG    R VV+  K
Sbjct: 189 ETYSEGQGIGRHVVIAPK 206
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 841> which encodes the amino acid sequence <SEQ ID 842>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3721(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 = 176/302 (58%), Positives = 223/302 (73%), Gaps = 32/302 (10%)

Query: 23  MVLFTGATVEEAIEKGLQELNISRLRAHIKVVSREKKGFLGFGKKPAKVEIEGITDEVTD 82
           MVLFTG TVEEAIE GLQEL +SRL+AHIKV+S+EKKGFLGFGKKPA+V+IEGI+D+
Sbjct: 1   MVLFTGKTVEEAIETGLQELGLSRLKAHIKVISKEKKGFLGFGKKPAQVDIEGISDKTVY 60

Query: 83  INESVALKNI------KNVPS--SVDVVEEYIEEVDETLEKEDVSQPELPKIDDK----- 129
             +  A + +      +N P+  S DV  E I+   + LE ED     L   D
Sbjct: 61  KADKKATRGVPEDINRQNTPAVNSADVEPEEIKAT-QRLEAEDTKVVPLMSEDSPAQTPS 119

Query: 130 ---NVVTTSEA------IEKIDL---------LPNIEVAAAQVTKYVENIIYEMDLDATI 171
               VT ++A      +E+ ++            +IE AA +V+ YV  IIYEMD++AT+
Sbjct: 120 NLAETVTETKAQQPSIPVEESEVPQDAGNDGFSKDIEKAAQEVSDYVTKIIYEMDIEATV 179

Query: 172 ETTTSKRQINLQIETPEAGRIIGYHGKVLKSLQLLAQNYLHDRFSKSFSVSINVHDYVEH 231
           ET+ ++RQINLQIETPEAGR+IGYHGKVLKSLQLLAQN+LHDR+SK+FSVS+NVHDYVEH
Sbjct: 180 ETSNNRRQINLQIETPEAGRVIGYHGKVLKSLQLLAQNFLHDRYSKNFSVSLNVHDYVEH 239

Query: 232 RTETLIDFSKKIARRVLETNEPYHMDPMSNSERKTVHKTIATIEGVESYSEGNDPNRFVV 291
           RTETLIDF++K+A+RVLE+ + Y MDPMSNSERK VHKT+++IEGV+SYSEGNDPNR+VV
Sbjct: 240 RTETLIDFTQKVAKRVLESGQDYTMDPMSNSERKIVHKTVSSIEGVDSYSEGNDPNRYVV 299

Query: 292 VT 293
           V+
Sbjct: 300 VS 301
```

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

Example 265

A DNA sequence (GBSx0290) was identified in *S.agalactiae* <SEQ ID 843> which encodes the amino acid sequence <SEQ ID 844>. This protein is predicted to be 60 kd inner-membrane protein (yidC). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> May be a lipoprotein
   INTEGRAL    Likelihood = -7.38   Transmembrane   54 -  70 (  52 -  75)
   INTEGRAL    Likelihood = -5.20   Transmembrane  193 - 209 ( 192 - 211)
   INTEGRAL    Likelihood = -3.61   Transmembrane  125 - 141 ( 124 - 144)
   INTEGRAL    Likelihood = -2.44   Transmembrane  168 - 184 ( 167 - 184)

----- Final Results -----
              bacterial membrane --- Certainty=0.3951(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:CAA78595 GB:Z14225 SpoIIIJ [Bacillus subtilis]
 Identities = 79/243 (32%), Positives = 142/243 (57%), Gaps = 5/243 (2%)

Query: 1   MKKKLKTFSLILLTGSLLVACG--RGEVSSHSATLWEQ-IVYAFAKSIQWLS--FNHSIG 55
           MK+++     ++    LL  C   +  +++ S   W++ +VY  ++ I +++     + G
Sbjct: 1   MKRRIGLLLSMVGVFMLLAGCSSVKEPITADSPHFWDKYVVYPLSELITYVAKLTGDNYG 60

Query: 56  LGIILFTLIIRAIMMPLYNMQMKSSQKMQEIQPRLKELQKKYPGKDPDNRLKLNDEMQSM 115
           L IIL T++IR +++PL   Q++SS+ MQ +QP +++L++KY  KD   + KL  E  ++
Sbjct: 61  LSIILVTILIRLLILPLMIKQLRSSKAMQALQPEMQKLKEKYSSKDQKTQQKLQQETMAL 120

Query: 116 YKAEGVNPYASVLPLLIQLPVLWALFQALTRVSFLKVGTFLSLELSQPDPYYILPVLAAL 175
           ++  GVNP A   P+LIQ+P+L   + A+ R   +   +FL  +L + DPYYILP++A +
Sbjct: 121 FQKHGVNPLAGCFPILIQMPILIGFYHAIMRTQAISEHSFLWFDLGEKDPYYILPIVAGV 180

Query: 176 FTFLSTWLTNKAAVEKNIALTLMTYVMPFIILVTSFNFASGVVLYWTVSNAFQVFQILLL 235
```

```
            TF+   L      ++N  + +M ++MP +I+V + NF + + LYW V N F + Q  L+
Sbjct: 181 ATFVQQKLMMAGNAQQNPQMAMMLWIMPIMIIVFAINFPAALSLYWVVGNLFMIAQTFLI 240

Query: 236 NNP 238
             P
Sbjct: 241 KGP 243
```

A related GBS sequence was identified <SEQ ID 10783> which encodes amino acid sequence <SEQ ID 10784>.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 845> which encodes the amino acid sequence <SEQ ID 846>. Analysis of this protein sequence reveals the following:

```
      Possible site: 49

   >>> May be a lipoprotein
      INTEGRAL    Likelihood = -6.32   Transmembrane  198 - 214 ( 197 - 220)
      INTEGRAL    Likelihood = -5.52   Transmembrane   59 -  75 (  57 -  80)
      INTEGRAL    Likelihood = -4.25   Transmembrane  130 - 146 ( 129 - 150)
      INTEGRAL    Likelihood = -2.28   Transmembrane  173 - 189 ( 170 - 189)

   ----- Final Results -----
                bacterial membrane --- Certainty=0.3527(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:BAA05234 GB:D26185 stage III sporulation [Bacillus subtilis]
 Identities = 90/249 (36%), Positives = 150/249 (60%), Gaps = 6/249 (2%)

Query: 16  IVPLVLLLVACG--RGEVTAQSSSGWDQ-LVYLFARAIQWLS--FDGSIGVGIILFTLTI 70
           +V + +LL  C   +  +TA S   WD+ +VY  +  I +++      + G+ IIL T+ I
Sbjct: 13  MVGVFMLLAGCSSVKEPITADSPHFWDKYVVYPLSELITYVAKLTGDNYGLSIILVTILI 72

Query: 71  RLMLMPLFNMQIKSSQKMQDIQPELRELQRKYAGKDTQTRMKLAEESQALYKKYGVNPYA 130
           RL+++PL   Q++SS+ MQ +QPE+++L+ KY+ KD +T+ KL +E+ AL++K+GVNP A
Sbjct: 73  RLLILPLMIKQLRSSKAMQALQPEMQKLKEKYSSKDQKTQQKLQQETMALFQKHGVNPLA 132

Query: 131 SLLPLLIQMPVMIALFQALTRVSFLKTGTFLWVELAQHDHLYLLPVLAAVFTFLSTWLTN 190
              P+LIQMP++I  + A+ R   +   +FLW +L + D  Y+LP++A V TF+   L
Sbjct: 133 GCFPILIQMPILIGFYHAIMRTQAISEHSFLWFDLGEKDPYYILPIVAGVATFVQQKLMM 192

Query: 191 LAAKEKNVMMTVMIYVMPLMIFFMGFNLASGVVLYWTVSNAFQVVQLLLLNNP-FKIIAE 249
              ++N  M +M+++MP+MI     N  + + LYW V N F + Q  L+  P  K   E
Sbjct: 193 AGNAQQNPQMAMMLWIMPIMIIVFAINFPAALSLYWVVGNLFMIAQTFLIKGPDIKKNPE 252

Query: 250 RQRLANEEK 258
           Q+   ++K
Sbjct: 253 PQKAGGKKK 261
```

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

```
 Identities = 172/270 (63%), Positives = 217/270 (79%), Gaps = 1/270 (0%)

Query: 1   MKKKLKTFSLILLTGSLLVACGRGEVSSHSATLWEQIVYAFAKSIQWLSFNHSIGLGIIL 60
           +KK +K   ++ L   LLVACGRGEV++ S++ W+Q+VY FA++IQWLSF+ SIG+GIIL
Sbjct: 7   VKKNIKIARIVPLV-LLLVACGRGEVTAQSSSGWDQLVYLFARAIQWLSFDGSIGVGIIL 65

Query: 61  FTLIIRAIMMPLYNMQMKSSQKMQEIQPRLKELQKKYPGKDPDNRLKLNDEMQSMYKAEG 120
           FTL IR ++MPL+NMQ+KSSQKMQ+IQP L+ELQ+KY GKD   R+KL +E Q++YK  G
Sbjct: 66  FTLTIRLMLMPLFNMQIKSSQKMQDIQPELRELQRKYAGKDTQTRMKLAEESQALYKKYG 125

Query: 121 VNPYASVLPLLIQLPVLWALFQALTRVSFLKVGTFLSLELSQPDPYYILPVLAALFTFLS 180
           VNPYAS+LPLLIQ+PV+ ALFQALTRVSFLK GTFL +EL+Q D  Y+LPVLAA+FTFLS
Sbjct: 126 VNPYASLLPLLIQMPVMIALFQALTRVSFLKTGTFLWVELAQHDHLYLLPVLAAVFTFLS 185
```

```
Query: 181 TWLTNKAAVEKNIALTLMTYVMPFIILVTSFNFASGVVLYWTVSNAFQVFQILLLNNPYK 240
           TWLTN AA EKN+ +T+M YVMP +I     FN ASGVVLYWTVSNAFQV Q+LLLNNP+K
Sbjct: 186 TWLTNLAAKEKNVMMTVMIYVMPLMIFFMGFNLASGVVLYWTVSNAFQVVQLLLLNNPFK 245

Query: 241 IIKVREEAVRVAHEKEQRVKRAKRKASKKR 270
           II  R+        E+  R +RA++KA K+
Sbjct: 246 IIAERQRLANEEKERRLRERRARKKAMKRK 275
```

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

A related GBS gene <SEQ ID 8527> and protein <SEQ ID 8528> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: 20   Crend: 5
McG: Discrim Score:      4.90
GvH: Signal Score (-7.5): -0.39
     Possible site: 42
>>> May be a lipoprotein
ALOM program   count: 4 value:  -7.38 threshold:  0.0
   INTEGRAL    Likelihood = -7.38   Transmembrane   54 -  70 (  52 -  75)
   INTEGRAL    Likelihood = -5.20   Transmembrane  193 - 209 ( 192 - 211)
   INTEGRAL    Likelihood = -3.61   Transmembrane  125 - 141 ( 124 - 144)
   INTEGRAL    Likelihood = -2.44   Transmembrane  168 - 184 ( 167 - 184)
   PERIPHERAL  Likelihood =  2.54     217
 modified ALOM score:   1.98

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3951(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:

```
32.8/62.3% over 242aa
                                                                  Bacillus subtilis
  EGAD|17722|  stage III sporulation protein j precursor Insert characterized
OMNI|NT01BS4782 -identity Insert characterized
  SP|Q01625|SP3J_BACSU STAGE III SPORULATION PROTEIN J PRECURSOR. Edit characterized
  GP|40023|emb|CAA44401.1||X62539 unnamed protein product Insert characterized
  GP|467388|dbj|BAA05234.1||D26185 stage III sporulation Insert characterized
  GP|2636651|emb|CAB16141.1||Z99124 alternate gene name: spo0J87 Insert characterized
  PIR|I40437|I40437 stage III sporulation protein spoIIIJ - Insert characterized

ORF02221(301 - 1014 of 1413)
EGAD|17722| S4098(3 - 245 of 261) stage III sporulation protein j precursor { acillus
subtilis}OMNI|NT01 S4782 -identitySP|Q01625|SP3J_ ACSU STAGE III SPORULATION PROTEIN J
PRECURSOR.GP|40023|emb|CAA44401.1||X62539 unnamed protein product { acillus
subtilis}GP|467388|dbj| AA05234.1||D26185 stage III sporulation { acillus
subtilis}GP|2636651|emb|CA 16141.1||Z99124 alternate gene name: spo0J87 { acillus
subtilis}PIR|I40437|I40437 stage III sporulation protein spoIIIJ - acillus subtilis
%Match = 17.0
%Identity = 32.8  %Similarity = 62.2
Matches = 79  Mismatches = 88  Conservative Sub.s = 71

219       249       279       309       339                393       420
DFVVIARKGVEELDYQALEKNLIHVLKIAGLI*KGIKLKKKLKTFSLILLTGSLLVACG--RGEVSSHSATLWEQ-IVYA
                                  : ||::: : ::    || |  : ::: |  :|:: :||
                                  MLLKRRIGLLLSMVGVFMLLAGCSSVKEPITADSPHFWDKYVVYP
                                          10        20        30        40

474       504       534       564       594       624       654
FAKSIQWLS--FNHSIGLGIILFTLIIRAIMMPLYNMQMKSSQKMQEIQPRLKELQKKYPGKDPDNRLKLNDEMQSMYKA
```

```
::: | :::     : || ||| |::|| :::||   |::||: || :|| :::|::||  ||   : ||  |  ::::
LSELITYVAKLTGDNYGLSIILVTILIRLLILPLMIKQLRSSKAMQALQPEMQKLKEKYSSKDQKTQQKLQQETMALFQK
             60        70        80        90       100       110       120

684       714       744       774       804       834       864       894
EGVNPYASVLPLLIQLPVLWALFQALTRVSFLKVGTFLSLELSQPDPYYILPVLAALFTFLSTWLTNKAAVEKNIALTLM
 |||| |  :|:|||:|:|  :::|: |   :   :|| ::| : |||||||::| : ||:   |      ::|  : :|
HGVNPLAGCFPILIQMPILIGFYHAIMRTQAISEHSFLWFDLGEKDPYYILPIVAGVATFVQQKLMMAGNAQQNPQMAMM
            140       150       160       170       180       190       200

924       954       984      1014      1044      1074      1104      1134
TYVMPFIILVTSFNFASGVVLYWTVSNAFQVFQILLLNNPYKIIKVREEAVRVAHEKEQRVKRAKRKASKKRK*ENHGII
 ::|| :|:| : || : : ||| | | | : | :|:  |       :     :
LWIMPIMIIVFAINFPAALSLYWVVGNLFMIAQTFLIKGPDIKKNPEPQKAGGKKK
            220       230       240       250       260
```

Example 266

A DNA sequence (GBSx0291) was identified in *S.agalactiae* <SEQ ID 847> which encodes the amino acid sequence <SEQ ID 848>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3778(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 9579> which encodes amino acid sequence <SEQ ID 9580> was also identified.

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

```
>GP:CAA44400 GB:X62539 homologous to E.coli rnpA [Bacillus subtilis]
 Identities = 52/109 (47%), Positives = 77/109 (69%), Gaps = 1/109 (0%)

Query: 21  LKKTYRVKSDKDFQMIFSRGKNVANRKFVIYYLEK-EQKHFRVGISVSKKLGNAVVRNAI 79
           LKK  R+K ++DFQ +F  G +VANR+FV+Y L++ E     RVG+SVSKK+GNAV+RN I
Sbjct: 4   LKKRNRLKKNEDFQKVFKHGTSVANRQFVLYTLDQPENDELRVGLSVSKKIGNAVMRNRI 63

Query: 80  KRKIRHVLLSQKTALQDYDFVVIARKGVEELDYQALEKNLIHVLKIAGL 128
           KR IR   L +K  L++ D+++IARK   +L Y+  +K+L H+ + + L
Sbjct: 64  KRLIRQAFLEEKERLKEKDYIIIARKPASQLTYEETKKSLQHLFRKSSL 112
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 849> which encodes the amino acid sequence <SEQ ID 850>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3820(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 = 73/109 (66%), Positives = 88/109 (79%)

Query: 21  LKKTYRVKSDKDFQMIFSRGKNVANRKFVIYYLEKEQKHFRVGISVSKKLGNAVVRNAIK 80
           LKKTYRVK +KDFQ IF  GK+ ANRKFVIY+L + Q HFRVGISV KK+GNAV RNA+K
Sbjct: 1   LKKTYRVKREKDFQAIFKDGKSTANRKFVIYHLNRGQDHFRVGISVGKKIGNAVTRNAVK 60

Query: 81  RKIRHVLLSQKTALQDYDFVVIARKGVEELDYQALEKNLIHVLKIAGLI 129
           RKIRHV+++    L+  DFVVIARKGV  L+YQ L++NL HVLK+A L+
```

```
Sbjct: 61  RKIRHVIMALGHQLKSEDFVVIARKGVHSLEYQELQQNLHHVLKLAQLL 109
```

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

Example 267

A DNA sequence (GBSx0292) was identified in *S.agalactiae* <SEQ ID 851> which encodes the amino acid sequence <SEQ ID 852>. This protein is predicted to be glycerol-3-phosphate dehydrogenase, NAD-dependent (gpsA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1429(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 8529> which encodes amino acid sequence <SEQ ID 8530> was also identified. There is a signal peptide at residues 1-19. The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA86746 GB:U32164 NAD(P)H-dependent dihydroxyacetone-phosphate
          reductase [Bacillus subtilis]
 Identities = 177/333 (53%), Positives = 241/333 (72%)

Query: 18  QKIAVLGPGSWGTALAQVLNDNGHEVRLWGNVVEQIEEINTNHTNQRYFKDITLDSKIKA 77
           +K+ +LG GSWGTALA VL DNG+EV +W +  + I +IN  H N+ Y  ++ L + IK
Sbjct: 2   KKVTMLGAGSWGTALALVLTDNGNEVCVWAHRADLIHQINELHENKDYLPNVKLSTSIKG 61

Query: 78  YTNLEEAINNVDSILFVVPTKVTRLVAKQVANLLKHKVVLMHASKGLEPGTHERLSTILE 137
            T+++EA+++ D I+  VPTK  R V +Q    +  K V +H SKG+EP +  R+S I+E
Sbjct: 62  TTDMKEAVSDADVIIVAVPTKAIREVLRQAVPFITKKAVFVHVSKGIEPDSLLRISEIME 121

Query: 138 EEISEQYRSDIVVVSGPSHAEEAIVRDITLITAASKDIEAAKYVQKLFSNHYFRLYTNTD 197
            E+    R DIVV+SGPSHAEE  +R  T +TA+SK + AA+ VQ LF NH FR+YTN D
Sbjct: 122 IELPSDVRRDIVVLSGPSHAEEVGLRHATTVTASSKSMRAAEEVQDLFINHNFRVYTNPD 181

Query: 198 VVGVETAGALKNIIAVGAGALHGLGYGDNAKAAIITRGLAEITRLGVQLGADPLTFSGLS 257
           ++GVE  GALKNIIA+ AG   GLGYGDNAKAA+ITRGLAEI RLG ++G +PLTFSGL+
Sbjct: 182 IIGVEIGGALKNIIALAAGITDGLGYGDNAKAALITRGLAEIARLGTKMGGNPLTFSGLT 241

Query: 258 GVGDLIVTGTSVHSRNWRAGDALGRGEKLEDIEKNMGMVIEGISTTKVAYEIAQNLNVYM 317
           GVGDLIVT TSVHSRNWRAG+ LG+G KLED+ + MGMV+EG+ TTK AY++++  +V M
Sbjct: 242 GVGDLIVTCTSVHSRNWRAGNLLGKGYKLEDVLEEMGMVVEGVRTTKAAYQLSKKYDVKM 301

Query: 318 PITEAIYKSIYEGANIKDSILDMMSNEFRSENE 350
           PITEA+++ ++ G  ++ ++  +M+     E E
Sbjct: 302 PITEALHQVLFNGQKVETAVESLMARGKTHEME 334
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 853> which encodes the amino acid sequence <SEQ ID 854>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0882(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 = 287/338 (84%), Positives = 316/338 (92%)

Query: 15  MTKQKIAVLGPGSWGTALAQVLNDNGHEVRLWGNVVEQIEEINTNHTNQRYFKDITLDSK 74
           MTKQK+A+LGPGSWGTAL+QVLNDNGH+VRLWGN+ +QIEEINT HTN+ YFKDI LD
Sbjct: 1   MTKQKVAILGPGSWGTALSQVLNDNGHDVRLWGNIPDQIEEINTKHTNRHYFKDIVLDKN 60

Query: 75  IKAYTNLEEAINNVDSILFVVPTKVTRLVAKQVANLLKHKVVLMHASKGLEPGTHERLST 134
           I A  +L +A+++VD++LFVVPTKVTRLVA+QVA +L HKVV+MHASKGLEP THERLST
Sbjct: 61  ITATLDLGQALSDVDAVLFVVPTKVTRLVARQVAAILDHKVVVMHASKGLEPETHERLST 120

Query: 135 ILEEEISEQYRSDIVVVSGPSHAEEAIVRDITLITAASKDIEAAKYVQKLFSNHYFRLYT 194
           ILEEEI   +RS++VVVSGPSHAEE IVRDITLITAASKDIEAAKYVQ LFSNHYFRLYT
Sbjct: 121 ILEEEIPAHFRSEVVVVSGPSHAEETIVRDITLITAASKDIEAAKYVQSLFSNHYFRLYT 180

Query: 195 NTDVVGVETAGALKNIIAVGAGALHGLGYGDNAKAAIITRGLAEITRLGVQLGADPLTFS 254
           NTDV+GVETAGALKNIIAVGAGALHGLGYGDNAKAA+ITRGLAEITRLGV+LGADPLT+S
Sbjct: 181 NTDVIGVETAGALKNIIAVGAGALHGLGYGDNAKAAVITRGLAEITRLGVKLGADPLTYS 240

Query: 255 GLSGVGDLIVTGTSVHSRNWRAGDALGRGEKLEDIEKNMGMVIEGISTTKVAYEIAQNLN 314
           GLSGVGDLIVTGTSVHSRNWRAG ALGRGEKLEDIE+NMGMVIEGI+TTKVAYEIAQ+L
Sbjct: 241 GLSGVGDLIVTGTSVHSRNWRAGAALGRGEKLEDIERNMGMVIEGIATTKVAYEIAQDLG 300

Query: 315 VYMPITEAIYKSIYEGANIKDSILDMMSNEFRSENEWH 352
           VYMPIT AIYKSIYEGA+IK+SIL MMSNEFRSENEWH
Sbjct: 301 VYMPITTAIYKSIYEGADIKESILGMMSNEFRSENEWH 338
```

SEQ ID 8530 (GBS291) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 5; MW 38.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 2; MW 64kDa).

GBS291-GST was purified as shown in Figure 226, lane 10-11.

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

Example 268

A DNA sequence (GBSx0293) was identified in *S.agalactiae* <SEQ ID 855> which encodes the amino acid sequence <SEQ ID 856>. This protein is predicted to be glucose-1-phosphate uridylyltransferase (gtaB). 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>
```

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

```
>GP:BAA28714 GB:AB001562 glucose-1-phosphate uridylyltransferase
           [Streptococcus mutans]
 Identities = 263/296 (88%), Positives = 285/296 (95%)

Query: 2   KVRKAVIPAAGLGTRFLPATKALAKEMLPIVDKPTIQFIVEEALKSGIEDILVVTGKSKR 61
           KVRKAVIPAAGLGTRFLPATKALAKEMLPIVDKPTIQFIVEEALKSGIEDILVVTGKSKR
Sbjct: 5   KVRKAVIPAAGLGTRFLPATKALAKEMLPIVDKPTIQFIVEEALKSGIEDILVVTGKSKR 64

Query: 62  SIEDHFDSNFELEYNLKEKGKNELLKLVDETTGIRLHFIRQSHPRGLGDAVLQAKAFVGN 121
           SIEDHFDSNFELEYNL++KGK +LLKLV++TT I LHFIRQSHPRGLGDAVLQAKAFVGN
Sbjct: 65  SIEDHFDSNFELEYNLEQKGKTDLLKLVNDTTAINLHFIRQSHPRGLGDAVLQAKAFVGN 124

Query: 122 EPFVVMLGDDLMDITNNKVIPLTKQLINDFEATHASTIAVMEVPHEDVSAYGVIAPQGEG 181
```

```
              EPFVVMLGDDLMDIT++K IPLT+QL+ND+E THASTIAVMEVPHEDVSAYGVIAPQGEG
Sbjct: 125 EPFVVMLGDDLMDITDDKAIPLTRQLMNDYEETHASTIAVMEVPHEDVSAYGVIAPQGEG 184

Query: 182 VNGLYSVNTFVEKPSPEEAPSNLAIIGRYLLTPEIFNILETQKPGAGNEIQLTDAIDTLN 241
              V+GLYSV+TFVEKP+P+EAPSNLAIIGRYLLTPEIF ILETQ+PGAGNE+QLTDAIDTLN
Sbjct: 185 VSGLYSVDTFVEKPAPKEAPSNLAIIGRYLLTPEIFTILETQEPGAGNEVQLTDAIDTLN 244

Query: 242 KTQRVFARKFTGDRYDVGDKFGFMKTSIDYALQHPQVKDDLKKYIIDLGKSLEKTS 297
              KTQRVFAR+F G RYDVGDKFGFMKTSIDYAL+HPQVK+DLK YII+LGK L++ S
Sbjct: 245 KTQRVFAREFKGKRYDVGDKFGFMKTSIDYALKHPQVKEDLKAYIIELGKKLDQKS 300
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 857> which encodes the amino acid sequence <SEQ ID 858>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have a cleavable N-term signal seq.

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

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

```
 Identities = 257/295 (87%), Positives = 277/295 (93%)

Query: 2   KVRKAVIPAAGLGTRFLPATKALAKEMLPIVDKPTIQFIVEEALKSGIEDILVVTGKSKR 61
            KVRKA+IPAAGLGTRFLPATKALAKEMLPIVDKPTIQFIVEEALKSGIE+ILVVTGK+KR
Sbjct: 3   KVRKAIIPAAGLGTRFLPATKALAKEMLPIVDKPTIQFIVEEALKSGIEEILVVTGKAKR 62

Query: 62  SIEDHFDSNFELEYNLKEKGKNELLKLVDETTGIRLHFIRQSHPRGLGDAVLQAKAFVGN 121
            SIEDHFDSNFELEYNL+ KGKNELLKLVDETT I LHFIRQSHPRGLGDAVLQAKAFVGN
Sbjct: 63  SIEDHFDSNFELEYNLQAKGKNELLKLVDETTAINLHFIRQSHPRGLGDAVLQAKAFVGN 122

Query: 122 EPFVVMLGDDLMDITNNKVIPLTKQLINDFEATHASTIAVMEVPHEDVSAYGVIAPQGEG 181
            EPFVVMLGDDLMDITN     PLTKQL+ D++ THASTIAVM+VPHEDVS+YGVIAPQG+
Sbjct: 123 EPFVVMLGDDLMDITNASAKPLTKQLMEDYDKTHASTIAVMKVPHEDVSSYGVIAPQGKA 182

Query: 182 VNGLYSVNTFVEKPSPEEAPSNLAIIGRYLLTPEIFNILETQKPGAGNEIQLTDAIDTLN 241
            V GLYSV+TFVEKP PE+APS+LAIIGRYLLTPEIF ILE Q PGAGNE+QLTDAIDTLN
Sbjct: 183 VKGLYSVDTFVEKPQPEDAPSDLAIIGRYLLTPEIFGILERQTPGAGNEVQLTDAIDTLN 242

Query: 242 KTQRVFARKFTGDRYDVGDKFGFMKTSIDYALQHPQVKDDLKKYIIDLGKSLEKT 296
            KTQRVFAR+F G+RYDVGDKFGFMKTSIDYAL+HPQVK+DLK YII LGK+LEK+
Sbjct: 243 KTQRVFAREFKGNRYDVGDKFGFMKTSIDYALEHPQVKEDLKNYIIKLGKALEKS 297
```

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

Example 269

A DNA sequence (GBSx0294) was identified in *S.agalactiae* <SEQ ID 859> which encodes the amino acid sequence <SEQ ID 860>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.94   Transmembrane   28 -  44 (  27 -  45)

----- Final Results -----
               bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB15143 GB:Z99120 similar to ABC transporter (lipoprotein)
          [Bacillus subtilis]
 Identities = 148/346 (42%), Positives = 222/346 (63%), Gaps = 16/346 (4%)

Query: 31  LTLLSLSVLTLTACGNRSDKSAN---KSDIKVAMVTNQGGVDDKSFNQSAWEGLQKWGKK 87
           ++L+  +   L ACGN    S +   K+   VAMVT+ GGVDDKSFNQSAWEG+Q +GK+
Sbjct: 1   MSLVIAAGTILGACGNSEKSSGSGEGKNKFSVAMVTDVGGVDDKSFNQSAWEGIQAFGKE 60

Query: 88  KGLTKG-NGFDYFQSSNESDHANNLDTAASSGYNLIFGIGFGLHDTIEKVSENNKDVKYV 146
            GL KG NG+DY QS +++D+  NL+  A   ++LI+G+G+ + D+I ++++  K+  +
Sbjct: 61  NGLKKGKNGYDYLQSKSDADYTTNLNKLARENFDLIYGVGYLMEDSISEIADQRKNTNFA 120

Query: 147 IVDDIIKGKENVASVTFADNEAAYLAGVAAAKTTKTKTVGFIGGMEGVVVKRFEAGFKAG 206
           I+D ++  K+NVAS+TF + E ++L GVAAA ++K+  +GF+GGME  ++K+FE GF+AG
Sbjct: 121 IIDAVVD-KDNVASITFKEQEGSFLVGVAAALSSKSGKIGFVGGMESELIKKFEVGFRAG 179

Query: 207 VKSIDPAIKVAVSYAGSFTDAAKGKTIAATQYATGVDVIYQAAGGTGAGIFSEAKTENET 266
           V++++P   V V YAG F  A  GK  A + Y +GVDVIY +AG TG G+F+EAK
Sbjct: 180 VQAVNPKAVVEVKYAGGFDKADVGKATAESMYKSGVDVIYHSAGATGTGVFTEAK---NL 236

Query: 267 RKESNK--VWVIGVDRDQSQEGNYVSKDGKKANFVLASTIKEVGKSLQSVAELTEKKQYP 324
           +KE  K  VWVIGVD+DQ  EG     +G   N  L S +K+V   ++ V +     ++P
Sbjct: 237 KKEDPKRDVWVIGVDKDQYAEGQV---EGTDDNVTLTSMVKKVDTVVEDVTKKASDGKFP 293

Query: 325 GGKVTVFGLKDSGVDI--KEHQLSSEGSVAVKKAKEDIVSGKIQVP 368
           GG+   +GL   GV I   +  LS +   AV K K+ I+ G +++P
Sbjct: 294 GGETLTYGLDQDGVGISPSKQNLSDDVIKAVDKWKKKIIDG-LEIP 338
```

There is also homology to SEQ ID 862.

A related GBS gene <SEQ ID 8531> and protein <SEQ ID 8532> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: 20   Crend: 3
       Sequence Pattern: CGNR
SRCFLG: 0
McG: Length of UR:   19
     Peak Value of UR:   2.31
     Net Charge of CR: 2
McG: Discrim Score:      5.09
GvH: Signal Score (-7.5): -3.29
     Possible site: 19
>>> May be a lipoprotein
Amino Acid Composition: calculated from 21
ALOM program   count: 0 value:   5.20 threshold:  0.0
   PERIPHERAL  Likelihood =  5.20      90
 modified ALOM score:  -1.54

*** Reasoning Step: 3

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

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

```
52.8/73.9% over 239aa
                                                                  Listeria
monocytogenes
  SP|Q48754| CD4+ T CELL-STIMULATING ANTIGEN PRECURSOR. Insert characterized
  GP|7240601|gb|AAB35725.2||S80336 CD4+ T cell-stimulating antigen Insert characterized

ORF02225(385 - 1086 of 1710)
SP|Q48754|TCSA_LISMO(8    -    247    of    268)    CD4+    T    CELL-STIMULATING    ANTIGEN
PRECURSOR.GP|7240601|gb|AAB35725.2||S80336  CD4+  T  cell-stimulating  antigen  (Listeria
monocytogenes)
```

```
%Match = 21.7
%Identity = 52.7  %Similarity = 73.8
Matches = 125  Mismatches = 59  Conservative Sub.s = 50

294       324       354       384       414       444       465       489
NFLWEK*NKVC*MIFLCYDRNLFLCDYNLLGGSFSVNRKIIGLTLLSLSVLTLTACGNRSD---KSANKS--DIKVAMVT
                                          : |:::  | : | |||: ||    | :||  |  |||||
                                       MKKRTFALALSMIIASGVILGACGSSSDDKKSSDDKSSKDFTVAMVT
                                          10        20        30        40

519       549       579       606       636       666       696       726
NQGGVDDKSFNQSAWEGLQKWGKKKGLTKG-NGFDYFQSSNESDHANNLDTAASSGYNLIFGIGFGLHDTIEKVSENNKD
: |||||:|||||||||||||:||   : || :|::|:||::|:|:  ||:||  | |:||:|||: | | ||:||:
DTGGVDDRSFNQSAWEGLQKFGKANDMEKGTDGYNYLQSASEADYKTNLNTAVRSDYDLIYGIGYKLKDAIEEVSKQKPK
          60        70        80        90       100       110       120

756       786       816       846       876       906       936       966
VKYVIVDDIIKGKENVASVTFADNEAAYLAGVAAAKTTKTKTVGFIGGMEGVVVKRFEAGFKAGVKSIDPAIKVAVSYAG
 :: |||| |  ::|| |: | ||: :|| || |  ||||  |||:||::| |: ||||||  ||||:::|  :: | ||
NQFAIVDDTIDDRDNVVSIGFKDNDGSYLVGVVAGLTTKTNKVGFVGGVKGTVIDRFEAGFTAGVKAVNPNAQIDVQYAN
          140       150       160       170       180       190       200

996      1026      1056      1086      1116      1146      1176      1206
SFTDAAKGKTIAATQYATGVDVIYQAAGGTGAGIFSEAKTENETRKESNKVWVIGVDRDQSQEGNYVSKDGKKANFVLAS
 |  | ||: ||:: |::|||||::|||||| |:|:|||   :   :       :
DFAKADKGQQIASSMYSSGVDVIFHAAGGTGNGVFAEAKNLKKKDLQMVPYGNSKLGCFGG
          220       230       240       250       260
```

A related GBS nucleic acid sequence <SEQ ID 10947> which encodes amino acid sequence <SEQ ID 10948> was also identified.

SEQ ID 8532 (GBS108) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 7; MW 39.6kDa). 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 9; MW 64.6kDa).

The GBS108-GST fusion product was purified (Figure 202, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 273), 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 270

A DNA sequence (GBSx0295) was identified in *S.agalactiae* <SEQ ID 863> which encodes the amino acid sequence <SEQ ID 864>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.74   Transmembrane  206 - 222 ( 197 - 224)
   INTEGRAL    Likelihood = -3.72   Transmembrane  174 - 190 ( 171 - 194)
   INTEGRAL    Likelihood = -3.19   Transmembrane   98 - 114 (  98 - 116)
   INTEGRAL    Likelihood = -1.54   Transmembrane  120 - 136 ( 120 - 139)
   INTEGRAL    Likelihood = -0.90   Transmembrane  157 - 173 ( 157 - 173)

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

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

```
>GP:CAB90755 GB:AJ400707 hypothetical protein [Streptococcus uberis]
 Identities = 126/218 (57%), Positives = 166/218 (75%)

Query: 8   KEYPTTVLLVSLTTLVFLLMQLTYGSQAESSQVIFQFGGIQGDYLKAYPTNLWRLISPIF 67
           KE P T   +8+T L+F++MQ+ YGS A+S QV+FQFGG+ G  +K+ P+ LWRL++PIF
Sbjct: 5   KEKPVTFFFLSVTILLFIVMQVFYGSWAKSPQVVFQFGGMFGLVVKSMPSQLWRLVTPIF 64

Query: 68  VHIGWEHFLLNGLALYFVGQMGESIWGSLRFLILYILSGLMGNIFTLFFTPHVVAAGAST 127
           +HIGWEHFL+N L LYFVGQ+ ESIWGS  FL+LY+LSG+MGN+ TLFFTPHVVAAGAST
Sbjct: 65  IHIGWEHFLINSLTLYFVGQLAESIWGSRFFLLLYVLSGIMGNVLTLFFTPHVVAAGAST 124

Query: 128 SLFGVFSAIAIAGYFGKNPYLKQVGKSYQVMILLNLFFNIFTPGVSLAGHVGGLVGGVLV 187
           SLFG+F+AI + GYFG N  LK +GKSYQ +I+LNL  N+F P V + GH+GG +GG L
Sbjct: 125 SLFGLFAAIVVVGYFGHNQLLKSIGKSYQTLIILNLVMNLFMPNVGIVGHLGGALGGALA 184

Query: 188 AIFLTKQNGSLLFKTWQSILALMIFIIVSISLIGLSLV 225
           A+FL      + LF    Q   AL+ ++ +++ LI LSL+
Sbjct: 185 AVFLPTLLDAELFTKKQKTSALLSYLTLALVLITLSLM 222
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 865> which encodes the amino acid sequence <SEQ ID 866>. Analysis of this protein sequence reveals the following:

```
     Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.92   Transmembrane  214 - 230 ( 212 - 232)
   INTEGRAL    Likelihood = -5.36   Transmembrane  135 - 151 ( 128 - 153)
   INTEGRAL    Likelihood = -1.81   Transmembrane  101 - 117 ( 100 - 117)
   INTEGRAL    Likelihood = -1.44   Transmembrane  183 - 199 ( 182 - 199)
   INTEGRAL    Likelihood = -0.53   Transmembrane  166 - 182 ( 166 - 182)

----- 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 databases:

```
>GP:CAB90755 GB:AJ400707 hypothetical protein [Streptococcus uberis]
 Identities = 72/128 (56%), Positives = 94/128 (73%)

Query: 106 FLLLYVLSGVMGNAFTFWLTPETVAAGASTSLFGLFAAIVVLSFLGKNQALKDLGKSYQT 165
           FLLLYVLSG+MGN  T + TP  VAAGASTSLFGLFAAIVV+ + G NQ LK +GKSYQT
Sbjct: 95  FLLLYVLSGIMGNVLTLFFTPHVVAAGASTSLFGLFAAIVVVGYFGHNQLLKSIGKSYQT 154

Query: 166 LIVVNLLMNLFMPNVSMAGHIGGVVGGALLSIVFPTKMRVITVKKTKRMLALVSYGIILV 225
           LI++NL+MNLFMPNV + GH+GG +GGAL ++  PT +       K ++  AL+SY  + +
Sbjct: 155 LIILNLVMNLFMPNVGIVGHLGGALGGALAAVFLPTLLDAELFTKKQKTSALLSYLTLAL 214

Query: 226 GVLVLGFL 233
           ++ L  +
Sbjct: 215 VLITLSLM 222
```

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

```
 Identities = 63/132 (47%), Positives = 92/132 (68%)

Query: 94  GSLRFLILYILSGLMGNIFTLFFTPHVVAAGASTSLFGVFSAIAIAGYFGKNPYLKQVGK 153
           G   FL+LY+LSG+MGN FT + TP  VAAGASTSLFG+F+AI +  + GKN  LK +GK
Sbjct: 102 GLTPFLLLYVLSGVMGNAFTFWLTPETVAAGASTSLFGLFAAIVVLSFLGKNQALKDLGK 161

Query: 154 SYQVMILLNLFFNIFTPGVSLAGHVGGLVGGVLVAIFLTKQNGSLLFKTWQSILALMIFI 213
           SYQ +I++NL  N+F P VS+AGH+GG+VGG L++I     +   K  + +LAL+ +
Sbjct: 162 SYQTLIVVNLLMNLFMPNVSMAGHIGGVVGGALLSIVFPTKMRVITVKKTKRMLALVSYG 221

Query: 214 IVSISLIGLSLV 225
           I+ + ++ L  +
Sbjct: 222 IILVGVLVLGFL 233
```

A further corresponding DNA sequence was identified in *S.pyogenes* <SEQ ID 9083> which encodes the amino acid sequence <SEQ ID 9084>. Analysis of this protein sequence reveals the following:

```
      Possible site: 52

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.70   Transmembrane   12 -  28 (   7 -  30)

----- Final Results -----
               bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS sequences follows:

```
 Score = 74.5 bits (180), Expect = 5e-16
 Identities = 37/96 (38%), Positives = 48/96 (49%)

Query: 1  MTQLLKRYPXXXXXXXXXXXXXXXAMQVVYGHLATGAQAIYQVGGMFGLLVKAMPDQLWRL 60
          M +  K YP                MQ+ YG  A  +Q I+Q GG+ G  +KA P  LWRL
Sbjct: 3  MKKFAKEYPTTVLLVSLTTLVFLLMQLTYGSQAESSQVIFQFGGIQGDYLKAYPTNLWRL 62

Query: 61 VTPXXXXXXXXXXXVNGLTLYFVGQIVEDLWGSRLF 96
          ++P           +NGL LYFVGQ+ E +WGS  F
Sbjct: 63 ISPIFVHIGWEHFLLNGLALYFVGQMGESIWGSLRF 98
```

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

Example 271

A DNA sequence (GBSx0296) was identified in *S.agalactiae* <SEQ ID 867> which encodes the amino acid sequence <SEQ ID 868>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2055(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:BAA28715 GB:AB001562 hypothetical protein [Streptococcus mutans]
 Identities = 96/173 (55%), Positives = 129/173 (74%)

Query: 1   MEKKLLRKEVLITLKSQPQAYKSEVDCKLLEAFIKTKAYQNSCVIATYLSFDYEYNTQLL 60
           M KK  R +V+  LK Q +A K   D +LLE  I+ +AYQ + VIATYL+F +E++T LL
Sbjct: 1   MMKKDYRTQVIEDLKKQDKAKKVLRDEQLLEELIQLEAYQKAHVIATYLAFPFEFDTSLL 60

Query: 61  IKQALCDGKRVLVPKTYPKGKMIFVDYQKDNLRTTPFGLLEPVNDRAVEKASIDLIHVPG 120
           I+QA  D K ++VPKTYP+ KMIFV Y + +L+ T FGL EP ++ A+EK++IDLIHVPG
Sbjct: 61  IEQAQRDNKSIVVPKTYPQRKMIFVVYDEADLQITKFGLKEPRSEEALEKSAIDLIHVPG 120

Query: 121 LIFNNKGFRIGYGAGYFDRYLSDFEGDTISTIYRCQRQDFVEEKHDVAVKEVL 173
           L FNN+G+RIG+GAGY+D+YL+DF+GDT+STIY  Q+  F     D+ VKEVL
Sbjct: 121 LAFNNEGYRIGFGAGYYDQYLADFQGDTVSTIYSFQQFTFEPSFFDIPVKEVL 173
```

A related GBS nucleic acid sequence <SEQ ID 10925> which encodes amino acid sequence <SEQ ID 10926> was also identified.

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

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

Example 272

A DNA sequence (GBSx0297) was identified in *S.agalactiae* <SEQ ID 869> which encodes the amino acid sequence <SEQ ID 870>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.44   Transmembrane  161 - 177 ( 161 - 177)
   INTEGRAL    Likelihood = -0.22   Transmembrane   29 -  45 (  28 -  45)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9305> which encodes amino acid sequence <SEQ ID 9306> was also identified.

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

```
>GP:AAD33517 GB:AF132127 glucose-6-phosphate isomerase
           [Streptococcus mutans]
 Identities = 344/401 (85%), Positives = 374/401 (92%)

Query: 1   MDLPENYDKEEFSRIQKAAEKIKSDSEVLVVIGIGGSYLGAKAAIDFLNNHFANLQTAEE 60
           ++LP+NYDKEEF+RI+KAAEKIKSDSEVLVVIGIGGSYLGA+AAIDFLN+ F NL+  EE
Sbjct: 49  LNLPQNYDKEEFARIKKAAEKIKSDSEVLVVIGIGGSYLGARAAIDFLNSSFVNLENKEE 108

Query: 61  RKAPQILYAGNSISSTYLADLVEYVQDKEFSVNVISKSGTTTEPAIAFRVFKELLVKKYG 120
           RKAPQILYAGNSISS YLADLV+YV DK+FSVNVISKSGTTTEPAIAFRVFK+LLVKKYG
Sbjct: 109 RKAPQILYAGNSISSNYLADLVDYVADKDFSVNVISKSGTTTEPAIAFRVFKDLLVKKYG 168

Query: 121 QEEANKRIYATTDKVKGAVKVEADANNWETFVVPDNVGGRFSVLTAVGLLPIAASGADIT 180
           QEEAN+RIYATTD+VKGAVKVEADAN WETFVVPD+VGGRF+VLTAVGLLPIAASGAD+
Sbjct: 169 QEEANQRIYATTDRVKGAVKVEADANGWETFVVPDSVGGRFTVLTAVGLLPIAASGADLD 228

Query: 181 ALMEGANAARKDLSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGEWWKQL 240
            LM GA AAR+D SS ++SEN AYQYAA+RN+LYRKGY+TE+LANYEPSLQYF EWWKQL
Sbjct: 229 QLMAGAEAARQDYSSAELSENEAYQYAAIRNILYRKGYVTEVLANYEPSLQYFSEWWKQL 288

Query: 241 AGESEGKDQKGIYPTSANFSTDLHSLGQFIQEGYRNLFETVVRVEKPRKNVTIPELTEDL 300
           AGESEGKDQKGIYPTSANFSTDLHSLGQFIQEG RNLFETV+RVEK RKN+ +PE  EDL
Sbjct: 289 AGESEGKDQKGIYPTSANFSTDLHSLGQFIQEGNRNLFETVIRVEKARKNILVPEAAEDL 348

Query: 301 DGLGYLQGKDVDFVNKKATDGVLLAHTDGGVPNMFVTLPTQDAYTLGYTIYFFELAIGLS 360
           DGL YLQGKDVDFVNKKATDGVLLAHTDGGVPN F+T+P QD +TLGY IYFFELAIGLS
Sbjct: 349 DGLAYLQGKDVDFVNKKATDGVLLAHTDGGVPNTFLTIPEQDEFTLGYVIYFFELAIGLS 408

Query: 361 GYLNSVNPFDQPGVEAYKRNMFALLGKPGFEELSAELNARL 401
           GYLN VNPFDQPGVEAYK+NMFALLGKPGFEEL AELNARL
Sbjct: 409 GYLNGVNPFDQPGVEAYKKNMFALLGKPGFEELGAELNARL 449
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 871> which encodes the amino acid sequence <SEQ ID 872>. Analysis of this protein sequence reveals the following:

```
      Possible site: 31

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.44   Transmembrane  209 - 225 ( 209 - 225)
```

```
   INTEGRAL    Likelihood = -0.22   Transmembrane   77 -  93 (  76 -  93)

----- Final Results -----
               bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAD33517 GB:AF132127 glucose-6-phosphate isomerase
          [Streptococcus mutans]
 Identities = 369/449 (82%), Positives = 408/449 (90%)

Query: 1   MSHITFDYSKVLESFAGQHEIDFLQGQVTEADKLLREGTGPGSDFLGWLDLPENYDKDEF 60
           M+HI FDYSKVL  F   HE+D++Q QVT AD+ LR+GTGPG++  GWL+LP+NYDK+EF
Sbjct: 1   MTHIKFDYSKVLGKFLASHELDYIQMQVTAADEALRKGTGPGAEMTGWLNLPQNYDKEEF 60

Query: 61  ARILTAAEKIKADSEVLVVIGIGGSYLGAKAAIDFLNHHFANLQTAKERKAPQILYAGNS 120
           ARI  AAEKIK+DSEVLVVIGIGGSYLGA+AAIDFLN  F NL+  +ERKAPQILYAGNS
Sbjct: 61  ARIKKAAEKIKSDSEVLVVIGIGGSYLGARAAIDFLNSSFVNLENKEERKAPQILYAGNS 120

Query: 121 ISSTYLADLVEYVQDKEFSVNVISKSGTTTEPAIAFRVFKELLVKKYGQEEANKRIYATT 180
           ISS YLADLV+YV DK+FSVNVISKSGTTTEPAIAFRVFK+LLVKKYGQEEAN+RIYATT
Sbjct: 121 ISSNYLADLVDYVADKDFSVNVISKSGTTTEPAIAFRVFKDLLVKKYGQEEANQRIYATT 180

Query: 181 DKVKGAVKVEADANNWETFVVPDNVGGRFSVLTAVGLLPIAASGADITALMEGANAARKD 240
           D+VKGAVKVEADAN WETFVVPD+VGGRF+VLTAVGLLPIAASGAD+  LM GA AAR+D
Sbjct: 181 DRVKGAVKVEADANGWETFVVPDSVGGRFTVLTAVGLLPIAASGADLDQLMAGAEAARQD 240

Query: 241 LSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGEWWKQLAGESEGKDQKGI 300
            SS ++SEN AYQYAA+RN+LYRKGY+TE+LANYEPSLQYF EWWKQLAGESEGKDQKGI
Sbjct: 241 YSSAELSENEAYQYAAIRNILYRKGYVTEVLANYEPSLQYFSEWWKQLAGESEGKDQKGI 300

Query: 301 YPTSANFSTDLHSLGQFIQEGYRNLFETVIRVDNPRKNVIIPELAEDLDGLGYLQGKDVD 360
           YPTSANFSTDLHSLGQFIQEG RNLFETVIRV+  RKN+++PE AEDLDGL YLQGKDVD
Sbjct: 301 YPTSANFSTDLHSLGQFIQEGNRNLFETVIRVEKARKNILVPEAAEDLDGLAYLQGKDVD 360

Query: 361 FVNKKATDGVLLAHTDGGVPNMFVTLPAQDEFTLGYTIYFFELAIAVSGYMNAVNPFDQP 420
           FVNKKATDGVLLAHTDGGVPN F+T+P QDEFTLGY IYFFELAI +SGY+N VNPFDQP
Sbjct: 361 FVNKKATDGVLLAHTDGGVPNTFLTIPEQDEFTLGYVIYFFELAIGLSGYLNGVNPFDQP 420

Query: 421 GVEAYKRNMFALLGKPGFEALSAELNARL 449
           GVEAYK+NMFALLGKPGFE L AELNARL
Sbjct: 421 GVEAYKKNMFALLGKPGFEELGAELNARL 449
```

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

```
>GP:CAB90755 GB:AJ400707 hypothetical protein [Streptococcus
          uberis]
 Identities = 58/91 (63%), Positives = 69/91 (75%)

Query: 6   KRYPITIFLLGLTGLIFIAMQVVYGHLATGAQAIYQVGGMFGLLVKAMPDQLWRLVTPIF 65
           K  P+T F L +T L+FI MQV YG  A   Q ++Q GGMFGL+VK+MP QLWRLVTPIF
Sbjct: 5   KEKPVTFFFLSVTILLFIVMQVFYGSWAKSPQVVFQFGGMFGLVVKSMPSQLWRLVTPIF 64

Query: 66 IHIGFGHFFVNGLTLYFVGQIVEDLWGSRLF 96
           IHIG+ HF +N LTLYFVGQ+ E +WGSR F
Sbjct: 65 IHIGWEHFLINSLTLYFVGQLAESIWGSRFF 95
```

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

```
 Identities = 380/401 (94%), Positives = 392/401 (96%)

Query: 1   MDLPENYDKEEFSRIQKAAEKIKSDSEVLVVIGIGGSYLGAKAAIDFLNNHFANLQTAEE 60
           +DLPENYDK+EF+RI  AAEKIK+DSEVLVVIGIGGSYLGAKAAIDFLN+HFANLQTA+E
Sbjct: 49  LDLPENYDKDEFARILTAAEKIKADSEVLVVIGIGGSYLGAKAAIDFLNHHFANLQTAKE 108
```

```
Query: 61  RKAPQILYAGNSISSTYLADLVEYVQDKEFSVNVISKSGTTTEPAIAFRVFKELLVKKYG 120
           RKAPQILYAGNSISSTYLADLVEYVQDKEFSVNVISKSGTTTEPAIAFRVFKELLVKKYG
Sbjct: 109 RKAPQILYAGNSISSTYLADLVEYVQDKEFSVNVISKSGTTTEPAIAFRVFKELLVKKYG 168

Query: 121 QEEANKRIYATTDKVKGAVKVEADANNWETFVVPDNVGGRFSVLTAVGLLPIAASGADIT 180
           QEEANKRIYATTDKVKGAVKVEADANNWETFVVPDNVGGRFSVLTAVGLLPIAASGADIT
Sbjct: 169 QEEANKRIYATTDKVKGAVKVEADANNWETFVVPDNVGGRFSVLTAVGLLPIAASGADIT 228

Query: 181 ALMEGANAARKDLSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGEWWKQL 240
           ALMEGANAARKDLSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGEWWKQL
Sbjct: 229 ALMEGANAARKDLSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGEWWKQL 288

Query: 241 AGESEGKDQKGIYPTSANFSTDLHSLGQFIQEGYRNLFETVVRVEKPRKNVTIPELTEDL 300
           AGESEGKDQKGIYPTSANFSTDLHSLGQFIQEGYRNLFETV+RV+ PRKNV IPEL EDL
Sbjct: 289 AGESEGKDQKGIYPTSANFSTDLHSLGQFIQEGYRNLFETVIRVDNPRKNVIIPELAEDL 348

Query: 301 DGLGYLQGKDVDFVNKKATDGVLLAHTDGGVPNMFVTLPTQDAYTLGYTIYFFELAIGLS 360
           DGLGYLQGKDVDFVNKKATDGVLLAHTDGGVPNMFVTLP QD +TLGYTIYFFELAI +S
Sbjct: 349 DGLGYLQGKDVDFVNKKATDGVLLAHTDGGVPNMFVTLPAQDEFTLGYTIYFFELAIAVS 408

Query: 361 GYLNSVNPFDQPGVEAYKRNMFALLGKPGFEELSAELNARL 401
           GY+N+VNPFDQPGVEAYKRNMFALLGKPGFE LSAELNARL
Sbjct: 409 GYMNAVNPFDQPGVEAYKRNMFALLGKPGFEALSAELNARL 449
```

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

Example 273

A DNA sequence (GBSx0298) was identified in *S.agalactiae* <SEQ ID 873> which encodes the amino acid sequence <SEQ ID 874>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.66   Transmembrane  654 - 670 ( 653 - 671)
   INTEGRAL    Likelihood = -1.65   Transmembrane  113 - 129 ( 113 - 129)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9463> which encodes amino acid sequence <SEQ ID 9464> was also identified.

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

```
>GP:AAA81906 GB:U04863 alcohol dehydrogenase 2 [Entamoeba
          histolytica]
 Identities = 536/864 (62%), Positives = 663/864 (76%), Gaps = 3/864 (0%)

Query: 20  ETTDVALAIDTLVQNGLKALDEMR--QLNQEQVDYIVAKASVAALDAHGELALHAVEETG 77
           +T  V   I+ LV+    AL E    + QE++DYIV KASVAALD H  LA  AVEETG
Sbjct: 5   QTMTVDEHINQLVRKAQVALKEYLKPEYTQEKIDYIVKKASVAALDQHCALAAAAVEETG 64

Query: 78  RGVFEDKATKNLFACEHVVNNMRHTKTVGVIEEDDVTGLTLIAEPVGVVCGITPTTNPTS 137
           RG+FEDKATKN+FACEHV + MRH KTVG+I  D + G+T IAEPVGVVCG+TP TNPTS
Sbjct: 65  RGIFEDKATKNIFACEHVTHEMRHAKTVGIINVDPLYGITEIAEPVGVVCGVTPVTNPTS 124

Query: 138 TAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAAIAAGAPENCVQWIEQPSIDAT 197
           TAIFKSLIS+KTRNPI+F+FHPSA + S  AA+IVRDAAIAAGAPENC+QWIE   I+A+
Sbjct: 125 TAIFKSLISIKTRNPIVFSFHPSALKCSIMAAKIVRDAAIAAGAPENCIQWIEFGGIEAS 184

Query: 198 NALMNHDGIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAHDIVMSK 257
           N LMNH G+ATILATGGNAMVKAAYS GKPALGVGAGNVP Y+EK+ NI+QAA+D+VMSK
```

```
Sbjct: 185 NKLMNHPGVATILATGGNAMVKAAYSSGKPALGVGAGNVPTYIEKTCNIKQAANDVVMSK 244

Query: 258 SFDNGMVCASEQAVIIDKEIYKEFVEEFKSYHTYFVNKKEKALLEEFCFGAKANSKNCAG 317
           SFDNGM+CASEQA IIDKEIY + VEE K+   YF+N++EKA LE+F FG  A S +
Sbjct: 245 SFDNGMICASEQAAIIDKEIYDQVVEEMKTLGAYFINEEEKAKLEKFMFGVNAYSADVNN 304

Query: 318 AKLNPNIVGKSAVWIAEQAGFTVPEGTNILAAECTEVSEKEPLTREKLSPVIAVLKAEST 377
           A+LNP   G S  W AEQ G  VPE  NI+ A C EV   EPLTREKLSPV+A+LKAE+T
Sbjct: 305 ARLNPKCPGMSPQWFAEQVGIKVPEDCNIICAVCKEVGPNEPLTREKLSPVLAILKAENT 364

Query: 378 EDGVEKARQMVEFNGLGHSAAIHTKDADLAREFGTRIRAIRVIWNSPSTFGGIGDVYNAF 437
           +DG++KA  MVEFNG GHSAAIH+ D  +  ++   ++A R++ N+PS+ GGIG +YN
Sbjct: 365 QDGIDKAEAMVEFNGRGHSAAIHSNDKAVVEKYALTMKACRILHNTPSSQGGIGSIYNYI 424

Query: 438 LPSLTLGCGSYGRNSVGDNVSAINLLNIKKVGRRRNNMQWFKVPSKTYFERDSIQYLQKC 497
            PS TLGCGSYG NSV  NV+  NLLNIK++  RRNN+QWF+VP K +FE  SI+YL +
Sbjct: 425 WPSFTLGCGSYGGNSVSANVTYHNLLNIKRLADRRNNLQWFRVPPKIFFEPHSIRYLAEL 484

Query: 498 RDVERVMIVTDHAMVELGFLDRIIEQLDLRRNKVVYQIFAEVEPDPDITTVMKGTDLMRT 557
           +++ ++ IV+D  M +LG++DR+++ L  R N+V  +IF +VEPDP I TV KG  +M T
Sbjct: 485 KELSKIFIVSDRMMYKLGYVDRVMDVLKRRSNEVEIEIFIDVEPDPSIQTVQKGLAVMNT 544

Query: 558 FKPDTIIAIGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFPELGKKTKFVA 617
           F PD IIA+GGGS MDAAK+MWL YE PE DF  + QKF+D+RKRAFKFP +GKK + +
Sbjct: 545 FGPDNIIAIGGGSAMDAAKIMWLLYEHPEADFFAMKQKFIDLRKRAFKFPTMGKKARLIC 604

Query: 618 IPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALVMTVPGFIAADTGMDV 677
           IPTTSGTGSEVTPFAVISD     +KYP+ADYSLTP+VAIVDP   M++P    ADTG+DV
Sbjct: 605 IPTTSGTGSEVTPFAVISDHETGKKYPLADYSLTPSVAIVDPMFTMSLPKRAIADTGLDV 664

Query: 678 LTHATEAYVSQMANDYTDGLALQAIKIVFDYLERSVKDADFEAREKMHNASTMAGMAFAN 737
           L HATEAYVS MAN+YTDGLA +A+K+VF+ L +S  + D EAREKMHNA+T+AGMAFA+
Sbjct: 665 LVHATEAYVSVMANEYTDGLARRAVKLVFENLLKSY-NGDLEAREKMHNAATIAGMAFAS 723

Query: 738 AFLGISHSMAHKIGAQFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKYNYYRADEKYQD 797
           AFLG+ HSMAHK+GA FH  HGR  A+LLP+VIRYNG +P K A WPKYN+Y+AD++Y +
Sbjct: 724 AFLGMDHSMAHKVGAAFHLPHGRCVAVLLPHVIRYNGQKPRKLAMWPKYNFYKADQRYME 783

Query: 798 IAKLLGLPAATPEEAVESYAKAVYDLGTRLGIKMNFRDQGIDEKEWKEKSRELAFLAYED 857
           +A+++GL   TP E VE++AKA  +L          F+   IDE  W  K  E+A LA+ED
Sbjct: 784 LAQMVGLKCNTPAEGVEAFAKACEELMKATETITGFKKANIDEAAWMSKVPEMALLAFED 843

Query: 858 QCSPANPRLPMVDHMQEIIEDAYY 881
           QCSPANPR+PMV  M++I++ AYY
Sbjct: 844 QCSPANPRVPMVKDMEKILKAAYY 867
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 875> which encodes the amino acid sequence <SEQ ID 876>. Analysis of this protein sequence reveals the following:

```
      Possible site: 55

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.66   Transmembrane  643 - 659 ( 642 - 660)
   INTEGRAL    Likelihood = -1.81   Transmembrane  102 - 118 ( 102 - 118)

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

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

```
>GP:AAA81906 GB:U04863 alcohol dehydrogenase 2 [Entamoeba
          histolytica]
 Identities = 535/870 (61%), Positives = 669/870 (76%), Gaps = 3/870 (0%)

Query: 6   NTVETTSVSVTIDALVQKGLAALEEMRKLD--QEQVDYIVAKASVAALDAHGELAKHAYE 63
```

```
            +T +T +V   I+ LV+K   AL+E  K +  QE++DYIV KASVAALD H  LA  A E
Sbjct: 2    STQQTMTVDEHINQLVRKAQVALKEYLKPEYTQEKIDYIVKKASVAALDQHCALAAAAVE 61

Query: 64   ETGRGVFEDKATKHLFACEHVVNNMRHQKTVGIIEEDDVTGLTLIAEPVGVICGITPTTN 123
            ETGRG+FEDKATK++FACEHV + MRH KTVGII  D + G+T IAEPVGV+CG+TP TN
Sbjct: 62   ETGRGIFEDKATKNIFACEHVTHEMRHAKTVGIINVDPLYGITEIAEPVGVVCGVTPVTN 121

Query: 124  PTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAAIAAGAPENCVQWVETPSL 183
            PTSTAIFKSLIS+KTRNPI+F+FHPSA + S  AA+IVRDAAIAAGAPENC+QW+E   +
Sbjct: 122  PTSTAIFKSLISIKTRNPIVFSFHPSALKCSIMAAKIVRDAAIAAGAPENCIQWIEFGGI 181

Query: 184  EATNALMNHDGIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAHDIV 243
            EA+N LMNH G+ATILATGGNAMVKAAYS GKPALGVGAGNVP Y+EK+ NI+QAA+D+V
Sbjct: 182  EASNKLMNHPGVATILATGGNAMVKAAYSSGKPALGVGAGNVPTYIEKTCNIKQAANDVV 241

Query: 244  MSKSFDNGMVCASEQAVIIDKEIYDDFVAEFKSYHTYFVNKKEKALLEEFCFGAKANSKN 303
            MSKSFDNGM+CASEQA IIDKEIYD  V E K+   YF+N++EKA LE+F FG  A S +
Sbjct: 242  MSKSFDNGMICASEQAAIIDKEIYDQVVEEMKTLGAYFINEEEKAKLEKFMFGVNAYSAD 301

Query: 304  CAGAKLNPNIVGKPATWIAEQAGFTVPEGTNILAAECKEVSENEPLTREKLSPVIAVLKS 363
               A+LNP   G    W AEQ G  VPE  NI+ A CKEV  NEPLTREKLSPV+A+LK+
Sbjct: 302  VNNARLNPKCPGMSPQWFAEQVGIKVPEDCNIICAVCKEVGPNEPLTREKLSPVLAILKA 361

Query: 364  ESREDGVEKARQMVEFNGLGHSAAIHTADAELAKEFGTRIRAIRVIWNSPSTFGGIGDVY 423
            E+ +DG++KA  MVEFNG GHSAAIH+ D  + +++   ++A R++ N+PS+ GGIG +Y
Sbjct: 362  ENTQDGIDKAEAMVEFNGRGHSAAIHSNDKAVVEKYALTMKACRILHNTPSSQGGIGSIY 421

Query: 424  NAFLPSLTLGCGSYGRNAVGDNVSAINLLNIKKVGRRRNNMQWFKVPSKTYFERDSIQYL 483
            N   PS TLGCGSYG N+V  NV+  NLLNIK++  RRNN+QWF+VP K +FE  SI+YL
Sbjct: 422  NYIWPSFTLGCGSYGGNSVSANVTYHNLLNIKRLADRRNNLQWFRVPPKIFFEPHSIRYL 481

Query: 484  QKCRDVERVMIVTDHAMVELGFLDRIIEQLDLRRNKVVYQIFAEVEPDPDITTVMKGTEL 543
             + +++ ++ IV+D  M +LG++DR+++ L  R N+V  +IF +VEPDP I TV KG  +
Sbjct: 482  AELKELSKIFIVSDRMMYKLGYVDRVMDVLKRRSNEVEIEIFIDVEPDPSIQTVQKGLAV 541

Query: 544  MRTFKPDTIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFPELGKKTK 603
            M TF PD IIA+GGGS MDAAK+MWL YE PE DF  + QKF+D+RKRAFKFP +GKK +
Sbjct: 542  MNTFGPDNIIAIGGGSAMDAAKIMWLLYEHPEADFFAMKQKFIDLRKRAFKFPTMGKKAR 601

Query: 604  FVAIPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALVLTVPGFIAADTG 663
             + IPTTSGTGSEVTPFAVISD    +KYP+ADYSLTP+VAIVDP   +++P    ADTG
Sbjct: 602  LICIPTTSGTGSEVTPFAVISDHETGKKYPLADYSLTPSVAIVDPMFTMSLPKRAIADTG 661

Query: 664  MDVLTHATEAYVSQMANDFTDGLALQAIKIVFDNLEKSVKTADFEAREKMHNASTMAGMA 723
            +DVL HATEAYVS MAN++TDGLA +A+K+VF+NL KS    D EAREKMHNA+T+AGMA
Sbjct: 662  LDVLVHATEAYVSVMANEYTDGLAREAVKLVFENLLKSY-NGDLEAREKMHNAATIAGMA 720

Query: 724  FANAFLGISHSMAHKIGAQFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKYNYYRADEK 783
            FA+AFLG+ HSMAHK+GA FH  HGR  A+LLP+VIRYNG +P K A WPKYN+Y+AD++
Sbjct: 721  FASAFLGMDHSMAHKVGAAFHLPHGRCVAVLLPHVIRYNGQKPRKLAMWPKYNFYKADQR 780

Query: 784  YQDIAKLLGLPASTPEEAVESYAKAVYDLGCRVGIQMNFKAQGIDENEWKEHSRELAYLA 843
            Y ++A+++GL  +TP E VE++AKA  +L          FK   IDE  W     E+A LA
Sbjct: 781  YMELAQMVGLKCNTPAEGVEAFAKACEELMKATETITGFKKANIDEAAWMSKVPEMALLA 840

Query: 844  YEDQCSPANPRLPMVDHMQEIIEDAYYGYA 873
            +EDQCSPANPR+PMV  M++I++ AYY  A
Sbjct: 841  FEDQCSPANPRVPMVKDMEKILKAAYYPIA 870
```

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

```
Identities = 827/880 (93%), Positives = 852/880 (95%)

Query: 12   MTEKTKAVETTDVALAIDTLVQNGLKALDEMRQLNQEQVDYIVAKASVAALDAHGELALH 71
            MTE    VETT V++ ID LVQ GL AL+EMR+L+QEQVDYIVAKASVAALDAHGELA H
Sbjct: 1    MTEGHNTVETTSVSVTIDALVQKGLAALEEMRKLDQEQVDYIVAKASVAALDAHGELAKH 60

Query: 72   AVEETGRGVFEDKATKNLFACEHVVNNMRHTKTVGVIEEDDVTGLTLIAEPVGVVCGITP 131
```

```
           A EETGRGVFEDKATK+LFACEHVVNNMRH KTVG+IEEDDVTGLTLIAEPVGV+CGITP
Sbjct: 61  AYEETGRGVFEDKATKHLFACEHVVNNMRHQKTVGIIEEDDVTGLTLIAEPVGVICGITP 120

Query: 132 TTNPTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAAIAAGAPENCVQWIEQ 191
           TTNPTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAAIAAGAPENCVQW+E
Sbjct: 121 TTNPTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAAIAAGAPENCVQWVET 180

Query: 192 PSIDATNALMNHDGIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAH 251
           PS++ATNALMNHDGIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAH
Sbjct: 181 PSLEATNALMNHDGIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAH 240

Query: 252 DIVMSKSFDNGMVCASEQAVIIDKEIYKEFVEEFKSYHTYFVNKKEKALLEEFCFGAKAN 311
           DIVMSKSFDNGMVCASEQAVIIDKEIY +FV EFKSYHTYFVNKKEKALLEEFCFGAKAN
Sbjct: 241 DIVMSKSFDNGMVCASEQAVIIDKEIYDDFVAEFKSYHTYFVNKKEKALLEEFCFGAKAN 300

Query: 312 SKNCAGAKLNPNIVGKSAVWIAEQAGFTVPEGTNILAAECTEVSEKEPLTREKLSPVIAV 371
           SKNCAGAKLNPNIVGK A WIAEQAGFTVPEGTNILAAEC EVSE EPLTREKLSPVIAV
Sbjct: 301 SKNCAGAKLNPNIVGKPATWIAEQAGFTVPEGTNILAAECKEVSENEPLTREKLSPVIAV 360

Query: 372 LKAESTEDGVEKARQMVEFNGLGHSAAIHTKDADLAREFGTRIRAIRVIWNSPSTFGGIG 431
           LK+ES EDGVEKARQMVEFNGLGHSAAIHT DA+LA+EFGTRIRAIRVIWNSPSTFGGIG
Sbjct: 361 LKSESREDGVEKARQMVEFNGLGHSAAIHTADAELAKEFGTRIRAIRVIWNSPSTFGGIG 420

Query: 432 DVYNAFLPSLTLGCGSYGRNSVGDNVSAINLLNIKKVGRRRNNMQWFKVPSKTYFERDSI 491
           DVYNAFLPSLTLGCGSYGRN+VGDNVSAINLLNIKKVGRRRNNMQWFKVPSKTYFERDSI
Sbjct: 421 DVYNAFLPSLTLGCGSYGRNAVGDNVSAINLLNIKKVGRRRNNMQWFKVPSKTYFERDSI 480

Query: 492 QYLQKCRDVERVMIVTDHAMVELGFLDRIIEQLDLRRNKVVYQIFAEVEPDPDITTVMKG 551
           QYLQKCRDVERVMIVTDHAMVELGFLDRIIEQLDLRRNKVVYQIFAEVEPDPDITTVMKG
Sbjct: 481 QYLQKCRDVERVMIVTDHAMVELGFLDRIIEQLDLRRNKVVYQIFAEVEPDPDITTVMKG 540

Query: 552 TDLMRTFKPDTIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFPELGK 611
           T+LMRTFKPDTIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFPELGK
Sbjct: 541 TELMRTFKPDTIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFPELGK 600

Query: 612 KTKFVAIPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALVMTVPGFIAA 671
           KTKFVAIPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALV+TVPGFIAA
Sbjct: 601 KTKFVAIPTTSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALVLTVPGFIAA 660

Query: 672 DTGMDVLTHATEAYVSQMANDYTDGLALQAIKIVFDYLERSVKDADFEAREKMHNASTMA 731
           DTGMDVLTHATEAYVSQMAND+TDGLALQAIKIVFD LE+SVK ADFEAREKMHNASTMA
Sbjct: 661 DTGMDVLTHATEAYVSQMANDFTDGLALQAIKIVFDNLEKSVKTADFEAREKMHNASTMA 720

Query: 732 GMAFANAFLGISHSMAHKIGAQFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKYNYYRA 791
           GMAFANAFLGISHSMAHKIGAQFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKYNYYRA
Sbjct: 721 GMAFANAFLGISHSMAHKIGAQFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKYNYYRA 780

Query: 792 DEKYQDIAKLLGLPAATPEEAVESYAKAVYDLGTRLGIKMNFRDQGIDEKEWKEKSRELA 851
           DEKYQDIAKLLGLPA+TPEEAVESYAKAVYDLG R+GI+MNF+ QGIDE EWKE SRELA
Sbjct: 781 DEKYQDIAKLLGLPASTPEEAVESYAKAVYDLGCRVGIQMNFKAQGIDENEWKEHSRELA 840

Query: 852 FLAYEDQCSPANPRLPMVDHMQEIIEDAYYGYEERPGRRK 891
           +LAYEDQCSPANPRLPMVDHMQEIIEDAYYGY ERPGRRK
Sbjct: 841 YLAYEDQCSPANPRLPMVDHMQEIIEDAYYGYAERPGRRK 880
```

A related GBS gene <SEQ ID 8533> and protein <SEQ ID 8534> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      -4.68
GvH: Signal Score (-7.5): -2.48
     Possible site: 21
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -2.66 threshold:  0.0
   INTEGRAL    Likelihood = -2.66   Transmembrane  100 - 116 (  99 - 117)
   PERIPHERAL  Likelihood =  3.61     173
 modified ALOM score:   1.03
```

```
*** Reasoning Step: 3

----- 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>
```

SEQ ID 8534 (GBS432) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 5; MW 66kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 7; MW 41kDa).

GBS432-GST was purified as shown in Figure 223, lane 9.

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

Example 274

A DNA sequence (GBSx0299) was identified in *S.agalactiae* <SEQ ID 877> which encodes the amino acid sequence <SEQ ID 878>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3444(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 880.

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

Example 275

A DNA sequence (GBSx0300) was identified in *S.agalactiae* <SEQ ID 881> which encodes the amino acid sequence <SEQ ID 882>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL     Likelihood = -8.39   Transmembrane   74 -  90 (  69 -  94)
   INTEGRAL     Likelihood = -5.31   Transmembrane  168 - 184 ( 163 - 186)
   INTEGRAL     Likelihood = -4.83   Transmembrane   34 -  50 (  29 -  52)
   INTEGRAL     Likelihood = -0.75   Transmembrane  202 - 218 ( 202 - 219)

----- Final Results -----
              bacterial membrane --- Certainty=0.4354(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:CAA17305 GB:AL021926 hypothetical protein Rv0111 [Mycobacterium
          tuberculosis]
 Identities = 70/218 (32%), Positives = 104/218 (47%), Gaps = 12/218 (5%)

Query: 9   VRITGLLLVLLYHFFKNSFPGGFVGVDIFFTFSGFLITALLIDEFSKTKKIDFVSFCRRR 68
           +R   + LVL  H       GGF+GVD FF  SGFLIT+LL+DE  +T +ID   F  RR
```

```
Sbjct: 39  LRAIAVALVLASHGGIPGMGGGFIGVDAFFVLSGFLITSLLLDELGRTGRIDLSGFWIRR 98

Query: 69  FYRIFPPLVLMVLVTIPFVFLVKSDFRASIGSQIMTALGFTSNFYEILTGGNYESQFI-P 127
             R+ P LVLMVL        L         + S  + A  +T+N+  +     +Y +Q   P
Sbjct: 99  ARRLLPALVLMVLTVSAARALFPDQALTGLRSDAIAAFLWTANWRFVAQNTDYFTQGAPP 158

Query: 128 HLFVHTWSLSIEVHFYVLWGL----TVWLLSKRSKDQKQLRGTLFLISMGIFGVSFLTMF 183
               HTWSL +E  +YV+W L        LL+ R++ ++  R T+  +     F ++ L
Sbjct: 159 SPLQHTWSLGVEEQYYVVWPLLLIGATLLLAARAR-RRCRRATVGGVRFAAFLIASLGTM 217

Query: 184 VRAFFVDNFST------IYFSTLSHIFPFFLGAMVATI 215
             A     F++       IYF T +      +G+  A +
Sbjct: 218 ASATAAVAFTSAATRDRIYFGTDTRAQALLIGSAAAAL 255
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 879> which encodes the amino acid sequence <SEQ ID 880>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.83   Transmembrane  325 - 341 ( 313 - 346)
   INTEGRAL    Likelihood = -9.29   Transmembrane  237 - 253 ( 234 - 258)
   INTEGRAL    Likelihood = -7.91   Transmembrane  166 - 182 ( 162 - 188)
   INTEGRAL    Likelihood = -6.10   Transmembrane   72 -  88 (  68 -  92)
   INTEGRAL    Likelihood = -4.09   Transmembrane  264 - 280 ( 260 - 281)
   INTEGRAL    Likelihood = -2.87   Transmembrane  371 - 387 ( 370 - 390)
   INTEGRAL    Likelihood = -2.66   Transmembrane   34 -  50 (  32 -  50)
   INTEGRAL    Likelihood = -1.91   Transmembrane    3 -  19 (   3 -  19)
   INTEGRAL    Likelihood = -0.85   Transmembrane  136 - 152 ( 136 - 154)

----- Final Results -----
              bacterial membrane --- Certainty=0.5331(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 = 167/226 (73%), Positives = 195/226 (85%)

Query: 1   MRIKWFSLVRITGLLLVLLYHFFKNSFPGGFVGVDIFFTFSGFLITALLIDEFSKTKKID 60
            MRIKWFS VR+TGLLLVLLYHFFKN FPGGF+GVDIFFTFSG+LITALLIDE++K + ID
Sbjct: 1   MRIKWFSFVRVTGLLLVLLYHFFKNVFPGGFIGVDIFFTFSGYLITALLIDEYTKKESID 60

Query: 61  FVSFCRRRFYRIFPPLVLMVLVTIPFVFLVKSDFRASIGSQIMTALGFTSNFYEILTGGN 120
             + F +RRFYRI PPLVLM+L+TIPF FL+K DF A+IGSQI    LGFT+N YEILTG +
Sbjct: 61  IIGFLKRRFYRIVPPLVLMILLTIPFTFLIKKDFIANIGSQITAVLGFTTNIYEILTGSS 120

Query: 121 YESQFIPHLFVHTWSLSIEVHFYVLWGLTVWLLSKRSKDQKQLRGTLFLISMGIFGVSFL 180
            YESQFIPHLFVHTWSL+IEVHFY+ WG+ VWLL++R + QKQLRG LFLIS+GIF +SFL
Sbjct: 121 YESQFIPHLFVHTWSLAIEVHFYLFWGVFVWLLARRKETQKQLRGLLFLISLGIFAISFL 180

Query: 181 TMFVRAFFVDNFSTIYFSTLSHIFPFFLGAMVATISGIREITGRFK 226
            +MF+R+F   NFS IYFS+LSH FPFFLGAM ATI+GI E T RF+
Sbjct: 181 SMFIRSFMTSNFSLIYFSSLSHSFPFFLGAMFATITGINETTVRFQ 226
```

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

Example 276

A DNA sequence (GBSx0302) was identified in *S.agalactiae* <SEQ ID 883> which encodes the amino acid sequence <SEQ ID 884>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have a cleavable N-term signal seq.
```

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

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

```
!GB:AE004818 hypothetical protein [Pseudomonas aerug...
!GB:AE004818 hypothetical protein [Pseudomonas aerug...
!GB:AE004818 hypothetical protein [Pseudomonas aerug...
!GB:AE004818 hypothetical protein [Pseudomonas aerug...
!GB:AE004818 hypothetical protein [Pseudomonas aerug...

>GP:AAG07403 GB:AE004818 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 33/80 (41%), Positives = 50/80 (62%)

Query: 45  KYVGSIVNHHMTGKGKLTYENGDYYKGDFVNGVFEGKGTFVSVHGWSYTGDFKKGQPDGQ 104
           +Y G +V+  + G+G+L Y+NG +Y G F +G+  G GT+    G  Y+G F  G  DGQ
Sbjct: 39  RYRGELVDGRLEGQGRLDYDNGAWYAGRFEHGLLHGHGTWQGADGSRYSGGFAAGLFDGQ 98

Query: 105 GRLNAKNKKVYKGTFKQGIY 124
           GRL   +  VY+G F+QG++
Sbjct: 99  GRLAMADGSVYQGGFRQGLF 118
 Identities = 31/91 (34%), Positives = 46/91 (50%), Gaps = 2/91 (2%)

Query: 34  QGVFSYDGGKIKYVGSIVNHHMTGKGKLTYENGDYYKGDFVNGVFEGKGTFVSVHGWSYT 93
           QG   YD G   Y G   +  + G G     +G  Y G F  G+F+G+G      G  Y
Sbjct: 52  QGRLDYDNGAW-YAGRFEHGLLHGHGTWQGADGSRYSGGFAAGLFDGQGRLAMADGSVYQ 110

Query: 94  GDFKKGQPDGQGRLNAKNKKVYKGTFKQGIY 124
           G F++G  DG+G L   +  Y+G F++G+Y
Sbjct: 111 GGFRQGLFDGEGSLEQQGTR-YRGGFRKGLY 140
 Identities = 31/91 (34%), Positives = 42/91 (46%), Gaps = 1/91 (1%)

Query: 32  SSQGVFSYDGGKIKYVGSIVNHHMTGKGKLTYENGDYYKGDFVNGVFEGKGTFVSVHGWS 91
           S QG      G  +Y GS      + G+G   +   G+ Y G F +G   GKG +    G
Sbjct: 141 SGQGTLDGSDGS-RYQGSFRQGRLEGEGSFSDSQGNQYAGTFRDGQLNGKGRWSGPDGDR 199

Query: 92  YTGDFKKGQPDGQGRLNAKNKKVYKGTFKQG 122
           Y G FK  Q  GQGR  + +  V+ G F +G
Sbjct: 200 YVGQFKDNQFHGQGRYESASGDVWIGRFSEG 230
 Identities = 31/91 (34%), Positives = 45/91 (49%), Gaps = 4/91 (4%)

Query: 34  QGVFSYDGGK----IKYVGSIVNHHMTGKGKLTYENGDYYKGDFVNGVFEGKGTFVSVHG 89
           QG+F  +G        +Y G        +G+G L   +G  Y+G F  G  EG+G+F    G
Sbjct: 115 QGLFDGEGSLEQQGTRYRGGFRKGLYSGQGTLDGSDGSRYQGSFRQGRLEGEGSFSDSQG 174

Query: 90  WSYTGDFKKGQPDGQGRLNAKNKKVYKGTFK 120
             Y G F+ GQ +G+GR +  +   Y G FK
Sbjct: 175 NQYAGTFRDGQLNGKGRWSGPDGDRYVGQFK 205
 Identities = 28/87 (32%), Positives = 45/87 (51%), Gaps = 1/87 (1%)

Query: 34  QGVFSYDGGKIKYVGSIVNHHMTGKGKLTYENGDYYKGDFVNGVFEGKGTFVSVHGWSYT 93
           +G FS   G  +Y G+  +  + GKG+ +  +GD Y G F +  F G+G + S  G  +
Sbjct: 166 EGSFSDSQGN-QYAGTFRDGQLNGKGRWSGPDGDRYVGQFKDNQFHGQGRYESASGDVWI 224

Query: 94  GDFKKGQPDGQGRLNAKNKKVYKGTFK 120
           G F +G  +G+G L   +   Y+G F+
Sbjct: 225 GRFSEGALNGPGELLGADGSRYRGGFQ 251
 Identities = 28/89 (31%), Positives = 43/89 (47%), Gaps = 2/89 (2%)

Query: 34  QGVFSYDGGKIKYVGSIVNHHMTGKGKLTYENGDYYKGDFVNGVFEGKGTFVSVHGWSYT 93
           QG  +  G + Y G        G+G L  + G  Y+G F  G+  G+G+F    G  Y
Sbjct: 98  QGRLAMADGSV-YQGGFRQGLFDGEGSLE-QQGTRYRGGFRKGLYSGQGTLDGSDGSRYQ 155

Query: 94  GDFKKGQPDGQGRLNAKNKKVYKGTFKQG 122
           G F++G  DG+G L   +  VY G F++G
Sbjct: 156 GSFRQGRLEGEGSFSDSQGNQYAGTFRDG 184
```

```
 Identities = 25/80 (31%), Positives = 37/80 (46%)

Query: 45  KYVGSIVNHHMTGKGKLTYENGDYYKGDFVNGVFEGKGTFVSVHGWSYTGDFKKGQPDGQ 104
           +YVG   ++   G+G+   +GD + G F  G   G G  +   G  Y G F+  +  GQ
Sbjct: 199 RYVGQFKDNQFHGQGRYESASGDVWIGRFSEGALNGPGELLGADGSRYRGGFQFWRFHGQ 258

Query: 105 GRLNAKNKKVYKGTFKQGIY 124
           G L   +   Y+G F  G Y
Sbjct: 259 GLLEQLDGTRYEGGFAAGAY 278
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 885> which encodes the amino acid sequence <SEQ ID 886>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.16   Transmembrane   20 -  36 (  12 -  41)

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

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

```
>GP:BAA16606 GB:D90899 hypothetical protein [Synechocystis sp.]
 Identities = 37/89 (41%), Positives = 49/89 (54%), Gaps = 6/89 (6%)

Query: 48  KGRMHYT------GYVINHKMNGEGKLVYPNGDIYEGTFKDGLFEGKGTFTAKTGWLYNG 101
           KG   YT      G V+   ++NG GK  Y NGD YEGT K+G  +G+G F    G  Y G
Sbjct: 141 KGTFIYTNGDRCSGTVVQGELNGSGKCEYNNGDQYEGTLKNGQPDGEGIFRFAAGGEYEG 200

Query: 102 EFHKGQANGKGVLKAKNNKVYKGIFKQGI 130
           EF  G+ +G+G      N   ++G FKQG+
Sbjct: 201 EFQSGEFSGQGTRIFANGNRFQGQFKQGL 229
```

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

```
 Identities = 68/126 (53%), Positives = 93/126 (72%)

Query: 1   MKNFKITRTHLEILSLIIIVVFGLSVFTLTTSSQGVFSYDGGKIKYVGSIVNHHMTGKGK 60
           +K + ITR  LEI+S+I+I+V  +SVF++  S++   +YD G++ Y G ++NH M G+GK
Sbjct: 8   VKKWSITRAKLEIVSVIVILVCAISVFSVRISNKTSLTYDKGRMHYTGYVINHKMNGEGK 67

Query: 61  LTYENGDYYKGDFVNGVFEGKGTFVSVHGWSYTGDFKKGQPDGQGRLNAKNKKVYKGTFK 120
           L Y NGD Y+G F +G+FEGKGTF +  GW Y G+F KGQ +G+G L AKN KVYKG FK
Sbjct: 68  LVYPNGDIYEGTFKDGLFEGKGTFTAKTGWLYNGEFHKGQANGKGVLKAKNNKVYKGIFK 127

Query: 121 QGIYQK 126
           QGI+QK
Sbjct: 128 QGIFQK 133
```

SEQ ID 884 (GBS139) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 3; MW 13kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 22 (lane 2; MW 38.2kDa), in Figure 24 (lane 7; MW 38kDa) and in Figure 33 (lane 7; MW 38.2kDa).

The GBS139-GST fusion product was purified (Figure 200, lane 2) and used to immunise mice. The resulting antiserum was used for FACS (Figure 287), 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 277

A DNA sequence (GBSx0303) was identified in *S.agalactiae* <SEQ ID 887> which encodes the amino acid sequence <SEQ ID 888>. This protein is predicted to be holliday junction dna helicase ruvb (ruvB). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4386(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:CAB75331 GB:Y15896 RuvB protein [Bacillus subtilis]
 Identities = 196/322 (60%), Positives = 254/322 (78%)

Query: 3   RFLDSDAMGDEELVERTLRPQYLREYIGQDKVKDQLKIFIEAAKLRDESLDHVLLFGPPG 62
           R + S+A   E ++E++LRPQ L +YIGQ KVK+ L++FI+AAK+R E+LDHVLL+GPPG
Sbjct: 4   RLVSSEADNHESVIEQSLRPQNLAQYIGQHKVKENLRVFIDAAKMRQETLDHVLLYGPPG 63

Query: 63  LGKTTMAFVIANELGVNLKQTSGPAIEKSGDLVAILNDLEPGDVLFIDEIHRMPMAVEEV 122
           LGKTT+A ++ANE+GV L+ TSGPAIE+ GDL AIL  LEPGDVLFIDEIHR+  ++EEV
Sbjct: 64  LGKTTLASIVANEMGVELRTTSGPAIERPGDLAAILTALEPGDVLFIDEIHRLHRSIEEV 123

Query: 123 LYSAMEDFYIDIMIGAGETSRSVHLDLPPFTLIGATTRAGMLSNPLRARFGITGHMEYYE 182
           LY AMEDF +DI+IG G ++RSV LDLPPFTL+GATTR G+L+ PLR RFG+   +EYY
Sbjct: 124 LYPAMEDFCLDIVIGKGPSARSVRLDLPPFTLVGATTRVGLLTAPLRDRFGVMSRLEYYT 183

Query: 183 ENDLTEIIERTADIFEMKITYEAASELARRSRGTPRIANRLLKRVRDYAQIMGDGLIDDN 242
           + +L +I+ RTAD+FE++I   +A E+ARRSRGTPR+ANRLL+RVRD+AQ++GD  I ++
Sbjct: 184 QEELADIVTRTADVFEVEIDKPSALEIARRSRGTPRVANRLLRRVRDFAQVLGDSRITED 243

Query: 243 ITDKALTMLDVDHEGLDYVDQKILRTMIEMYNGGPVGLGTLSVNIAEERDTVEDMYEPYL 302
           I+  AL  L VD  GLD++D K+L  MIE +NGGPVGL T+S  I EE  T+ED+YEPYL
Sbjct: 244 ISQNALERLQVDRLGLDHIDHKLLMGMIEKFNGGPVGLDTISATIGEESHTIEDVYEPYL 303

Query: 303 IQKGFIMRTRTGRVATVKAYEH 324
           +Q GFI RT  GR+ T   Y H
Sbjct: 304 LQIGFIQRTPRGRIVTPAVYHH 325
```

A related GBS nucleic acid sequence <SEQ ID 10943> which encodes amino acid sequence <SEQ ID 10944> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 889> which encodes the amino acid sequence <SEQ ID 890>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0686(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 = 282/327 (86%), Positives = 306/327 (93%)
```

```
Query: 1   MTRFLDSDAMGDEELVERTLRPQYLREYIGQDKVKDQLKIFIEAAKLRDESLDHVLLFGP 60
           M R LD++ MG+EE  +RTLRPQYL EYIGQDKVK+Q  IFIEAAK RDESLDHVLLFGP
Sbjct: 25  MARILDNNVMGNEEFSDRTLRPQYLHEYIGQDKVKEQFAIFIEAAKRRDESLDHVLLFGP 84

Query: 61  PGLGKTTMAFVIANELGVNLKQTSGPAIEKSGDLVAILNDLEPGDVLFIDEIHRMPMAVE 120
           PGLGKTTMAFVIANELGVNLKQTSGPA+EK+GDLVAILN+LEPGD+LFIDEIHRMPM+VE
Sbjct: 85  PGLGKTTMAFVIANELGVNLKQTSGPAVEKAGDLVAILNELEPGDILFIDEIHRMPMSVE 144

Query: 121 EVLYSAMEDFYIDIMIGAGETSRSVHLDLPPFTLIGATTRAGMLSNPLRARFGITGHMEY 180
           EVLYSAMEDFYIDIMIGAG+TSRS+HLDLPPFTLIGATTRAGMLSNPLRARFGITGHMEY
Sbjct: 145 EVLYSAMEDFYIDIMIGAGDTSRSIHLDLPPFTLIGATTRAGMLSNPLRARFGITGHMEY 204

Query: 181 YEENDLTEIIERTADIFEMKITYEAASELARRSRGTPRIANRLLKRVRDYAQIMGDGLID 240
           Y+E DLTEI+ERTA IFE+KI +EAA +LA RSRGTPRIANRLLKRVRDYAQI+GDG+I
Sbjct: 205 YQEKDLTEIVERTATIFEIKIDHEAARKLACRSRGTPRIANRLLKRVRDYAQIIGDGIIT 264

Query: 241 DNITDKALTMLDVDHEGLDYVDQKILRTMIEMYNGGPVGLGTLSVNIAEERDTVEDMYEP 300
             ITD+ALTMLDVD EGLDY+DQKILRTMIEMY GGPVGLGTLSVNIAEER+TVE+MYEP
Sbjct: 265 AQITDRALTMLDVDREGLDYIDQKILRTMIEMYQGGPVGLGTLSVNIAEERNTVEEMYEP 324

Query: 301 YLIQKGFIMRTRTGRVATVKAYEHLGY 327
           YLIQKGF+MRTRTGRVAT KAY HLGY
Sbjct: 325 YLIQKGFLMRTRTGRVATQKAYRHLGY 351
```

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

Example 278

A DNA sequence (GBSx0304) was identified in *S.agalactiae* <SEQ ID 891> which encodes the amino acid sequence <SEQ ID 892>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.87   Transmembrane  157 - 173 ( 157 - 174)
   INTEGRAL    Likelihood = -1.49   Transmembrane  205 - 221 ( 205 - 222)

----- Final Results -----
               bacterial membrane --- Certainty=0.2147(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 893> which encodes the amino acid sequence <SEQ ID 894>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3097(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 = 130/303 (42%), Positives = 202/303 (65%)

Query: 1   MLKHFGSKVRNLRVTRNITREDFCGDETELSVRQLARIESGQSIPNLTKAHYIAKQLNVK 60
           ML+HFG KV+ LR+ + I+RED CGDE+ELSVRQLARIE GQSIP+L+K  +IAK LNV
Sbjct: 1   MLEHFGGKVKVLRLEKRISREDLCGDESELSVRQLARIELGQSIPSLSKVIFIAKALNVS 60

Query: 61  LDILTGGESLELPKRYKELKYLILRIPTYADAERLKLRECQFDHIFEEFYDNLPEDECLA 120
           +  LT G  LELPKRYKELKYLILR PTY D  +L++RE QFD IFE++YD LPE+E +
```

```
Sbjct: 61  VGYLTDGADLELPKRYKELKYLILRTPTYMDDGKLQVREEQFDEIFEDYYDKLPEEEKII 120

Query: 121 IDSLQAKFEVYQTGDINFGVEVLCECFDKVKYKEKYTLNDLIIIDLFLTCAVVSKFNNRA 180
           ID LQA  +   + + NFG+++L E F+++K K ++  NDLI+++L+L    +   + +
Sbjct: 121 IDCLQATLDTLLSENTNFGIDLLQEYFNQIKTKVRFRQNDLILLELYLAYLDIEGMDGQY 180

Query: 181 FTKEVFQTICKTLISQNHKLTAEDLFWFNHVLLNCVFVGLCLNSEECLAEMLEVSRQTMV 240
             K  + ++   L  Q  +   ++LF  N ++++   + L  N  + L + +E+S++ M
Sbjct: 181 SDKIFYDSLLDNLSEQFEQFELDELFIVNKIIIDISSLSLKNNRLDNLEKAIEMSQKIMA 240

Query: 241 STHDFHKMPLYFMYQWKYFITIDNDIKSAENAYQQSIMFSKMIDDKHLIKKLELEWQEDI 300
             D+++MP+  + +WKYF+    DI  AE ++ ++ +F++M  D++L  KL  EW++D+
Sbjct: 241 KIQDWNRMPILKLIEWKYFLIKQKDIIKAEQSFMKACLFAQMTADQYLENKLIQEWEKDV 300

Query: 301 TGH 303
           +
Sbjct: 301 KSY 303
```

SEQ ID 892 (GBS319) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 40 (lane 4; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 7; MW 62kDa).

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

Example 279

A DNA sequence (GBSx0305) was identified in *S.agalactiae* <SEQ ID 895> which encodes the amino acid sequence <SEQ ID 896>. This protein is predicted to be adenylosuccinate lyase (purB). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3358(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:BAB04344 GB:AP001509 adenylosuccinate lyase [Bacillus halodurans]
 Identities = 326/430 (75%), Positives = 366/430 (84%)

Query: 1   MIERYSRPEMAAIWTEENKYRAWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL 60
           MIERY+RPEM AIWTEEN+Y+AWLEVEI+A EAWAELGEIPKEDV KIRE A FD++RIL
Sbjct: 1   MIERYTRPEMGAIWTEENRYQAWLEVEIVACEAWAELGEIPKEDVKKIREHASFDVERIL 60

Query: 61  EIEQDTRHDVVAFTRAVSETLGEERKWVHYGLTSTDVVDTAYGYLYKQANDIIRRDLENF 120
           EIEQ+TRHDVVAFTRAVSETLGEERKWVHYGLTSTDVVDTA  YL KQAN+II  DL  F
Sbjct: 61  EIEQETRHDVVAFTRAVSETLGEERKWVHYGLTSTDVVDTALSYLLKQANEIIEADLVRF 120

Query: 121 TNIVADKAKEHKFTIMMGRTHGVHAEPTTFGLKLATWYSEMKRNIERFEHAAAGVEAGKI 180
            +I+ +KA EHK+T+MMGRTHGVHAEPTTFGLKLA WY EMKRN+ERF  AA GV  GK+
Sbjct: 121 LDILKEKALEHKYTVMMGRTHGVHAEPTTFGLKLALWYEEMKRNLERFRLAAEGVRVGKL 180

Query: 181 SGAVGNFANIPPFVEQYVCDKLGIRPQEISTQVLPRDLHAEYFAVLASIATSIERMATEI 240
           SGAVG +ANI PFVEQYVC+KLG+    ISTQ L RD HAEY A LA IATSIE+ A EI
Sbjct: 181 SGAVGTYANIDPFVEQYVCEKLGLERAPISTQTLQRDRHAEYMATLALIATSIEKFAVEI 240

Query: 241 RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENVALWHE 300
           RGLQKSE REVEE+FAKGQKGSSAMPHKRNPIGSENMTG+ARV+RGHM+ AYENV LWHE
Sbjct: 241 RGLQKSETREVEEYFAKGQKGSSAMPHKRNPIGSENMTGIARVVRGHMLAAYENVPLWHE 300
```

```
Query: 301 RDISHSSAERIITPDTTILIDYMLNRFGNIVKNLTVFPENMMRNMESTFGLIYSQRVMLK 360
           RDISHSSAERII PD TI I+YMLNRFGNIVKNLTVFPENM RNM  T+GLIYSQRV+L
Sbjct: 301 RDISHSSAERIILPDATIAINYMLNRFGNIVKNLTVFPENMKRNMTRTYGLIYSQRVLLS 360

Query: 361 LIEKGMTREEAYDLVQPKTAYSWDNQVDFKPLLEEDTKVTSCLTQEEIDELFNPIYYTKR 420
           LI+KGM REEAYDLVQPK   +W+  V F+ L+E++ ++TS L+ EEI+  F+  ++ K
Sbjct: 361 LIDKGMVREEAYDLVQPKAMEAWEKGVQFRELVEQEERITSVLSPEEIEACFDYNHHLKH 420

Query: 421 VDDIFERLGL 430
           VD IFERLGL
Sbjct: 421 VDTIFERLGL 430
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 897> which encodes the amino acid sequence <SEQ ID 898>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3358(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 = 422/430 (98%), Positives = 428/430 (99%)

Query: 1   MIERYSRPEMAAIWTEENKYRAWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL 60
           M+ERYSRPEMAAIWTEENKY AWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL
Sbjct: 1   MLERYSRPEMAAIWTEENKYHAWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL 60

Query: 61  EIEQDTRHDVVAFTRAVSETLGEERKWVHYGLTSTDVVDTAYGYLYKQANDIIRRDLENF 120
           EIEQDTRHDVVAFTRAVSETLGEERKWVHYGLTSTDVVDTAYGYLYKQANDIIRRDLENF
Sbjct: 61  EIEQDTRHDVVAFTRAVSETLGEERKWVHYGLTSTDVVDTAYGYLYKQANDIIRRDLENF 120

Query: 121 TNIVADKAKEHKFTIMMGRTHGVHAEPTTFGLKLATWYSEMKRNIERFEHAAAGVEAGKI 180
           TNIVADKA+EHK TIMMGRTHGVHAEPTTFGLKLATWYSEMKRNIERFEHAAAGVEAGKI
Sbjct: 121 TNIVADKAREHKMTIMMGRTHGVHAEPTTFGLKLATWYSEMKRNIERFEHAAAGVEAGKI 180

Query: 181 SGAVGNFANIPPFVEQYVCDKLGIRPQEISTQVLPRDLHAEYFAVLASIATSIERMATEI 240
           SGAVGNFANIPPFVE+YVCDKLGIRPQEISTQVLPRDLHAEYFAVLASIATSIERMATEI
Sbjct: 181 SGAVGNFANIPPFVEEYVCDKLGIRPQEISTQVLPRDLHAEYFAVLASIATSIERMATEI 240

Query: 241 RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENVALWHE 300
           RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENV+LWHE
Sbjct: 241 RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENVSLWHE 300

Query: 301 RDISHSSAERIITPDTTILIDYMLNRFGNIVKNLTVFPENMMRNMESTFGLIYSQRVMLK 360
           RDISHSSAERIITPDTTILIDYMLNRFGNIVKNLTVFPENMMRNMESTFGLIYSQRVMLK
Sbjct: 301 RDISHSSAERIITPDTTILIDYMLNRFGNIVKNLTVFPENMMRNMESTFGLIYSQRVMLK 360

Query: 361 LIEKGMTREEAYDLVQPKTAYSWDNQVDFKPLLEEDTKVTSCLTQEEIDELFNPIYYTKR 420
           LIEKGMTREEAYDLVQPKTAYSWDNQVDFKPLLEEDTKVTSCLTQEEIDELFNPIYYTKR
Sbjct: 361 LIEKGMTREEAYDLVQPKTAYSWDNQVDFKPLLEEDTKVTSCLTQEEIDELFNPIYYTKR 420

Query: 421 VDDIFERLGL 430
           VDDIF+RLG+
Sbjct: 421 VDDIFKRLGI 430
```

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

Example 280

A DNA sequence (GBSx0306) was identified in *S.agalactiae* <SEQ ID 899> which encodes the amino acid sequence <SEQ ID 900>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-16.24   Transmembrane  145 - 161 ( 119 - 167)
   INTEGRAL    Likelihood = -9.98   Transmembrane  125 - 141 ( 119 - 144)
   INTEGRAL    Likelihood = -9.29   Transmembrane   28 -  44 (  23 -  51)
   INTEGRAL    Likelihood = -7.01   Transmembrane  196 - 212 ( 193 - 220)
   INTEGRAL    Likelihood = -6.21   Transmembrane   96 - 112 (  88 - 116)
   INTEGRAL    Likelihood = -5.79   Transmembrane  249 - 265 ( 246 - 266)
   INTEGRAL    Likelihood = -2.87   Transmembrane  222 - 238 ( 222 - 238)
   INTEGRAL    Likelihood = -2.28   Transmembrane  279 - 295 ( 278 - 295)

----- Final Results -----
              bacterial membrane --- Certainty=0.7496(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:BAB13498 GB:AB028634 RNA polymerase [Flammulina velutipes]
 Identities = 83/336 (24%), Positives = 150/336 (43%), Gaps = 40/336 (11%)

Query: 152 ILLLIAFVSIGKNR-VYNFVQNLNYFEEVIWNYFEENPVKIKEKSLIIK-----FLLTIS 205
           IL L    SI  NR +  ++ N     ++  N+F+ + +K   K L+I      F++ +S
Sbjct: 133 ILFLYLIYSILINRFILKWLDNSGIIYKININWFKNHMIKHINKMLVINIKFFNFIIKLS 192

Query: 206 FVFVIDFAMVRL-----LNFNIKFSTILACSAILLAWLYQN-----------KSVTEPFL 249
            + +I  +++ L     +NF+I+   I     I    ++               S+   F
Sbjct: 193 IITIIGISIMELFGIFGINFDIRIIIINYLKTINSGKIHLTTINMDQYSVLENSIHTIFY 252

Query: 250 LKKLVIYFIFFIATLIGNLKN-ELSILETPLLFISIFFTMDRIIALSKEMRDLI--ISKS 306
           +  L+I+ IF    L  N+KN + +I    +L+I IF      I     ++DL+  ++K
Sbjct: 253 INLLIIFLIFISLILYRNVKNIDTNIKRWIILYILIPLINIIFIFNHIYIKDLMDNLNKY 312

Query: 307 ILFYYDHENIKPSILLSEIKEIKYLENVDIGE---LELVRQMVIRLRLELEEEFLILSDI 363
           IL Y D   I  S+ L    ++K L+ ++I +      V+ + I+ ++E     L +  I
Sbjct: 313 ILDYMDLHIIVNSLFLFNKFDVK-LKRINIYKSYSTVTVKDLEIKSKIEERSNELDIKLI 371

Query: 364 YMKNG-YEKYIQFVQGNVYFINLE--LDKIPNYTNLKLILESIFD----HNNQKIFIPKL 416
             K G YE YI  ++ N+  ++ E  L   P Y N    +E + +      +   F+ K+
Sbjct: 372 IAKYGSYENYINSIE-NINIVDEEFILKNYPEYINDSKFIEFLMELEPLFRDHTEFVKKI 430

Query: 417 YEEYIYILISLGEVEKAKEIL---KEVSDYLTEESL 449
           YE        L  +   K+IL   KE+ DY+ + +L
Sbjct: 431 YENLNSTNEKLEFLLANKDILSENKEIFDYVLQLNL 466
```

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 281

A DNA sequence (GBSx0308) was identified in *S.agalactiae* <SEQ ID 901> which encodes the amino acid sequence <SEQ ID 902>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3307(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 282

A DNA sequence (GBSx0309) was identified in *S.agalactiae* <SEQ ID 903> which encodes the amino acid sequence <SEQ ID 904>. This protein is predicted to be purK (purK). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0334(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 9461> which encodes amino acid sequence <SEQ ID 9462> was also identified.

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

```
>GP:CAA04376 GB:AJ000883 purK [Lactococcus lactis]
 Identities = 208/347 (59%), Positives = 258/347 (73%), Gaps = 3/347 (0%)

Query: 14  NSFKTIGIIGGGQLGQMMAIAAIYMGHKVITLDPASDCPASRVS-EVIVAPYDDVEALGT 72
           N+ +TIGIIGGGQLGQMMAIAA YMGHKVITLDP  +C A++VS E+IVAPYDDVE L
Sbjct: 4   NTKQTIGIIGGGQLGQMMAIAAQYMGHKVITLDPNPNCSAAKVSDELIVAPYDDVENLLR 63

Query: 73  LAARCDVLTYEFENVDADGLDAVVSAGQLPQGTDLLRISQNRIFEKDFLANKAGVTVAPY 132
           LA  CDV+TYEFENV A  L  +    ++PQG  LL I+QNR FEK+FL N+A V VAP+
Sbjct: 64  LAYACDVITYEFENVSAKALHEIEGCVRIPQGIRLLEITQNRRFEKEFLTNEAKVNVAPW 123

Query: 133 KVVTSSLDLEGLDLTKTYVLKTATGGYDGHGQKVIRSAEDLPEAQQLANSAQCVLEEFVN 192
           ++V S+  L    +T+  VLKT TGGYDGHGQ V+ + E L  A+ L   ++CVLE+F++
Sbjct: 124 QLVDSAEKLPET-VTRKQVLKTTTGGYDGHGQVVLNTDEKLSAAKSLTELSECVLEDFIS 182

Query: 193 FDLEISVIVSGNGQDVTVFPVQENIHRNNILSKTIVPARISDQLADKAKEMAVQIAKKLQ 252
           F+ EISVI+SGNG +  VFP+ EN HR NIL +TI PARIS ++ + A ++A  IA+KL+
Sbjct: 183 FEREISVIISGNGHEYVVFPLAENEHRENILHQTISPARISAEITENAYKIATSIAEKLE 242

Query: 253 LSGTLCVEMFATAD-DIIVNEIAPRPHNSGHYSIEACDFSQFDTHILGVLGAPLPPIKLH 311
           LSG LCVEMF TAD  I VNE+APRPHNSGH++IEACDF+QFD HI G+LG  LP  KL
Sbjct: 243 LSGVLCVEMFLTADGQIYVNELAPRPHNSGHFTIEACDFNQFDLHIKGILGEDLPEPKLL 302

Query: 312 APAVMFNVLGQHVQQAIDHVAQNPSAHLHMYGKLEAKHNRKMGHVTV 358
            PA+M NVLGQHV+        ++   H H YGK +AKHNRKMGHVT+
Sbjct: 303 KPAIMLNVLGQHVEAVKKLNHEHADWHQHDYGKADAKHNRKMGHVTI 349
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 905> which encodes the amino acid sequence <SEQ ID 906>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

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

```
bacterial cytoplasm --- Certainty=0.0334(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 = 344/369 (93%), Positives = 353/369 (95%)

Query: 1   MRNKEKSQRSQAMNSFKTIGIIGGGQLGQMMAIAAIYMGHKVITLDPASDCPASRVSEVI 60
            MRNKEKSQRSQ +NSFKTIGIIGGGQLGQMMAIAAIYMGHKVITLDPASD PASRVSEVI
Sbjct: 1   MRNKEKSQRSQVVNSFKTIGIIGGGQLGQMMAIAAIYMGHKVITLDPASDSPASRVSEVI 60

Query: 61  VAPYDDVEALGTLAARCDVLTYEFENVDADGLDAVVSAGQLPQGTDLLRISQNRIFEKDF 120
            VAPYDDVEALG LAARCDVLTYEFENVDADGLDAVVSA QLPQGTDLLRISQNRI EKDF
Sbjct: 61  VAPYDDVEALGQLAARCDVLTYEFENVDADGLDAVVSACQLPQGTDLLRISQNRIVEKDF 120

Query: 121 LANKAGVTVAPYKVVTSSLDLEGLDLTKTYVLKTATGGYDGHGQKVIRSAEDLPEAQQLA 180
            LANKAGVTVAPYKVVTSSLDL GLDLTKTYVLKT TGGYDGHGQK+IRSAEDLPEAQQLA
Sbjct: 121 LANKAGVTVAPYKVVTSSLDLGGLDLTKTYVLKTETGGYDGHGQKIIRSAEDLPEAQQLA 180

Query: 181 NSAQCVLEEFVNFDLEISVIVSGNGQDVTVFPVQENIHRNNILSKTIVPARISDQLADKA 240
            NSAQCVLEEFVNFDLEISVIVSGNG+DVTVFPVQENIHRNNILSKTIVPARISDQLADKA
Sbjct: 181 NSAQCVLEEFVNFDLEISVIVSGNGKDVTVFPVQENIHRNNILSKTIVPARISDQLADKA 240

Query: 241 KEMAVQIAKKLQLSGTLCVEMFATADDIIVNEIAPRPHNSGHYSIEACDFSQFDTHILGV 300
            K+ AVQIAKKLQLSGTLCVEMF TADDIIVNEIAPRPHNSG YSIEACDFSQFDTHILGV
Sbjct: 241 KKTAVQIAKKLQLSGTLCVEMFTTADDIIVNEIAPRPHNSGRYSIEACDFSQFDTHILGV 300

Query: 301 LGAPLPPIKLHAPAVMFNVLGQHVQQAIDHVAQNPSAHLHMYGKLEAKHNRKMGHVTVFS 360
            LGAPLP I+LHAPAVM NVLGQHVQQA D+VA+NPSAHLHMYGKLEAKHNRKMGHVTVF+
Sbjct: 301 LGAPLPQIQLHAPAVMLNVLGQHVQQATDYVAKNPSAHLHMYGKLEAKHNRKMGHVTVFA 360

Query: 361 DVPDEVEEF 369
               DEV+EF
Sbjct: 361 KDADEVKEF 369
```

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

Example 283

A DNA sequence (GBSx0310) was identified in *S.agalactiae* <SEQ ID 907> which encodes the amino acid sequence <SEQ ID 908>. This protein is predicted to be phosphoribosylaminoimidazole carboxylase catalytic subunit (purE). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3572(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:CAB12462 GB:Z99107 phosphoribosylaminoimidazole carboxylase I
           [Bacillus subtilis]
 Identities = 106/162 (65%), Positives = 128/162 (78%)

Query: 33  MQPIISIIMGSKSDWTTMQKTAEVLDNFGIAYEKKVVSAHRTPDLMFKHAEEARGRGIKI 92
            MQP++ IIMGS SDW TM+   ++LD   + YEKKVVSAHRTPD MF++AE AR RGIK+
Sbjct: 1   MQPLVGIIMGSTSDWETMKHACDILDELNVPYEKKVVSAHRTPDFMFEYAETARERGIKV 60

Query: 93  IIAGAGGAAHLPGMVAAKTTLPVIGVPVKSRALSGLDSLYSIVQMPGGVPVATMAIGEAG 152
            IIAGAGGAAHLPGM AAKTTLPVIGVPV+S+AL+G+DSL SIVQMPGGVPVAT +IG+AG
```

```
Sbjct: 61  IIAGAGGAAHLPGMTAAKTTLPVIGVPVQSKALNGMDSLLSIVQMPGGVPVATTSIGKAG 120

Query: 153 ATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESSNELI 194
           A NA L A +ILS  D++LA  L    E   +   ESS++L+
Sbjct: 121 AVNAGLLAAQILSAFDEDLARKLDERRENTKQTVLESSDQLV 162
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 909> which encodes the amino acid sequence <SEQ ID 910>. Analysis of this protein sequence reveals the following:

```
     Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.08   Transmembrane   36 -  52 (  34 -  52)

----- Final Results -----
               bacterial membrane --- Certainty=0.2232(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:CAA04375 GB:AJ000883 purE [Lactococcus lactis]
 Identities = 105/158 (66%), Positives = 131/158 (82%)

Query: 46  ISIIMGSKSDWATMQKTAEVLDNFGIAYEKKVVSAHRTPDLMFKHAEEARGRGIKIIIAG 105
           ++IIMG  SDWATM++TA++LD+FG+AYEKKVVSAHRTP LM + + +AR RG K+IIAG
Sbjct: 4   VAIIMGCSSDWATMKETAKILDDFGLAYEKKVVSAHRTPALMAEFSSQARERGYKVIIAG 63

Query: 106 AGGAAHLPGMVAAKTTLPVIGVPVKSRALSGLDSLYSIVQMPGGVPVATMAIGEAGATNA 165
           AGGAAHLPGMV+A+T +PVIGVP+KSRALSGLDSLYSIVQMP GVPVATMAIGEAGA NA
Sbjct: 64  AGGAAHLPGMVSAQTLVPVIGVPIKSRALSGLDSLYSIVQMPAGVPVATMAIGEAGAKNA 123

Query: 166 ALTALRILSIEDQNLADALAHFHEEQGKIAEESSGELI 203
           AL AL++L+  ++NL   L +     ++ EES+  L+
Sbjct: 124 ALFALQLLANTNENLIQKLLVYRAAAQEMVEESNKALL 161
```

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

```
 Identities = 162/169 (95%), Positives = 164/169 (96%), Gaps = 1/169 (0%)

Query: 27  PLYLNIMQ-PIISIIMGSKSDWTTMQKTAEVLDNFGIAYEKKVVSAHRTPDLMFKHAEEA 85
           PL + IM+ PIISIIMGSKSDW TMQKTAEVLDNFGIAYEKKVVSAHRTPDLMFKHAEEA
Sbjct: 35  PLCILIMKTPIISIIMGSKSDWATMQKTAEVLDNFGIAYEKKVVSAHRTPDLMFKHAEEA 94

Query: 86  RGRGIKIIIAGAGGAAHLPGMVAAKTTLPVIGVPVKSRALSGLDSLYSIVQMPGGVPVAT 145
           RGRGIKIIIAGAGGAAHLPGMVAAKTTLPVIGVPVKSRALSGLDSLYSIVQMPGGVPVAT
Sbjct: 95  RGRGIKIIIAGAGGAAHLPGMVAAKTTLPVIGVPVKSRALSGLDSLYSIVQMPGGVPVAT 154

Query: 146 MAIGEAGATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESSNELI 194
           MAIGEAGATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESS ELI
Sbjct: 155 MAIGEAGATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESSGELI 203
```

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

Example 284

A DNA sequence (GBSx0311) was identified in *S.agalactiae* <SEQ ID 911> which encodes the amino acid sequence <SEQ ID 912>. This protein is predicted to be phosphoribosylglycinamide synthetase (purD). Analysis of this protein sequence reveals the following:

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

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

```
bacterial cytoplasm --- Certainty=0.1966(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:CAA04374 GB:AJ000883 purD [Lactococcus lactis]
 Identities = 236/419 (56%), Positives = 298/419 (70%), Gaps = 7/419 (1%)

Query: 1   MKLLVVGSGGREHAIAKKLLASKDVDQVFVAPGNDGMTLDGLDLVNIGISEHSRLIDFVK 60
           MK+LV+GSGGREHA+AKK + S  V++VFVAPGN GM  DG+ +V+I    + +L+ F +
Sbjct: 1   MKILVIGSGGREHALAKKFMESPQVEEVFVAPGNSGMEKDGIQIVHISELSNDKLVKFAQ 60

Query: 61  ENEIAWTLIGPDDALAAGIVDGFNSAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA 120
              I  T +GP+ AL  G+VD F  A L  FGP K AAELE SKDFAK IM KY VPTA
Sbjct: 61  NQNIGLTFVGPETALMNGVVDAFIKAELPIFGPNKMAAELEGSKDFAKSIMKKYGVPTAD 120

Query: 121 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 180
           Y TF   E A AY++E+G P+V+KADGLA GKGV VA  +E A  A  ++     F  S
Sbjct: 121 YATFDSLEPALAYLDEKGVPLVIKADGLAAGKGVTVAFDIETAKSALADI-----FSGSQ 175

Query: 181 ARVVIEEFLDGEEFSLFAFANGDKFYIMPTAQDHKRAYDGDKGLNTGGMGAYAPVPHLPQ 240
            +VVIEEFLDGEEFSLF+F +  K Y MP AQDHKRA+D DKG NTGGMGAY+PV H+ +
Sbjct: 176 GKVVIEEFLDGEEFSLFSFIHDGKIYPMPIAQDHKRAFDEDKGPNTGGMGAYSPVLHISK 235

Query: 241 SVVDTAVETIVKPVLEGMIAEGRPYLGVLYAGLILTADGPKVIEFNSRFGDPETQIILPR 300
           VV+ A+E +VKP + GMI EG+ + GVLYAGLILT DG K IEFN+RFGDPETQ++LPR
Sbjct: 236 EVVNEALEKVVKPTVAGMIEEGKSFTGVLYAGLILTEDGVKTIEFNARFGDPETQVVLPR 295

Query: 301 LTSDFAQNIDDIMMGIEPYITWQKDGVTLGVVVASEGYPLDYEKGVPLPEKTDGDIITYY 360
           L SD AQ I DI+ G EP + W + GVTLGVVVA+EGYP   + G+ LPE  +G +  YY
Sbjct: 296 LKSDLAQAIIDILAGNEPTLEWLESGVTLGVVVAAEGYPSQAKLGLILPEIPEG-LNVYY 354

Query: 361 AGAKFAENSKALLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQQDTTGLFYRNDIGSKAI 419
           AG    EN++ L+S+GGRVY++  T + VK+ Q  +Y +L + +  G FYR+DIGS+AI
Sbjct: 355 AGVSKNENNQ-LISSGGRVYLVSETGEDVKSTQKLLYEKLDKLENDGFFYRHDIGSRAI 412
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 913> which encodes the amino acid sequence <SEQ ID 914>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.80   Transmembrane   5 -  21 (  5 -  21)

----- Final Results -----
              bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAA04374 GB:AJ000883 purD [Lactococcus lactis]
 Identities = 236/419 (56%), Positives = 301/419 (71%), Gaps = 7/419 (1%)

Query: 50  LKLLVVGSGGREHAIAKKLLASKGVDQVFVAPGNDGMTLDGLDLVNIVVSEHSRLIAFAK 109
           +K+LV+GSGGREHA+AKK + S  V++VFVAPGN GM  DG+ +V+I    + +L+ FA+
Sbjct: 1   MKILVIGSGGREHALAKKFMESPQVEEVFVAPGNSGMEKDGIQIVHISELSNDKLVKFAQ 60

Query: 110 ENEISWAFIGPDDALAAGIVDDFNSAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA 169
              I   F+GP+ AL  G+VD F  A L  FGP K AAELE SKDFAK IM KY VPTA
Sbjct: 61  NQNIGLTFVGPETALMNGVVDAFIKAELPIFGPNKMAAELEGSKDFAKSIMKKYGVPTAD 120

Query: 170 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 229
           Y TF   E A AY++E+G P+V+KADGLA GKGV VA  +E A  A  ++     F  S
Sbjct: 121 YATFDSLEPALAYLDEKGVPLVIKADGLAAGKGVTVAFDIETAKSALADI-----FSGSQ 175

Query: 230 ARVVIEEFLDGEEFSLFAFANGDKFYIMPTAQDHKRAFDGDKGPNTGGMGAYAPVPHLPQ 289
```

```
              +VVIEEFLDGEEFSLF+F +  K Y MP AQDHKRAFD DKGPNTGGMGAY+PV H+ +
Sbjct: 176 GKVVIEEFLDGEEFSLFSFIHDGKIYPMPIAQDHKRAFDEDKGPNTGGMGAYSPVLHISK 235

Query: 290 SVVDTAVEMIVRPVLEGMVAEGRPYLGVLYVGLILTADGPKVIEFNSRFGDPETQIILPR 349
            VV+ A+E +V+P + GM+ EG+ + GVLY GLILT DG K IEFN+RFGDPETQ++LPR
Sbjct: 236 EVVNEALEKVVKPTVAGMIEEGKSFTGVLYAGLILTEDGVKTIEFNARFGDPETQVVLPR 295

Query: 350 LTSDFAQNIDDIMMGIEPYITWQKDGVTLGVVVASEGYPFDYEKGVPLPEKTDGDIITYY 409
            L SD AQ I DI+ G EP + W + GVTLGVVVA+EGYP   + G+ LPE  +G +  YY
Sbjct: 296 LKSDLAQAIIDILAGNEPTLEWLESGVTLGVVVAAEGYPSQAKLGLILPEIPEG-LNVYY 354

Query: 410 AGVKFSENSELLLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQQDTTGLFYRNDIGSKAI 468
            AGV  +EN++ L+S+GGRVY++  T + VK+ Q  +Y +L + +  G FYR+DIGS+AI
Sbjct: 355 AGVSKNENNQ-LISSGGRVYLVSETGEDVKSTQKLLYEKLDKLENDGFFYRHDIGSRAI 412
```

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

```
 Identities = 399/421 (94%), Positives = 408/421 (96%)

Query: 1   MKLLVVGSGGREHAIAKKLLASKDVDQVFVAPGNDGMTLDGLDLVNIGISEHSRLIDFVK 60
            +KLLVVGSGGREHAIAKKLLASK VDQVFVAPGNDGMTLDGLDLVNI +SEHSRLI F K
Sbjct: 50  LKLLVVGSGGREHAIAKKLLASKGVDQVFVAPGNDGMTLDGLDLVNIVVSEHSRLIAFAK 109

Query: 61  ENEIAWTLIGPDDALAAGIVDGFNSAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA 120
            ENEI+W  IGPDDALAAGIVD FNSAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA
Sbjct: 110 ENEISWAFIGPDDALAAGIVDDFNSAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA 169

Query: 121 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 180
            YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG
Sbjct: 170 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 229

Query: 181 ARVVIEEFLDGEEFSLFAFANGDKFYIMPTAQDHKRAYDGDKGLNTGGMGAYAPVPHLPQ 240
            ARVVIEEFLDGEEFSLFAFANGDKFYIMPTAQDHKRA+DGDKG NTGGMGAYAPVPHLPQ
Sbjct: 230 ARVVIEEFLDGEEFSLFAFANGDKFYIMPTAQDHKRAFDGDKGPNTGGMGAYAPVPHLPQ 289

Query: 241 SVVDTAVETIVKPVLEGMIAEGRPYLGVLYAGLILTADGPKVIEFNSRFGDPETQIILPR 300
            SVVDTAVE IV+PVLEGM+AEGRPYLGVLY GLILTADGPKVIEFNSRFGDPETQIILPR
Sbjct: 290 SVVDTAVEMIVRPVLEGMVAEGRPYLGVLYVGLILTADGPKVIEFNSRFGDPETQIILPR 349

Query: 301 LTSDFAQNIDDIMMGIEPYITWQKDGVTLGVVVASEGYPLDYEKGVPLPEKTDGDIITYY 360
            LTSDFAQNIDDIMMGIEPYITWQKDGVTLGVVVASEGYP DYEKGVPLPEKTDGDIITYY
Sbjct: 350 LTSDFAQNIDDIMMGIEPYITWQKDGVTLGVVVASEGYPFDYEKGVPLPEKTDGDIITYY 409

Query: 361 AGAKFAENSKALLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQQDTTGLFYRNDIGSKAIKE 421
            AG KF+ENS+ LLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQQDTTGLFYRNDIGSKAI+E
Sbjct: 410 AGVKFSENSELLLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQQDTTGLFYRNDIGSKAIRE 470
```

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

Example 285

A DNA sequence (GBSx0312) was identified in *S.agalactiae* <SEQ ID 915> which encodes the amino acid sequence <SEQ ID 916>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.28   Transmembrane  235 - 251 ( 235 - 251)

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

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

```
>GP:AAA23257 GB:M81878 unknown [Clostridium perfringens]
 Identities = 66/258 (25%), Positives = 119/258 (45%), Gaps = 9/258 (3%)

Query: 1   MTIYDQIESALDLMTDLEREIACYFMGQPISKDALASTIVTKQLHISQAALTRFAKKCGF 60
           M I +Q+E+     T  E+ +  Y      +    + +I+ K+  + +A +TRF KK GF
Sbjct: 1   MGILEQLENPKFKATKSEKTLIEYIKSDLDNIIYKSISIIAKESGVGEATITRFTKKLGF 60

Query: 61  KGYREFVFEYLKS-HETISQQLYGLQNDNTKKVFMNYQEMISKSADI-------IDEEQL 112
            G+++F      K      +  +  L     + V     +M+  S +I        ID + +
Sbjct: 61  NGFQDFKVTLAKEISNKKNTSIINLHVHRDESVTETANKMLKSSINILEQTVKQIDLDLM 120

Query: 113 LEVSHMIEQADRVYFYGKGSSSLVAKEFKIRLMRLGVICEALDDTDSFSWTNSIVNDRCL 172
            +   +I  A RVYF G G S + A +    + MR+G      + D+ +     +SI ND  +
Sbjct: 121 CKCRDLIMNAKRVYFIGIGYSGIAATDINYKFMRIGFTTVPVTDSHTMVIMSSITNDDDV 180

Query: 173 VIAFSLSGNTNSVIGALKIASCHGAKTVLFTK-QPHTIDYAFDKIIQVASARHLDYGNRI 231
           ++A S SG T  VI  +K A  +G K +  T+   + +    D  +   SA  +     I
Sbjct: 181 IVAISNSGTTKEVIKTVKQAKENGTKIITLTEDSDNPLRKLSDYELTYTSAETIFETGSI 240

Query: 232 SPQIPMLIMVDIIYAQFL 249
           S +IP + ++D++Y + +
Sbjct: 241 SSKIPQIFLLDLLYTEVI 258
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 917> which encodes the amino acid sequence <SEQ ID 918>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood = -4.88   Transmembrane  243 - 259 ( 242 - 261)

----- Final Results -----
               bacterial membrane --- Certainty=0.2954(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 9093> which encodes the amino acid sequence <SEQ ID 9094>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 56
>>> Seems to have no N-terminal signal sequence
     INTEGRAL     Likelihood = -4.88   Transmembrane  239 - 255 ( 238 - 257)

----- Final Results -----
               bacterial membrane --- Certainty= 0.295(Affirmative) < succ>
                bacterial outside --- Certainty= 0.000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

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

```
 Identities = 138/263 (52%), Positives = 189/263 (71%), Gaps = 2/263 (0%)

Query: 6   QIESALDLMTDLEREIACYFMGQPISKDALASTIVTKQLHISQAALTRFAKKCGFKGYRE 65
           +IE++L+ MT LE+ IA +F+   ++    L ++ + K+LHISQAALTRFAKKCGF GYR
Sbjct: 14  KIEASLEHMTSLEKGIAHFFITTDLTPQELTASEIVKRLHISQAALTRFAKKCGFTGYRA 73

Query: 66  FVFEYLKSHETISQQLYGLQNDNTKKVFMNYQEMISKSADIIDEEQLLEVSHMIEQADRV 125
           F F+YL S +    +    + TK+V M+Y  +I+K+ ++++EE+LL ++ +I+ ++RV
Sbjct: 74  FAFDYLHSLQESQETFQSIHLELTKRVLMDYDALINKTYELVNEEKLLNLAKLIDSSERV 133

Query: 126 YFYGKGSSSLVAKEFKIRLMRLGVICEALDDTDSFSWTNSIVNDRCLVIAFSLSGNTNSV 185
           YF+GKGSS LVA+E K+R MRLG+IC+A  DTD F+W NS+VN+ CLV  FSLSG TNSV
Sbjct: 134 YFFGKGSSGLVAREMKLRFMRLGLICDAYSDTDGFTWANSLVNENCLVFGFSLSGKTNSV 193

Query: 186 IGALKIASCHGAKTVLFTKQPHT-IDYAFDKIIQVASARHLDYGNRISPQIPMLIMVDII 244
           I AL  AS  GAKTVL T    T  D + D II V+S   L YGNR+SPQ P+LIM+DII
Sbjct: 194 ITALHQASQRGAKTVLLTTDNQTEFDDSLD-IIPVSSTHQLHYGNRVSPQFPLLIMMDII 252
```

```
Query: 245 YAQFLDINKIEKERIFRETIIQR 267
          YA  L I+K  KE+IF+ TII +
Sbjct: 253 YAYVLAIDKPHKEKIFKNTIIDK 275
```

SEQ ID 916 (GBS320) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 40 (lane 5; MW 33kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 7; MW 58kDa) and in Figure 160 (lane 7 & 8; MW 58kDa).

GBS320-GST was purified as shown in Figure 224, lane 3-4.

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

Example 286

A DNA sequence (GBSx0313) was identified in *S.agalactiae* <SEQ ID 919> which encodes the amino acid sequence <SEQ ID 920>. This protein is predicted to be xylan esterase 1 (cephalosporin-C). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4981(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:AAB68821 GB:AF001926 xylan esterase 1 [Thermoanaerobacterium sp.
          'JW/SL YS485']
 Identities = 133/299 (44%), Positives = 188/299 (62%), Gaps = 1/299 (0%)

Query: 5   MSLDDMREYLGQDQIPEDFDDFWKKQTMKYQG-NIEYRLDKKDFNITFAQAYDLHFKGSN 63
           M L  +REY G +  PEDFD++W +   + +  + +  L +  F ++FA+ YDL+F G
Sbjct: 6   MPLQKLREYTGTNPCPEDFDEYWNRALDEMRSVDPKIELKESSFQVSFAECYDLYFTGVR 65

Query: 64  NSIVYAKCLFPKTNKPYPVVFYFHGYQNQSPDWSDQLNYVAAGYGVVSMDVRGQAGQSQD 123
            + ++AK + PKT   +P +  FHGY + S DW+D+LNYVAAG+ VV+MDVRGQ GQSQD
Sbjct: 66  GARIHAKYIKPKTEGKHPALIRFHGYSSNSGDWNDKLNYVAAGFTVVAMDVRGQGGQSQD 125

Query: 124 KGHFDGITVKGQIVRGMISGPNHLFYKDIYLDVFQLIDIIATLESVDSNQLYSYGWSQGG 183
            G   G T+ G I+RG+     +++ ++ I+LD  QL  I+  +  VD +++   G SQGG
Sbjct: 126 VGGVTGNTLNGHIIRGLDDDADNMLFRHIFLDTAQLAGIVMNMPEVDEDRVGVMGPSQGG 185

Query: 184 ALALIAAALNPKIVKTVAVYPFLSDFRRVLDLGGVSEPYDELFRYFKYSDPFHKTENNVL 243
            L+L  AAL P++ K V+ YPFLSD++RV DL      Y E+  YF+  DP H+ EN V
Sbjct: 186 GLSLACAALEPRVRKVVSEYPFLSDYKRVWDLDLAKNAYQEITDYFRLFDPRHERENEVF 245

Query: 244 KTLAYIDVKNFAHRISCPVVLLTALKDDICPPSTQFAIFNRLTSTKKHLLLPDYGHDPM 302
             L YIDVKN A RI    V++    L D +CPPST FA +N + S K    + PDYGH+PM
Sbjct: 246 TKLGYIDVKNLAKRIKGDVLMCVGLMDQVCPPSTVFAAYNNIQSKKDIKVYPDYGHEPM 304
```

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 287

A DNA sequence (GBSx0314) was identified in *S.agalactiae* <SEQ ID 921> which encodes the amino acid sequence <SEQ ID 922>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.73   Transmembrane  128 - 144 ( 126 - 145)

----- Final Results -----
               bacterial membrane --- Certainty=0.3293(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:AAA23256 GB:M81878 unknown [Clostridium perfringens]
 Identities = 78/160 (48%), Positives = 110/160 (68%)

Query: 131 CLTIGTGIGGCLIIDKTVFHGFSNSACEVGYMHLSDGDFQDLASTTALIADVAKAHGDEI 190
           CLTIGTGIGG LIID  V HGFSNSA E+GYM ++  + QD+AS +AL+ +VA   G E
Sbjct: 18  CLTIGTGIGGALIIDGKVLHGFSNSAGEIGYMMVNGENIQDIASASALVKNVALRKGVEP 77

Query: 191 SRWDGRRIFQEAKKGNEKCIASIDRMINYLGQGIANMVYVVNPEKVVLGGGIMAQKDYLQ 250
           S  DGR +    + G+  C   ++++ + L  GI+N+VY++NPE VVLGGGIMA+++  +
Sbjct: 78  SSIDGRYVLDNYENGDLICKEEVEKLADNLALGISNIVYLINPEVVVLGGGIMAREEVFR 137

Query: 251 DKLSESLKRNLVTSLAEKTAIVFAQHENQAGMLGAYYHFK 290
             +  SL++ L+ S+   T I FA+ +N AGM GAYY+FK
Sbjct: 138 PLIENSLRKYLIESVYNNTKIAFAKLKNTAGMKGAYYNFK 177
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 923> which encodes the amino acid sequence <SEQ ID 924>. Analysis of this protein sequence reveals the following:

```
     Possible site: 22

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.30   Transmembrane  128 - 144 ( 127 - 145)
   INTEGRAL    Likelihood = -0.11   Transmembrane  227 - 243 ( 227 - 243)

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

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

```
>GP:BAB04516 GB:AP001509 glucose kinase [Bacillus halodurans]
 Identities = 97/291 (33%), Positives = 155/291 (52%), Gaps = 14/291 (4%)

Query: 5   LAIDIGGTAIKYGLISETGDLLEKEEMATEAYKGGPSILEKVKGLVKTYQDQMDLAGVAI 64
           + ID+GGT IK  L+S+ G+++  +E  TEA +G   ++ K+  L +   D    AG+ I
Sbjct: 3   VGIDLGGTKIKAALVSDAGEIISVQECPTEAAQGPEEVMNKMMSLTEKVTDHQPFAGIGI 62

Query: 65  SSAGMVNPDEGEIFYAGPQIPNYAGTQFKKEIEETFGLPCEVENDVNCAGLAEAISGSAK 124
            + G ++  EG I  + P +P +          +E F  P +++ND N A LAEA+ GS +
Sbjct: 63  GAPGPLSSTEGTIL-SPPNLPGWDHIHLVDRFQEQFQCPVKLDNDANVAALAEALLGSGQ 121

Query: 125 DYPVALCLTIGTGIGGCLLFNSQVFHGSSHSACEVG----------YLHLSDGQFQDLAS 174
            +     LTI TGIGG  + +  + HG+S  A E+G          + +L+ G  + LAS
Sbjct: 122 GFTSVFYLTISTGIGGGYVLDGSIVHGASDYAGEIGNMIVQPNGYQHANLNPGSLEGLAS 181

Query: 175 TTALVQEVVLAYGDDISQWDGRRIFEQAKAGDAICIAAISKQVDYLGQGIANICYVVNPN 234
            TA+ +     +G    +   R +F+Q + GD       + + +DYL  GIANI + +NP+
Sbjct: 182 GTAIGRMARERFG---VEGGTREVFDQIRRGDHDMQRLVEEAMDYLAIGIANIAHTINPD 238
```

```
Query: 235 VVVLGGGIMAQKDYLADKLKTALDSYLVSSLAKKTQLKFASHGNNAGILGA 285
           V VLGGG+M   D +   +K  +  YL   LA+ T +  A  G ++G+LGA
Sbjct: 239 VFVLGGGVMNADDLILPIVKEKVSRYLYPGLAQSTTIVKAKLGGDSGVLGA 289
```

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

```
 Identities = 192/292 (65%), Positives = 237/292 (80%)

Query: 1   MTRTVAIDIGGTMIKHGIVDNLGCIVEASELATEAYKGGPGILQKVCQIIDNYLAEGSID 60
           M   +AIDIGGT IK+G++   G ++E  E+ATEAYKGGP IL+KV  ++  Y  +  +
Sbjct: 1   MKHYLAIDIGGTAIKYGLISETGDLLEKEEMATEAYKGGPSILEKVKGLVKTYQDQMDLA 60

Query: 61  GIAISSAGMVDPDEGCIFYSGPQIPNYAGTQFKKVLEDTYQVRTEIENDVNCAGLAEAVS 120
           G+AISSAGMV+PDEG IFY+GPQIPNYAGTQFKK +E+T+ +  E+ENDVNCAGLAEA+S
Sbjct: 61  GVAISSAGMVNPDEGEIFYAGPQIPNYAGTQFKKEIEETFGLPCEVENDVNCAGLAEAIS 120

Query: 121 GSAKDSSIALCLTIGTGIGGCLIIDKTVFHGFSNSACEVGYMHLSDGDFQDLASTTALIA 180
           GSAKD  +ALCLTIGTGIGGCL+ +  VFHG S+SACEVGY+HLSDG FQDLASTTAL+
Sbjct: 121 GSAKDYPVALCLTIGTGIGGCLLFNSQVFHGSSHSACEVGYLHLSDGQFQDLASTTALVQ 180

Query: 181 DVAKAHGDEISRWDGRRIFQEAKKGNEKCIASIDRMINYLGQGIANMVYVVNPEKVVLGG 240
           +V  A+GD+IS+WDGRRIF++AK G+  CIA+I + ++YLGQGIAN+ YVVNP  VVLGG
Sbjct: 181 EVVLAYGDDISQWDGRRIFEQAKAGDAICIAAISKQVDYLGQGIANICYVVNPNVVVLGG 240

Query: 241 GIMAQKDYLQDKLSESLKRNLVTSLAEKTAIVFAQHENQAGMLGAYYHFKNR 292
           GIMAQKDYL DKL  +L   LV+SLA+KT + FA H N AG+LGAYYHFK +
Sbjct: 241 GIMAQKDYLADKLKTALDSYLVSSLAKKTQLKFASHGNNAGILGAYYHFKQK 292
```

SEQ ID 922 (GBS331) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 2; MW 35.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 3; MW 61kDa).

The GBS331-GST fusion product was purified (Figure 209, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 309), 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 288

A DNA sequence (GBSx0315) was identified in *S.agalactiae* <SEQ ID 925> which encodes the amino acid sequence <SEQ ID 926>. This protein is predicted to be a acylneuraminate lyase (nanA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0894(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:CAA69950 GB:Y08695 putative acylneuraminate lyase [Clostridium
          tertium]
 Identities = 162/225 (72%), Positives = 191/225 (84%)

Query: 1   MKDLQKYQGIIPAFYACYDDKGDICPERVKALTNYFIDKGVQGLYVNGSSGECIYQSVAD 60
           M++L+KY+GIIPAFYACYDD+G I PER +  T Y IDKGV+GLYV GSSGECIYQS  +
Sbjct: 1   MRNLEKYKGIIPAFYACYDDEGKISPERTQMFTQYLIDKGVKGLYVCGSSGECIYQSKEE 60
```

```
Query: 61  RKLVLENVMSVAKGKLTVIAHVACNNTKDSVELAMHAEAIGVDAIAAIPPIYFRLPEYAI 120
           RK+ LENVM VAKGK+T+IAHV CNNT+DS ELA HAE+IGVDAIA+IPPIYF LP+Y+I
Sbjct: 61  RKITLENVMKVAKGKITIIAHVGCNNTRDSEELAEHAESIGVDAIASIPPIYFHLPDYSI 120

Query: 121 ADYWNTISQAAPQTDFIIYNIPQLAGVALTSDLYRKMLQNPQVIGVKNSSMPVQDIQNFV 180
           A+YWN IS AAP TDFIIYNIPQLAGV L  +LY++ML+NP+VIGVKNSSMPVQDIQ F
Sbjct: 121 AEYWNDISNAAPNTDFIIYNIPQLAGVGLGINLYKQMLKNPRVIGVKNSSMPVQDIQMFK 180

Query: 181 AIGGENHIVFNGPDEQFLGGRLMGAAAGIGGTYGVMPELYLTLNQ 225
            I G+  +VFNGPDEQF+ GR+MGA  GIGGTY VMPEL+L  ++
Sbjct: 181 DISGDESVVFNGPDEQFVAGRIMGADGGIGGTYAVMPELFLAADK 225
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 927> which encodes the amino acid sequence <SEQ ID 928>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0981(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 = 238/304 (78%), Positives = 263/304 (86%)

Query: 1   MKDLQKYQGIIPAFYACYDDKGDICPERVKALTNYFIDKGVQGLYVNGSSGECIYQSVAD 60
           M DL KYQGIIPAFYACYDD+G+I PERV+ALT Y+IDKGVQGLY+NGSSGECIYQSV D
Sbjct: 1   MTDLTKYQGIIPAFYACYDDQGNISPERVRALTQYYIDKGVQGLYINGSSGECIYQSVFD 60

Query: 61  RKLVLENVMSVAKGKLTVIAHVACNNTKDSVELAMHAEAIGVDAIAAIPPIYFRLPEYAI 120
           R+LVLENVM+VAKGKLT+I HVACNNTKDS+ELA H+E +GVDAIAAIPPIYFRLPEYA+
Sbjct: 61  RQLVLENVMAVAKGKLTIINHVACNNTKDSIELAAHSERLGVDAIAAIPPIYFRLPEYAV 120

Query: 121 ADYWNTISQAAPQTDFIIYNIPQLAGVALTSDLYRKMLQNPQVIGVKNSSMPVQDIQNFV 180
           ADYWN IS AAP TDFIIYNIPQLAGVALT  LY+ ML N +VIGVKNSSMPVQDIQ F
Sbjct: 121 ADYWNAISSAAPHTDFIIYNIPQLAGVALTPSLYKTMLANKRVIGVKNSSMPVQDIQTFC 180

Query: 181 AIGGENHIVFNGPDEQFLGGRLMGAAAGIGGTYGVMPELYLTLNQLIVDKDLEKARELQF 240
           AIGG++HIVFNGPDEQFLGGRLMGAAAGIGGTYG MPEL+L LNQLI DKDLEKA+ LQ+
Sbjct: 181 AIGGDDHIVFNGPDEQFLGGRLMGAAAGIGGTYGAMPELFLRLNQLIADKDLEKAKALQY 240

Query: 241 TINDIITKLCSGHGNMYAVIKAVLEINEQLTIGSVRLPLASVTEEDKPIIKEAAEMIRHA 300
           TIN+II  L S HGNMY VIK VL INE L IGSVR PLA + EED+ I + AA +I  A
Sbjct: 241 TINEIIGVLVSAHGNMYGVIKEVLRINEGLDIGSVRSPLAELVEEDRVICQRAAALINQA 300

Query: 301 KKQF 304
           K+ F
Sbjct: 301 KETF 304
```

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

Example 289

A DNA sequence (GBSx0317) was identified in *S.agalactiae* <SEQ ID 929> which encodes the amino acid sequence <SEQ ID 930>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.45   Transmembrane   82 -  98 (  79 - 111)
   INTEGRAL    Likelihood = -6.85   Transmembrane   24 -  40 (  21 -  52)
   INTEGRAL    Likelihood = -5.26   Transmembrane  180 - 196 ( 172 - 200)
```

```
    INTEGRAL    Likelihood = -5.10   Transmembrane  160 - 176 ( 158 - 179)
    INTEGRAL    Likelihood = -4.35   Transmembrane  110 - 126 ( 106 - 130)

----- Final Results -----
               bacterial membrane --- Certainty=0.4779(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:BAB05827 GB:AP001514 unknown conserved protein in B. subtilis
          [Bacillus halodurans]
 Identities = 40/148 (27%), Positives = 74/148 (49%), Gaps = 4/148 (2%)

Query: 14  VNNPFMQGCNVVFDLALLNLLFMI-TCLPLVTIG--AAKISLYRTLWQKLEGD-QTNLLI 69
           +++ F Q C+ ++ LA +NLL++  T L LV +G   A  +++  L +   G+    +
Sbjct: 6   MSSRFYQTCDWIWKLAYINLLWLSGTLLGLVVLGFLPATTAMFTVLRKWFTGNPDVAITR 65

Query: 70  LYIKHLKKEWFQGMLLGLVELSILVVIIFDLTILHYQIGFIVSFLKITCYAFLLLTVMTS 129
            + +  K E+ +  LLG V L    ++ F+   L    G +  L +  YAFL+L ++T
Sbjct: 66  TFFQAYKNEFLKINLLGAVLLLGAYILYFNYMYLGTVEGTVHMVLSLGWYAFLILYIITL 125

Query: 130 IYLFPMAARYEMSLLDTVKKSFIMACLN 157
            Y+ P    Y + L   +K + I+  +N
Sbjct: 126 FYIIPAYVHYNLKLFQYIKTALIIGFVN 153
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 931> which encodes the amino acid sequence <SEQ ID 932>. Analysis of this protein sequence reveals the following:

```
     Possible site: 24
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.86   Transmembrane  117 - 133 ( 108 - 139)
   INTEGRAL    Likelihood = -7.48   Transmembrane   30 -  46 (  21 -  54)
   INTEGRAL    Likelihood = -6.90   Transmembrane   88 - 104 (  83 - 105)
   INTEGRAL    Likelihood = -6.26   Transmembrane  165 - 181 ( 151 - 187)
   INTEGRAL    Likelihood = -5.89   Transmembrane  189 - 205 ( 182 - 207)

----- Final Results -----
               bacterial membrane --- Certainty=0.6944(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:BAB05582 GB:AP001513 unknown conserved protein in bacilli
          [Bacillus halodurans]
 Identities = 59/194 (30%), Positives = 93/194 (47%), Gaps = 11/194 (5%)

Query: 17  SKWMRASAALFDLLVFNLLFVL-SCLPLLTIGV--AKMALYASLLDWREGQVS-QLVTTY 72
           +K M+    +  L+  NLL++L S +  + +GV  A  +L+A    W + +    L  TY
Sbjct: 8   TKIMKLFEWIMRLVYLNLLWLLFSFIGGIILGVMPATASLFAVFRKWYQKEDDFPLFQTY 67

Query: 73  SSHFKYYFKSGLRLGLIELGIMTICLLDLFLIRNQSGLVFQGFKVLCVAVLFLVVILFLY 132
            + FK  FK    +GL  + I  I  LD+ L+   S  + Q   +  A+ F+ ++  LY
Sbjct: 68  LNEFKRSFKIANLVGLTLVLIGGILYLDVLLLLGTSHWIGQLLLMGVGALSFIYLVTLLY 127

Query: 133 AYPQAVKRDLSLSTLFKRSFLLAGLFFPWSFAFLAFICLTIFSLQL----SLLTLFGGVS 188
            +P  V  DLS    FK SFLL G+  P+    L  I L++ +L        LL LF   S
Sbjct: 128 IFPTLVHFDLSYKQYFKHSFLL-GVLQPFR-TLLLMITLSLSALLFLTFPILLPLF-AAS 184

Query: 189 LLAIIGISSLTYLY 202
           +A + + S  + Y
Sbjct: 185 FMAALTMWSFLFGY 198
```

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

```
 Identities = 68/210 (32%), Positives = 117/210 (55%)
```

```
Query: 3   KANQLIAAIFDVNNPFMQGCNVVFDLALLNLLFMITCLPLVTIGAAKISLYRTLWQKLEG 62
           K   L+ ++F +++ +M+     +FDL + NLLF+++CLPL+TIG AK++LY +L    EG
Sbjct: 4   KKQGLLHSLFKLDSKWMRASAALFDLLVFNLLFVLSCLPLLTIGVAKMALYASLLDWREG 63

Query: 63  DQTNLLILYIKHLKKEWFQGMLLGLVELSILVVIIFDLTILHYQIGFIVSFLKITCYAFL 122
             + L+  Y  H K  +  G+ LGL+EL I+ + + DL ++  Q G +     K+ C A L
Sbjct: 64  QVSQLVTTYSSHFKYYFKSGLRLGLIELGIMTICLLDLFLIRNQSGLVFQGFKVLCVAVL 123

Query: 123 LLTVMTSIYLFPMAARYEMSLLDTVKKSFIMACLNLKWTGVLMFLLIMTWFIMVQSSLLF 182
            L V+  +Y +P A + ++SL    K+SF++A L   W+   +  + +T F +  S L
Sbjct: 124 FLVVILFLYAYPQAVKRDLSLSTLFKRSFLLAGLFFPWSFAFLAFICLTIFSLQLSLLTL 183

Query: 183 MLTVSAIFIFAYTAFAYFKIIILQKQFAYF 212
              VS + I    ++  Y  +II++     F
Sbjct: 184 FGGVSLLAIIGISSLTYLYLIIMESLLRRF 213
```

A related GBS gene <SEQ ID 8535> and protein <SEQ ID 8536> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 2
McG: Discrim Score:       3.27
GvH: Signal Score (-7.5): -4.23
     Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 5 value:  -9.45 threshold:  0.0
   INTEGRAL    Likelihood = -9.45   Transmembrane   82 -  98 (  79 - 111)
   INTEGRAL    Likelihood = -6.85   Transmembrane   24 -  40 (  21 -  52)
   INTEGRAL    Likelihood = -5.26   Transmembrane  180 - 196 ( 172 - 200)
   INTEGRAL    Likelihood = -5.10   Transmembrane  160 - 176 ( 158 - 179)
   INTEGRAL    Likelihood = -4.35   Transmembrane  110 - 126 ( 106 - 130)
   PERIPHERAL  Likelihood =  5.89     142
 modified ALOM score:   2.39

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4779(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:

```
ORF00072(364 - 828 of 1260)
EGAD|108353|BS3003(14   -   171   of   222)   hypothetical   protein   {Bacillus   subtilis}
OMNI|NT01BS3507 conserved hypothetical protein GP|2635493|emb|CAB14987.1||Z99119 similar to
hypothetical      proteins      from      B.      subtilis      {Bacillus      subtilis}
GP|2293197|gb|AAC00275.1||AF008220 YteU {Bacillus subtilis} PIR|D69991|D69991 conserved
hypothetical protein yteU - Bacillus subtilis
%Match = 5.9
%Identity = 26.6  %Similarity = 50.6
Matches = 42  Mismatches = 74  Conservative Sub.s = 38

270       300       330       360       390       417       441       471
IMSKKGY*KC*WRKKYREYIVKKANQLIAAIFDVNNPFMQGCNVVFDLALLNLLFMI-TCLPLVTIG--AAKISLYRTLW
                                          |  :  :|  |||::  | | |   |   |  :|:  :
                                MEHDGSLGRMLRFCEWIMRFAYTNLLWLFFTLLGLGVFGIMPATAALFAVMR
                                         10        20        30        40        50

498       528       558       588       618       648       678       708
QKLEG-DQTNLLILYIKHLKKEWFQGMLLGLVELSILVVIIFDLTILHYQIGFIVSFLKITCYAFLLLTVMTSIYLFPMA
: ::| |   :|  : :  | |:|:  ||| |   | |:|  || :: |   |::  |:      | :| |    |:||:
KWIQGQDNVPVLKTFWQEYKGEFFRSNLLGAVLALIGVIIYIDLALI-YPSHFLLHILRFAIMIFGFLFVSMLFYVFPLL
                  70        80        90        100       110       120       130

738       768       798       828       858       888       918       948
ARYEMSLLDTVKKSFIMACLNLKWTGVLMFLLIMTWFIMVQSSLLFMLTVSAIFIFAYTAFAYFKIIILQKQFAYFSKQQ
```

```
  ::      || |::::   |::|  :: | :  :|::
VHFDWKKRLYVKFSLLLSVAYLQYTLTMLALTVALFFLLAYLPGIVPFFSVSLISYCHMRIVYAVLLKVEQHGGEPQRKS
                150       160       170       180       190       200       210
```

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

Example 290

A DNA sequence (GBSx0318) was identified in *S.agalactiae* <SEQ ID 933> which encodes the amino acid sequence <SEQ ID 934>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1827(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:AAC44392 GB:U43526 ORF-1 [Streptococcus pneumoniae]
 Identities = 48/151 (31%), Positives = 66/151 (42%), Gaps = 5/151 (3%)

Query: 1   MIYDHLLNLTHYKDINPNLDLAIDYLLSHDLRNLDIGTYHISPEVILMVQSNQLSES-FD 59
           MI   + L  Y  +NP+    ID+L    L NL  G+  I    +        L++
Sbjct: 1   MIITKISRLGTYVGVNPHFATLIDFLEKTGLENLTEGSIAIDGNRLFGNCFTYLADGQAG 60

Query: 60  HIFEYHKKYLDIHYVIEGHEVIKLGKGDKVEV-EEY--LGDIGFIKCSEETSFDLRDNYI 116
             FE H+KYLDIH V+E  E + +   + V V +EY    DI       E    LR
Sbjct: 61  AFFETHQKYLDIHLVLENEEAMAVTSPENVSVTQEYDEEKDIELYTGKVEQLVHLRAGEC 120

Query: 117 AFFFPEEAHQPNGMGSLGNYVKKGVLKVLMA 147
              FPE+ HQP  +      VKK V KV ++
Sbjct: 121 LITFPEDLHQPK-VRINDEPVKKVVFKVAIS 150
```

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 291

A DNA sequence (GBSx0319) was identified in *S.agalactiae* <SEQ ID 935> which encodes the amino acid sequence <SEQ ID 936>. This protein is predicted to be sugar ABC transporter, permease protein (araQ). Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.38   Transmembrane  245 - 261 ( 239 - 265)
   INTEGRAL    Likelihood = -3.72   Transmembrane  140 - 156 ( 139 - 158)
   INTEGRAL    Likelihood = -3.61   Transmembrane   76 -  92 (  71 -  94)
   INTEGRAL    Likelihood = -2.81   Transmembrane  112 - 128 ( 107 - 128)
   INTEGRAL    Likelihood = -1.59   Transmembrane  188 - 204 ( 186 - 204)

----- Final Results -----
              bacterial membrane --- Certainty=0.3951(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:AAD35515 GB:AE001721 sugar ABC transporter, permease protein
          [Thermotoga maritima]
 Identities = 94/262 (35%), Positives = 158/262 (59%), Gaps = 1/262 (0%)

Query: 15  LILCLLTVLFIFPFYWIMTGAFKSQPDTIIIPPQWWPKAPTLENFKALTVQNPALRWLWN 74
           + +  + V+F+ P ++ +  +FK   +    PP  +PK P+LE +  +  +   L +L N
Sbjct: 9   IFIVFMLVVFMLPVFYAVVSSFKPMSEIYSYPPTIFPKKPSLEGYINVIKEYDLLTYLRN 68

Query: 75  SVFISIMTMFLVCCTSSMAGYVLAKKRFYGQKILFSLFIAAMALPKQVVLVPLVRIINFM 134
           ++F++ +    S M GY LAK +F+G + + S+F   M +  QV++VPL  +I  +
Sbjct: 69  TLFVATVATVITVLVSVMTGYGLAKGKFWGIRPVNSMFTMTMFVSAQVIMVPLFVVIRSL 128

Query: 135 GIHDTLWAVILPLVGWPFGVFLMKQFSENIPTELLESAKIDGCGEIRTFINVAFPIVKPG 194
           G+ ++LW +I+P V  P G+F+  Q+ ++IP ELLESAKIDG  E + F  + FP+ KP
Sbjct: 129 GLINSLWGLIIPAVYTPTGMFMAVQYMKDIPDELLESAKIDGANEWQIFWRIVFPLSKPL 188

Query: 195 FAALAIFTFINTWNDYFMQLVMLTSRNNLTISLGVATMQAEM-ATNYGLIMAGAALAAVP 253
            AALAIF+F   WND+ + L+++  RN  T+ L +AT+Q E     + I+A + L  +P
Sbjct: 189 VAALAIFSFTWRWNDFVLPLLVVNRRNLYTLQLALATIQEEYGGAEWNTILAFSTLTIIP 248

Query: 254 IVTVFLVFQKSFTQGITMGAVK 275
           + +FL+FQ+ F +GI  G +K
Sbjct: 249 TLIIFLLFQRLFMKGIMAGGLK 270
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 937> which encodes the amino acid sequence <SEQ ID 938>. Analysis of this protein sequence reveals the following:

```
     Possible site: 40
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.37   Transmembrane  245 - 261 ( 240 - 265)
   INTEGRAL    Likelihood = -5.15   Transmembrane  140 - 156 ( 139 - 158)
   INTEGRAL    Likelihood = -2.97   Transmembrane  111 - 127 ( 107 - 128)
   INTEGRAL    Likelihood = -2.87   Transmembrane   76 -  92 (  75 -  93)
   INTEGRAL    Likelihood = -1.59   Transmembrane  188 - 204 ( 186 - 204)

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

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

```
>GP:CAB59597 GB:AL132662 probable sugar transport inner membrane
          protein [Streptomyces coelicolor A3(2)]
 Identities = 88/262 (33%), Positives = 147/262 (55%)

Query: 15  VMLCVLTILFIFPFYWIMTGAFKAQADTIMIPPQWWPKAPTIENFKALVVQNPALKWLWN 74
           ++L  L ++F  P  W++  +  + A+    PP  WP    +  ++ ++      +W  N
Sbjct: 38  LLLAPLALVFAVPLVWLVLSSVMSNAEINRFPPALWPSGIDLGGYRYVLGNAMFPRWFVN 97

Query: 75  SVFISVATMFLVCGTSSLAGYALAKKRFYGQRLLFSIFIAAMALPKQVVLVPLVRIVNFM 134
           S+ +S  T+       SLAGYA A+ RF G R+L  + +A MA+P Q+ ++P   ++  +
Sbjct: 98  SLIVSAVTVAANLVFGSLAGYAFARMRFAGSRVLMGLMLATMAVPFQLTMIPTFLVMKKL 157

Query: 135 GIHDTLAAVILPLVGWPFGVFLMKQFSENIPTELLESAKIDGCGEIRTFFNVAFPIVKPG 194
           G+ DTL A+I+P +  PF VFL++QF  ++P EL E+A IDGC  +R  + +  P+ +P
Sbjct: 158 GLIDTLGALIVPSLVTPFAVFLLRQFFLSLPRELEEAAWIDGCSRLRVLWRIVLPLSRPA 217

Query: 195 FAALAIFTFINTWNDYFMQLVMLTSRENLTISLGVATMQAEMATNYGLIMAGAAMAAVPI 254
            A +A+ TF+ TWND    L+ +      T+ LG+ T Q +  T +  +MAG  +  +P+
Sbjct: 218 LATVAVLTFLTTWNDLTWPLIAINHDTQYTLQLGLTTFQGQHHTQWAAVMAGNVITVLPV 277

Query: 255 VTVFLVFQKSFTQGITMGAVKG 276
           +  FL  QK+F Q IT   +KG
Sbjct: 278 LLAFLGAQKTFIQSITSSGLKG 299
```

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

```
 Identities = 245/276 (88%), Positives = 262/276 (94%)

Query: 1   MKKKTFSAYNFLTALILCLLTVLFIFPFYWIMTGAFKSQPDTIIIPPQWWPKAPTLENFK 60
           M KK  +A + LT ++LC+LT+LFIFPFYWIMTGAFK+Q DTI+IPPQWWPKAPT+ENFK
Sbjct: 1   MTKKKLTASDILTTVMLCVLTILFIFPFYWIMTGAFKAQADTIMIPPQWWPKAPTIENFK 60

Query: 61  ALTVQNPALRWLWNSVFISIMTMFLVCCTSSMAGYVLAKKRFYGQKILFSLFIAAMALPK 120
           AL VQNPAL+WLWNSVFIS+ TMFLVC TSS+AGY LAKKRFYGQ++LFS+FIAAMALPK
Sbjct: 61  ALVVQNPALKWLWNSVFISVATMFLVCGTSSLAGYALAKKRFYGQRLLFSIFIAAMALPK 120

Query: 121 QVVLVPLVRIINFMGIHDTLWAVILPLVGWPFGVFLMKQFSENIPTELLESAKIDGCGEI 180
           QVVLVPLVRI+NFMGIHDTL AVILPLVGWPFGVFLMKQFSENIPTELLESAKIDGCGEI
Sbjct: 121 QVVLVPLVRIVNFMGIHDTLAAVILPLVGWPFGVFLMKQFSENIPTELLESAKIDGCGEI 180

Query: 181 RTFINVAFPIVKPGFAALAIFTFINTWNDYFMQLVMLTSRNNLTISLGVATMQAEMATNY 240
           RTF NVAFPIVKPGFAALAIFTFINTWNDYFMQLVMLTSR NLTISLGVATMQAEMATNY
Sbjct: 181 RTFFNVAFPIVKPGFAALAIFTFINTWNDYFMQLVMLTSRENLTISLGVATMQAEMATNY 240

Query: 241 GLIMAGAALAAVPIVTVFLVFQKSFTQGITMGAVKG 276
           GLIMAGAA+AAVPIVTVFLVFQKSFTQGITMGAVKG
Sbjct: 241 GLIMAGAAMAAVPIVTVFLVFQKSFTQGITMGAVKG 276
```

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

Example 292

A DNA sequence (GBSx0320) was identified in *S.agalactiae* <SEQ ID 939> which encodes the amino acid sequence <SEQ ID 940>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.83  Transmembrane   74 -  90 (  64 -  96)
   INTEGRAL    Likelihood = -6.37  Transmembrane  108 - 124 ( 107 - 126)
   INTEGRAL    Likelihood = -5.84  Transmembrane  270 - 286 ( 265 - 290)
   INTEGRAL    Likelihood = -5.20  Transmembrane  161 - 177 ( 156 - 182)
   INTEGRAL    Likelihood = -0.16  Transmembrane  219 - 235 ( 219 - 235)

----- Final Results -----
              bacterial membrane --- Certainty=0.5331(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:BAB05584 GB:AP001513 sugar transport system (permease) (binding
          protein dependent transporter) [Bacillus halodurans]
 Identities = 106/289 (36%), Positives = 168/289 (57%), Gaps = 6/289 (2%)

Query: 9   RETMIAYAFLAPILLFFLIFVFAPMVMGFVTSFFNYSM-TQFTFIGLANYNRMF-HDSIF 66
           +E    Y F+AP ++ F IF   PM+     SF ++ +  +  + G  NY R+F  D +F
Sbjct: 25  KEYFWGYLFIAPPIIGFAIFALGPMLYSIYVSFTDFDLYNEPVWTGADNYYRLFVTDDLF 84

Query: 67  MKSLINTVIIVIGSVPVVVFFSLFVAANTYEKNVFSRSFYRCVFFLPVVTGSVAVTVVWK 126
            K++ NT    +G +P+ +  SL +A    +K V   + +R  FFLP V+  VA+T++W+
Sbjct: 85  RKTVFNTFYAALG-IPIGMAVSLGIAVALNQK-VKGIALFRTAFFLPAVSSVVAITLLWR 142

Query: 127 WIYDPMSGILNYILKSGHVIEQNISWLGDKHWALLAIIIILLTTSVGQPIILYIAAMGNI 186
           WI++   G+LN +L   +V      WL D+ WA+ A+II   +   +G  +ILY+AA+  +
Sbjct: 143 WIFNADFGLLNIMLN--YVGIHGPGWLSDEKWAMPAMIIQGVWGGLGINMILYLAALQGV 200

Query: 187 DNSLCEAARVDGANEMQVFWQIKWPSLLPTTLYIAVITTINSFQCFALIQLLTSGGPNYS 246
           + +L EAA +DG N  Q F  I  PS+ PTT +I + +TI + Q F    ++T GGPNYS
Sbjct: 201 NPALYEAADIDGGNAWQKFIHITVPSISPTTFFILITSTIGALQDFQRFMIMTEGGPNYS 260

Query: 247 TSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIALISFAQFKILGNDVEY 295
```

```
           T+T++YYL+  AF+  E GYA+ M   L ++I +I+   FK+    V Y
Sbjct: 261 TTTVVYYLFLNAFRYMEMGYASAMAWVLGIIILIITIINFKLAKKWVHY 309
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 941> which encodes the amino acid sequence <SEQ ID 942>. Analysis of this protein sequence reveals the following:

```
     Possible site: 13
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.74   Transmembrane   55 -  71 (  44 -  78)
   INTEGRAL    Likelihood =-10.83   Transmembrane  109 - 125 (  98 - 130)
   INTEGRAL    Likelihood = -6.21   Transmembrane  304 - 320 ( 299 - 324)
   INTEGRAL    Likelihood = -6.00   Transmembrane  142 - 158 ( 141 - 160)
   INTEGRAL    Likelihood = -5.04   Transmembrane  196 - 212 ( 190 - 216)
   INTEGRAL    Likelihood = -0.16   Transmembrane  253 - 269 ( 253 - 269)

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

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

```
>GP:BAB05584 GB:AP001513 sugar transport system (permease) (binding
          protein dependent transporter) [Bacillus halodurans]
 Identities = 113/310 (36%), Positives = 176/310 (56%), Gaps = 9/310 (2%)

Query: 25  KVEQKKEVFQVNVNKLKMR---ETLISYAFLAPVLVFFVIFVLIPMIMGFVTSFFNYSM- 80
           +VE  +E        K K R    E     Y F+AP ++ F IF L PM+      SF ++ +
Sbjct: 4   EVETPRETKTTKARKQKRRLNKEYFWGYLFIAPPIIGFAIFALGPMLYSIYVSFTDFDLY 63

Query: 81  TEFTFVGFANYARMF-QDPIFMKSLINTLIIVIGSVPVVVFFSLFVAAKTYDKNVVARSF 139
            E  + G  NY R+F  D +F K++ NT     +G +P+ +  SL +A      K V   +
Sbjct: 64  NEPVWTGADNYYRLFVTDDLFRKTVFNTFYAALG-IPIGMAVSLGIAVALNQK-VKGIAL 121

Query: 140 YRAVFFLPVVTGSVAVTVVWKWIYDPMSGILNYVLKYAHVIEQNISWLGDKHWALLAIIV 199
           +R  FFLP V+  VA+T++W+WI++   G+LN +L Y  +      WL D+ WA+ A+I+
Sbjct: 122 FRTAFFLPAVSSVVAITLLWRWIFNADFGLLNIMLNYVGI--HGPGWLSDEKWAMPAMII 179

Query: 200 ILLTTSVGQPIILYIAAMGNIDNSLVEAARVDGATEFQVFWNIKWPSLLPTTLYIAVITT 259
             +   +G  +ILY+AA+  ++ +L EAA +DG   +Q F +I  PS+ PTT +I + +T
Sbjct: 180 QGVWGGLGINMILYLAALQGVNPALYEAADIDGGNAWQKFIHITVPSISPTTFFILITST 239

Query: 260 INSFQCFALIQLLTSGGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIAIISFAQ 319
           I + Q F    ++T GGPNYST+T++YYL+  AF+  E GYA+ M   L ++I II+
Sbjct: 240 IGALQDFQRFMIMTEGGPNYSTTTVVYYLFLNAFRYMEMGYASAMAWVLGIIILIITIIN 299

Query: 320 FKILGNDVEY 329
           FK+    V Y
Sbjct: 300 FKLAKKWVHY 309
```

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

```
 Identities = 263/295 (89%), Positives = 278/295 (94%)

Query: 1   MRTNKLKMRETMIAYAFLAPILLFFLIFVFAPMVMGFVTSFFNYSMTQFTFIGLANYNRM 60
           + NKLKMRET+I+YAFLAP+L+FF+IFV  PM+MGFVTSFFNYSMT+FTF+G ANY RM
Sbjct: 35  VNVNKLKMRETLISYAFLAPVLVFFVIFVLIPMIMGFVTSFFNYSMTEFTFVGFANYARM 94

Query: 61  FHDSIFMKSLINTVIIVIGSVPVVVFFSLFVAANTYEKNVFSRSFYRCVFFLPVVTGSVA 120
           F D IFMKSLINT+IIVIGSVPVVVFFSLFVAA TY+KNV +RSFYR VFFLPVVTGSVA
Sbjct: 95  FQDPIFMKSLINTLIIVIGSVPVVVFFSLFVAAKTYDKNVVARSFYRAVFFLPVVTGSVA 154

Query: 121 VTVVWKWIYDPMSGILNYILKSGHVIEQNISWLGDKHWALLAIIIILLTTSVGQPIILYI 180
           VTVVWKWIYDPMSGILNY+LK  HVIEQNISWLGDKHWALLAII+ILLTTSVGQPIILYI
Sbjct: 155 VTVVWKWIYDPMSGILNYVLKYAHVIEQNISWLGDKHWALLAIIVILLTTSVGQPIILYI 214

Query: 181 AAMGNIDNSLCEAARVDGANEMQVFWQIKWPSLLPTTLYIAVITTINSFQCFALIQLLTS 240
```

```
            AAMGNIDNSL EAARVDGA E QVFW IKWPSLLPTTLYIAVITTINSFQCFALIQLLTS
Sbjct: 215 AAMGNIDNSLVEAARVDGATEFQVFWNIKWPSLLPTTLYIAVITTINSFQCFALIQLLTS 274

Query: 241 GGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIALISFAQFKILGNDVEY 295
           GGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIA+ISFAQFKILGNDVEY
Sbjct: 275 GGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIAIISFAQFKILGNDVEY 329
```

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

Example 293

A DNA sequence (GBSx0321) was identified in *S.agalactiae* <SEQ ID 943> which encodes the amino acid sequence <SEQ ID 944>. 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>
```

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

```
>GP:CAB12516 GB:Z99107 similar to sugar-binding protein [Bacillus subtilis]
 Identities = 54/187 (28%), Positives = 90/187 (47%), Gaps = 14/187 (7%)

Query: 19  MFACVDSSQSVMAAEKD-KVEITWWAFPTFTQEKAKDGVGTYEKKVIKAFEKKNPNIKVK 77
           MF+   + +     ++D  + I WW       +   D    Y  KVI+ +EKKNP++ ++
Sbjct: 1   MFSGCSAGEEASGKKEDVTLRIAWWG-----GQPRHD----YTTKVIELYEKKNPHVHIE 51

Query: 78  LETIDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNGKLADLNDLFTDQFIKDVN-- 135
            E  ++    +K+     AG  PDV+      + QYGK  +L DL     D  I DV+
Sbjct: 52  AEFANWDDYWKKLAPMSAAGQLPDVIQMDTAYLAQYGKKNQLEDLTPYTKDGTI-DVSSI 110

Query: 136 NKNIIQASKSGDKAYMYPISSAPFYMAFNKKMLKDAGVLKLVKEGWTTSDFEKVLKALKN 195
           ++N++   K  +K Y + +      +  N+ +LK AGV  + +E WT  D+EK+   L+
Sbjct: 111 DENMLSGGKIDNKLYGFTLGVNVLSVIANEDLLKKAGV-SINQENWTWEDYEKLAYDLQE 169

Query: 196 KGYTPGS 202
           K   GS
Sbjct: 170 KAGVYGS 176
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 945> which encodes the amino acid sequence <SEQ ID 946>. Analysis of this protein sequence reveals the following:

```
      Possible site: 20

>>> 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:

```
!GB:Z99107 similar to sugar-binding protein [Bacillu...     82  2e-14

>GP:CAB12516 GB:Z99107 similar to sugar-binding protein [Bacillus subtilis]
 Identities = 105/446 (23%), Positives = 176/446 (38%), Gaps = 71/446 (15%)

Query: 24  GKSQKEAGASKSDTAKTEITWWAFPVFTQEKAEDGVGTYEKKLIAAFEKANPEIKVKLET 83
           G S  E  + K +     I WW       +   D    Y  K+I  +EK NP + ++ E
```

```
Sbjct: 4   GCSAGEEASGKKEDVTLRIAWWG-----GQPRHD----YTTKVIELYEKKNPHVHIEAEF 54

Query: 84  IDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNGKLADLNDLFTEEFTKDVN--NDK 141
            ++    +K+     AG  PDV+      + QYGK  +L DL   +T++ T DV+  ++
Sbjct: 55  ANWDDYWKKLAPMSAAGQLPDVIQMDTAYLAQYGKKNQLEDLTP-YTKDGTIDVSSIDEN 113

Query: 142 LIQASKAGDTAYMYPISSAPFYMALNKKMLKDAGVLDLVKEGWTTDDFEKVLKALKDK-- 199
            ++   K  +  Y + +      +  N+ +LK AGV  + +E WT +D+EK+   L++K
Sbjct: 114 MLSGGKIDNKLYGFTLGVNVLSVIANEDLLKKAGV-SINQENWTWEDYEKLAYDLQEKAG 172

Query: 200 -----GYNPGSFFANGQGGDQGPRAFFANLYSSHITDDKV---------------TKYTT 239
                G +P   F       +G R +  +       DD++                T  T
Sbjct: 173 VYGSNGMHPPDIFFPYYLRTKGERFYKEDGTGLAYQDDQLFVDYFERQLRLVKAKTSPTP 232

Query: 240 DDANSIKAMTKISNWIKDGLMMNGSQYDGSADIQNFANGQTSFTILWAPAQPGIQAKLLE 299
            D++  IK M       +D  ++ G     SA   N++N    F             A+L +
Sbjct: 233 DESAQIKGM-------EDDFIVKGK----SAITWNYSNQYLGF-----------ARLTD 269

Query: 300 ASKVDYLEIPFPSDDGKPELEYLVNGFAVFNNKDEQKVAASKTFIQFIADDKEWGPKNVV 359
            +    YL    P    + L    +         E K  A+K FI F  +++E   + +
Sbjct: 270 SPLSLYLP---PEQMQEKALTLKPSMLFSIPKSSEHKKEAAK-FINFFVNNEE-ANQLIK 324

Query: 360 RTGAFPVRTSYGDLYKDKRMEK---IAEWTKFYSPYYNTID-----GFAEMRTLWFPMVQ 411
                PV     D  K K  E+   I E+ +  S   +  D     G AE+  L
Sbjct: 325 GERGVPVSDKVADAIKPKLNEEETNIVEYVETASKNISKADPPEPVGSAEVIKLLKDTSD 384

Query: 412 AVSNGDEKPEDALKAFTEKANKTIKK 437
            +      PE A K F +KAN+ +++
Sbjct: 385 QILYQKVSPEKAAKTFRKKANEILER 410
```

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

```
Identities = 352/438 (80%), Positives = 384/438 (87%), Gaps = 4/438 (0%)

Query: 1   MSIKKSVIGFCLGAAALSMFACVDSSQSVMAAEKD---KVEITWWAFPTFTQEKAKDGVG 57
            M++KK      LGA+ L + AC   SQ    A K    K EITWWAFP FTQEKA+DGVG
Sbjct: 1   MNMKKLASLAMLGASVLGLAACGGKSQKEAGASKSDTAKTEITWWAFPVFTQEKAEDGVG 60

Query: 58  TYEKKVIKAFEKKNPNIKVKLETIDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNG 117
            TYEKK+I AFEK NP IKVKLETIDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNG
Sbjct: 61  TYEKKLIAAFEKANPEIKVKLETIDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNG 120

Query: 118 KLADLNDLFTDQFIKDVNNKNIIQASKSGDKAYMYPISSAPFYMAFNKKMLKDAGVLKLV 177
            KLADLNDLFT++F KDVNN  +IQASK+GD AYMYPISSAPFYMA NKKMLKDAGVL LV
Sbjct: 121 KLADLNDLFTEEFTKDVNNDKLIQASKAGDTAYMYPISSAPFYMALNKKMLKDAGVLDLV 180

Query: 178 KEGWTTSDFEKVLKALKNKGYTPGSFFANGQGGDQGPRAFFANLYSAPITDKEVTKYTTD 237
            KEGWTT DFEKVLKALK+KGY PGSFFANGQGGDQGPRAFFANLYS+ ITD +VTKYTTD
Sbjct: 181 KEGWTTDDFEKVLKALKDKGYNPGSFFANGQGGDQGPRAFFANLYSSHITDDKVTKYTTD 240

Query: 238 TKNSVKSMKKIVEWIKKGYLMNGSQYDGSADIQNFANGQTAFTILWAPAQPKTQAKLLES 297
              NS+K+M KI  WIK G +MNGSQYDGSADIQNFANGQT+FTILWAPAQP  QAKLLE+
Sbjct: 241 DANSIKAMTKISNWIKDGLMMNGSQYDGSADIQNFANGQTSFTILWAPAQPGIQAKLLEA 300

Query: 298 SKVDYLEVPFPSEDGKPDLEYLVNGFAVFNNKDENKVKASKKFITFIADDKKWGPKDVIR 357
            SKVDYLE+PFPS+DGKP+LEYLVNGFAVFNNKDE KV ASK FI FIADDK+WGPK+V+R
Sbjct: 301 SKVDYLEIPFPSDDGKPELEYLVNGFAVFNNKDEQKVAASKTFIQFIADDKEWGPKNVVR 360

Query: 358 TGAFPVRTSFGDLYKGDKRMMKISKWTQYYSPYYNTIDGFSEMRTLWFPMVQSVSNGDEK 417
            TGAFPVRTS+GDLYK DKRM KI++WT++YSPYYNTIDGF+EMRTLWFPMVQ+VSNGDEK
Sbjct: 361 TGAFPVRTSYGDLYK-DKRMEKIAEWTKFYSPYYNTIDGFAEMRTLWFPMVQAVSNGDEK 419

Query: 418 PADALKDFTQKANDTIKK 435
            P DALK FT+KAN TIKK
Sbjct: 420 PEDALKAFTEKANKTIKK 437
```

A related GBS gene <SEQ ID 8537> and protein <SEQ ID 8538> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:      5.05
GvH: Signal Score (-7.5): 4.69
     Possible site: 31
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   7.69 threshold:  0.0
   PERIPHERAL  Likelihood =  7.69      90
 modified ALOM score:  -2.04

*** 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:

```
28.8/48.4% over 409aa
                                                                    Bacillus subtilis
  EGAD|107689| hypothetical protein Insert characterized
  GP|2633010|emb|CAB12516.1||Z99107 similar to sugar-binding protein Insert characterized
  PIR|F69796|F69796 sugar-binding protein homolog yesO - Insert characterized

ORF01146(355 - 1605 of 1914)
EGAD|107689|BS0697(1    -    410    of    412)    hypothetical    protein    (Bacillus
subtilis)GP|2633010|emb|CAB12516.1||Z99107 similar to sugar-bindin
g protein (Bacillus subtilis)PIR|F69796|F69796 sugar-binding protein homolog yesO -
Bacillus subtilis
%Match = 5.4
%Identity = 28.8  %Similarity = 48.3
Matches = 69  Mismatches = 116  Conservative Sub.s = 47

318       348       378                 435       465       495       525
RGIVMSIKKSVIGFCLGAAALSMFACVDSSQSVMAAEKD-KVEITWWAFPTFTQEKAKDGVGTYEKKVIKAFEKKNPNIK
                      ||:   : :     ::|  : | ||              |  |||: :|||||::
                      MFSGCSAGEEASGKKEDVTLRIAWW---------GGQPRHDYTTKVIELYEKKNPHVH
                                10        20                30        40

555       585       615       645       675       705       732       762
VKLETIDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNGKLADLNDLFTDQFIKDVN-NKNIIQASKSGDKAYMYPI
:: |  ::    :|:     ||  |||:      : ||||  :| ||     |  |   : ::|::   |  :| | : :
IEAEFANWDDYWKKLAPMSAAGQLPDVIQMDTAYLAQYGKKNQLEDLTPYTKDGTIDVSSIDENMLSGGKIDNKLYGFTL
         60        70        80        90       100       110       120

792       822       852       882       912       942       972
SSAPFYMAFNKKMLKDAGVLKLVKEGWTTSDFEKVLKALKNKGYTPGSFFANGQGGDQGPRAFFANLYSA----------
    : :  |: :|| |||  : :| ||  |:||:   |: |    | : :||    : |  ||
GVNVLSVIANEDLLKKAGV-SINQENWTWEDYEKLAYDLQEK---AGVYGSNGM---HPPDIFFPYYLRTKGERFYKEDG
        140       150       160       170             180       190       200

990       1020      1050      1080
----------------------------------PITDKEVTKYTTDTKNSVKSMKKIVEWIKKGYLMNGSQYDGSA~~~
                                   |:   || |   || : :   |
TGLAYQDDQL~~~~NIVEYVETASKNISKADPPEPVGSAEVIKLLKDTSDQILYQKV----------------------~~~
             350       360       370       380       390

1515      1545      1575      1605      1635      1665      1695      1725
FSEMRTLWFPMVQSVSNGDEKPADALKDFTQKANDTIKKAAK*LRRLLFYGQSHIGIEEEFLVKLRCKGEYRMRTNKLK
                       | | | | :|||: :::
--------------------SPEKAAKTFRKKANEILERNN
```

SEQ ID 944 (GBS16) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 9; MW 49kDa).

The GBS16-His fusion product was purified (Figure 92A; see also Figure 189, lane 9) and used to immunise mice (lane 1 + 2 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 92B), FACS (Figure 92C), 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 294

A DNA sequence (GBSx0322) was identified in *S.agalactiae* <SEQ ID 947> which encodes the amino acid sequence <SEQ ID 948>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> 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>
```

A related GBS nucleic acid sequence <SEQ ID 9459> which encodes amino acid sequence <SEQ ID 9460> was also identified.

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

```
>GP:AAC66999 GB:AE001166 conserved hypothetical protein [Borrelia
          burgdorferi]
 Identities = 107/225 (47%), Positives = 147/225 (64%), Gaps = 6/225 (2%)

Query: 12  QIKNGIIVSCQALPGEPLYTESGGVMPLLALAAQEAGAVGIRANSVRDIKEIQEVTNLPI 71
           +IK G+IVSCQAL  EPL+  S  +M  +ALAA+  GA+GIRAN V DI +I+   +LPI
Sbjct: 6   KIKRGLIVSCQALENEPLH--SSFIMSKMALAAKIGGAIGIRANGVNDISQIKLEVDLPI 63

Query: 72  IGIIKREYPPQEPFITATMTEVDQLASLDIAVIALDCTLRERHDGLSVVEFIQKIKRKYP 131
           IGIIK+ Y   + FIT TM E+D+L +  + +IALD T R R DG+ + +F + IK+KYP
Sbjct: 64  IGIIKKNYNNCDVFITPTMKEIDELCNEGVDIIALDATFRNRPDGVLLDDFFENIKKKYP 123

Query: 132 EQLLMADISTFEEGKNAFEAGVDFVGTTLSGYTDYSR--QEEGPDIELLNKLCQAGI--D 187
           +Q LMADIS+ +E  NA + G DF+GTTL GYT  +         D   L  L + +
Sbjct: 124 KQCLMADISSLDEAINADKLGFDFIGTTLYGYTKNTNGLNIADNDFNFLRTLLNSNLKST 183

Query: 188 VIAEGKIHTPKQANEINHIGVAGIVVGGAITRPKEIAERFISGLS 232
           +I EGKI TP +A +   +GV  +VVGGAITRP EI ++F+  ++
Sbjct: 184 LIVEGKIDTPLKAQKCFEMGVDLVVVGGAITRPAEITKKFVEKIN 228
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 949> which encodes the amino acid sequence <SEQ ID 950>. Analysis of this protein sequence reveals the following:

```
     Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.49   Transmembrane  175 - 191 ( 175 - 192)

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

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

```
>GP:AAD28762 GB:AF130859 putative N-acetylmannosamine-6-P epimerase
          [Clostridium perfringens]
 Identities = 113/225 (50%), Positives = 148/225 (65%), Gaps = 5/225 (2%)

Query: 10  LMEQLKGGIIVSCQALPGEPLYSETGGIMPLMAKAAQEAGAVGIRANSVRDIKEIQAITD 69
           +++ +KG +IVSCQAL  EPL+S    IM  MA AA++ GA  IRA  + DI EI+ +T
Sbjct: 1   MLDVVKGNLIVSCQALSDEPLHSSF--IMGRMAIAAKQGGAAAIRAQGIDDINEIKEVTK 58

Query: 70  LPIIGIIKKDYPPQEPFITATMTEVDQLAALNIAVIAMDCTKRDRHDGLDIASFIRQVKE 129
           LPIIGIIK++Y   E +IT TM EVD+L   +  +I +D TKR R +G +I   +  +
Sbjct: 59  LPIIGIIKRNYDDSEIYITPTMKEVDELLKTDCEMIGLDATKRKRPNGENIKDLVDAIHA 118

Query: 130 KYPNQLLMADISTFDEGLVAHQAGIDFVGTTLSGYTPYSRQEAGPDVALIEALCK-AGIA 188
           K   +L MADIST +EG+ A + G D V TTLSGYTPYS+Q    D  L+E L K   I
Sbjct: 119 K--GRLAMADISTLEEGIEAEKLGFDCVSTTLSGYTPYSKQSNSVDFELLEELVKTVKIP 176

Query: 189 VIAEGKIHSPEEAKKINDLGVAGIVVGGAITRPKEIAERFIEALK 233
           VI EG+I++PEE KK  DLG    VVGGAITRP++I +RF + LK
Sbjct: 177 VICEGRINTPEELKKALDLGAYSAVVGGAITRPQQITKRFTDILK 221
```

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

```
 Identities = 172/227 (75%), Positives = 202/227 (88%)

Query: 5   SKEAFKKQIKNGIIVSCQALPGEPLYTESGGVMPLLALAAQEAGAVGIRANSVRDIKEIQ 64
           +KE   +Q+K GIIVSCQALPGEPLY+E+GG+MPL+A AAQEAGAVGIRANSVRDIKEIQ
Sbjct: 6   TKEKLMEQLKGGIIVSCQALPGEPLYSETGGIMPLMAKAAQEAGAVGIRANSVRDIKEIQ 65

Query: 65  EVTNLPIIGIIKREYPPQEPFITATMTEVDQLASLDIAVIALDCTLRERHDGLSVVEFIQ 124
            +T+LPIIGIIK++YPPQEPFITATMTEVDQLA+L+IAVIA+DCT R+RHDGL +  FI+
Sbjct: 66  AITDLPIIGIIKKDYPPQEPFITATMTEVDQLAALNIAVIAMDCTKRDRHDGLDIASFIR 125

Query: 125 KIKRKYPEQLLMADISTFEEGKNAFEAGVDFVGTTLSGYTDYSRQEEGPDIELLNKLCQA 184
           ++K KYP QLLMADISTF+EG  A +AG+DFVGTTLSGYT YSRQE GPD+ L+  LC+A
Sbjct: 126 QVKEKYPNQLLMADISTFDEGLVAHQAGIDFVGTTLSGYTPYSRQEAGPDVALIEALCKA 185

Query: 185 GIDVIAEGKIHTPKQANEINHIGVAGIVVGGAITRPKEIAERFISGL 231
           GI VIAEGKIH+P++A +IN +GVAGIVVGGAITRPKEIAERFI  L
Sbjct: 186 GIAVIAEGKIHSPEEAKKINDLGVAGIVVGGAITRPKEIAERFIEAL 232
```

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

Example 295

A DNA sequence (GBSx0323) was identified in *S.agalactiae* <SEQ ID 951> which encodes the amino acid sequence <SEQ ID 952>. This protein is predicted to be group B streptococcal surface immunogenic protein. 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 DNA sequence was identified in *S.pyogenes* <SEQ ID 953> which encodes the amino acid sequence <SEQ ID 954>. 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>
```

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

```
 Identities = 182/437 (41%), Positives = 240/437 (54%), Gaps = 53/437 (12%)

Query: 1   MKMNKKVLLTSTMAASLLSVASVQAQETDTTWTARTVSEVKADLVKQDNKSSYTVKYGDT 60
           M + KK L  +++A SL+ +A+ QAQE     WT R+V+E+K++LV  DN  +YTVKYGDT
Sbjct: 1   MIITKKSLFVTSVALSLVPLATAQAQE----WTPRSVTEIKSELVLVDNVFTYTVKYGDT 56

Query: 61  LSVISEAMSIDMNVLAKINNIADINLIYPETTLTVTYDQKSHTATSMKIETPATNAAGQT 120
           LS I+EAM ID++VL  IN+IA+I+LI+P+T LT  Y+Q    AT++ ++ PA++ A  +
Sbjct: 57  LSTIAEAMGIDVHVLGDINHIANIDLIFPDTILTANYNQHGQ-ATNLTVQAPASSPASVS 115

Query: 121 TATVDLKTNQVSVADQKVSLNTISEGMTP-EAATTIVSPMKTYSSAPALKSKEVLAQEQA 179
                  Q S   Q      ++   TP +  TT  +  K  SS  A  S E+ +
Sbjct: 116 HVPSSEPLPQASATSQPTV--PMAPPATPSDVPTTPFASAKPDSSVTA--SSELTSSTND 171

Query: 180 VSQAAANEQVSPAPVKSITSEVPAAKEEVKPTQTSVSQSTTVSPASVAAETPAPVAKVAP 239
           VS   ++E      V       P A E      T V   T +S A  +A  P P    +
Sbjct: 172 VSTELSSESQKQPEVPQEAVPTPKAAE-----TTEVEPKTDISEAPTSANRPVPNESASE 226

Query: 240 VRTVAAPRVASVKVVTPKVETGASPEHVSAPAVP---VTTTSPATDSKLQATEVKSVPVA 296
             + AAP            +  A  E  SAPA       TTS AT + L
Sbjct: 227 EVSSAAP----------AQAPAEKEETSAPAAQKAVADTTSVATSNGL---------- 264

Query: 297 QKAPTATPVAQPASTTNAVAAHPENAGLQPHVAAYKEKVASTYGVNEFSTYRAGDPGDHG 356
             AP               A +P NAGLQP  AA+KE+VAS +G+  FS YR GDPGDHG
Sbjct: 265 SYAPNH----------AYNPMNAGLQPQTAAFKEEVASAFGITSFSGYRPGDPGDHG 311

Query: 357 KGLAVDFIVGTNQALGNKVAQYSTQNMAANNISYVIWQQKFYSNTNSIYGPANTWNAMPD 416
           KGLA+DF+V  N ALG++VAQY+  +MA   ISYVIW+Q+FY+   SIYGPA TWN MPD
Sbjct: 312 KGLAIDFMVPENSALGDQVAQYAIDHMAERGISYVIWKQRFYAPFASIYGPAYTWNPMPD 371

Query: 417 RGGVTANHYDHVHVSFN 433
           RG +T NHYDHVHVSFN
Sbjct: 372 RGSITENHYDHVHVSFN 388
```

A related GBS gene <SEQ ID 8539> and protein <SEQ ID 8540> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length of UR:  20
     Peak Value of UR:   1.96
     Net Charge of CR: 2
McG: Discrim Score:      2.95
GvH: Signal Score (-7.5): 3.84
     Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 24
ALOM program   count: 0 value:   4.29 threshold:  0.0
   PERIPHERAL  Likelihood =  4.29      58
 modified ALOM score:  -1.36

*** Reasoning Step: 3

Rule gpo1

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

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

SEQ ID 8540 (GBS322) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 9; MW 52kDa). The GBS322-His fusion product was purified (Figure 214, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 267), 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 296

A DNA sequence (GBSx0324) was identified in *S.agalactiae* <SEQ ID 955> which encodes the amino acid sequence <SEQ ID 956>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.86   Transmembrane    5 -  21 (   4 -  21)

----- Final Results -----
               bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAC46072 GB:U50357 zoocin A endopeptidase [Streptococcus
          zooepidemicus]
 Identities = 163/274 (59%), Positives = 196/274 (71%), Gaps = 11/274 (4%)

Query: 25  VLADTYVRPIDNGRITTGFNGYPGHCGVDYAVPTGTIIRAVADGTVKFAGAGANFSWMTD 84
           V A TY RP+D G ITTGFNGYPGH GVDYAVP GT +RAVA+GTVKFAG GAN  WM
Sbjct: 21  VSAATYTRPLDTGNITTGFNGYPGHVGVDYAVPVGTPVRAVANGTVKFAGNGANHPWMLW 80

Query: 85  LAGNCVMIQHADGMHSGYAHMSRVVARTGEKVKQGDIIGYVGATGMATGPHLHFEFLPAN 144
           +AGNCV+IQHADGMH+GYAH+S++   T   VKQG IIGY GATG  TGPHLHFE LPAN
Sbjct: 81  MAGNCVLIQHADGMHTGYAHLSKISVSTDSTVKQGQIIGYTGATGQVTGPHLHFEMLPAN 140

Query: 145 PNFQNGFHGRINPTSLIANVATFSGKTQASAPSIKPLQSAPVQNQSSKLKVYRVDELQKV 204
           PN+QNGF GRI+PT  IAN   F+G T           + P  N    LK+Y+VD+LQK+
Sbjct: 141 PNWQNGFSGRIDPTGYIANAPVFNGTTPTE-------PTTPTTN----LKIYKVDDLQKI 189

Query: 205 NGVWLVKNNTLTPTGFDWNDNGIPASEIDEVDANGNLTADQVLQKGGYFIFNPKTLKTVE 264
           NG+W V+NN L PT F W DNGI A ++ EV +NG  T+DQVLQKGGYF+ NP  +K+V
Sbjct: 190 NGIWQVRNNILVPTDFTWVDNGIAADDVIEVTSNGTRTSDQVLQKGGYFVINPNNVKSVG 249

Query: 265 KPIQGTAGLTWAKTRFANGSSVWLRVDNSQELLY 298
            P++G+ GL+WA+  F  G +VWL   +   LLY
Sbjct: 250 TPMKGSGGLSWAQVNFTTGGNVWLNTTSKDNLLY 283
```

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

A related GBS gene <SEQ ID 8541> and protein <SEQ ID 8542> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:      6.63
GvH: Signal Score (-7.5): -2.97
     Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -1.86 threshold:  0.0
   INTEGRAL    Likelihood = -1.86   Transmembrane    5 -  21 (   4 -  21)
   PERIPHERAL  Likelihood =  5.57      50
```

```
 modified ALOM score:   0.87

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
GP|2804351|gb|AAC46072.1||U50357(21 - 283 of 285) zoocin A endopeptidase (Streptococcus
zooepidemicus)
%Match = 34.2
%Identity = 61.3  %Similarity = 74.4
Matches = 163  Mismatches = 65  Conservative Sub.s = 35

144       174       204       234       264       294       324       354
VV*VFLS*LRYTTYILKTFLFIKPPKYSSR*VLFLIF*FKFSNKLIASV*ALHYINSIWRFFLNKWLVKASSLVVLGGMV

                                                                    MKRIFFAFLSLCLF
                                                                              10

384       414       444       474       504       534       564       594
LSAGSRVLADTYVRPIDNGRITTGFNGYPGHCGVDYAVPTGTIIRAVADGTVKFAGAGANFSWMTDLAGNCVMIQHADGM
:     | | || ||:| | |||||||||||| ||||||| || :||||:||||||| ||| || :|||||:|||||||
IFGTQTVSAATYTRPLDTGNITTGFNGYPGHVGVDYAVPVGTPVRAVANGTVKFAGNGANHPWMLWMAGNCVLIQHADGM
               30        40        50        60        70        80        90

624       654       684       714       744       774       804       834
HSGYAHMSRVVARTGEKVKQGDIIGYVGATGMATGPHLHFEFLPANPNFQNGFHGRINPTSLIANVATFSGKTQASAPSI
|:||||:|::   |   |||| |||| |||| |||||||| ||||||:|||| |||:|| |||   |:| |
HTGYAHLSKISVSTDSTVKQGQIIGYTGATGQVTGPHLHFEMLPANPNWQNGFSGRIDPTGYIANAPVFNGTT-------
             110       120       130       140       150       160

864       894       924       954       984      1014      1044      1074
KPLQSAPVQNQSSKLKVYRVDELQKVNGVWLVKNNTLTPTGFDWNDNGIPASEIDEVDANGNLTADQVLQKGGYFIFNPK
 |  : |    :: ||:|:||:|||:||:| |:|| | || | | |||| | :: || :||  |:||||||||||: ||
-P--TEP-TTPTTNLKIYKVDDLQKINGIWQVRNNILVPTDFTWVDNGIAADDVIEVTSNGTRTSDQVLQKGGYFVINPN
             180       190       200       210       220       230       240

1104      1134      1164      1194      1224      1254      1284      1314
TLKTVEKPIQGTAGLTWAKTRFANGSSVWLRVDNSQELLYK*FEVLIHCFK*QLCY*LSTISLNRLKIIL*SSKV*YYSL
 :|:|  |::|: ||:||:  |  | :|||  :  |||
NVKSVGTPMKGSGGLSWAQVNFTTGGNVWLNTTSKDNLLYGK
               260       270       280
```

SEQ ID 8542 (GBS36) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 4; MW 34.1kDa).

GBS36-His was purified as shown in Figure 192, lane 7.

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

Example 297

A DNA sequence (GBSx0325) was identified in *S.agalactiae* <SEQ ID 957> which encodes the amino acid sequence <SEQ ID 958>. This protein is predicted to be phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohyd. Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2815(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:BAB04352 GB:AP001509 phosphoribosylaminoimidazolecarboxamide
          formyltransferase/IMP cyclohydrolase [Bacillus halodurans]
 Identities = 310/515 (60%), Positives = 390/515 (75%), Gaps = 4/515 (0%)

Query: 1   MTKRALISVSDKSGIIDFAKELKNLGWDIISTGGTKVALDDAGVETIAIDDVTGFPEMMD 60
           M +RAL+SVS+K GI+ FAK L      +I+STGGTK AL +AG+     I DVTGFPE++D
Sbjct: 1   MKRRALVSVSNKEGIVPFAKALVEHEVEIVSTGGTKRALQEAGIPVTGISDVTGFPEILD 60

Query: 61  GRVKTLHPNIHGGLLARRDADSHLQAAKDNNIELIDLVVVNLYPFKETILRPDVTYDLAV 120
           GRVKTLHPNIHGGLLA R+ D HL     +++I  ID VVVNLYPF++TI +P+ T+  A+
Sbjct: 61  GRVKTLHPNIHGGLLAMRERDEHLAQLNEHHIRPIDFVVVNLYPFQQTIAKPEATFADAI 120

Query: 121 ENIDIGGPSMLRSAAKNHASVTVVVDSADYATVLGELADASQTTFKTRQRLAAKAFRHTA 180
           ENIDIGGPSMLR+AAKNH  VTVVVD  DY TVL ELAD       +T++RLAAK FRHTA
Sbjct: 121 ENIDIGGPSMLRAAAKNHQHVTVVVDPVDYETVLKELADQGNVATETKRRLAAKVFRHTA 180

Query: 181 AYDALIAEYFTAQVGEAKPEKLTITYDLKQAMRYGENPQQDADFYQKALPTDYSIASAKQ 240
           AYDA+IAEY T  VGE  PE LT+T++ KQ +RYGENP Q A FYQK L    SIA AKQ
Sbjct: 181 AYDAMIAEYLTDAVGEESPESLTVTFEKKQDLRYGENPHQKATFYQKPLGAKASIAHAKQ 240

Query: 241 LNGKELSFNNIRDADAAIRIIRDFKDSPTVVALKHMNPCGIGQADDIETAWDYAYEADPV 300
           L+GKELS+NNI DADAA+ I+++FK+ P  VA+KHMNPCG+G  + I+ A+D AYEADPV
Sbjct: 241 LHGKELSYNNINDADAALSIVKEFKE-PAAVAVKHMNPCGVGTGETIKEAFDKAYEADPV 299

Query: 301 SIFGGIVVLNREVDAATAEKMHPIFLEIIIAPSYSEEALAILTNKKKNLRILELPFDAQA 360
           SIFGGI+ LNREVD  TA+ +  IFLEIIIAPS+SEEAL +LT+ KKNLR+L LP + +
Sbjct: 300 SIFGGIIALNREVDVETAKTLKEIFLEIIIAPSFSEEALDVLTS-KKNLRLLTLPLNEE- 357

Query: 361 ASEVEAEYTGVVGGLLVQNQDVVAENPSDWQVVTDRQPTEQEATALEFAWKAIKYVKSNG 420
            ++ E   T + GG LVQ +D      ++ ++ T R+PTE E  AL+ AW+ +K+VKSN
Sbjct: 358 -NQAEKRITSIHGGALVQEEDTYGFEEAEIKIPTKREPTEAEWEALKLAWRVVKHVKSNA 416

Query: 421 IIITNDHMTLGLGAGQTNRVGSVKIAIEQAKDHLDGAVLASDAFFPFADNIEEIAAAGIK 480
           I++ +  MT+G+GAGQ NRVG+ KIAIEQA +   G+V+ SDAFFP  D +E  A AGI
Sbjct: 417 IVLADGQMTVGVGAGQMNRVGAAKIAIEQAGEKAAGSVMGSDAFFPMGDTVELAAKAGIT 476

Query: 481 AIIQPGGSVRDQESIDAANKHGLTMIFTGVRHFRH 515
           AIIQPGGS+RD+ESI+ A+KHG+ M+FTGVRHF+H
Sbjct: 477 AIIQPGGSIRDEESIENADKHGIAMVFTGVRHFKH 511
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 959> which encodes the amino acid sequence <SEQ ID 960>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2932(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 = 500/515 (97%), Positives = 507/515 (98%)

Query: 1   MTKRALISVSDKSGIIDFAKELKNLGWDIISTGGTKVALDDAGVETIAIDDVTGFPEMMD 60
           MTKRALISVSDKSGI+DFAKELKNLGWDIISTGGTKV LDDAGVETIAIDDVT FPEMMD
Sbjct: 1   MTKRALISVSDKSGIVDFAKELKNLGWDIISTGGTKVTLDDAGVETIAIDDVTRFPEMMD 60

Query: 61  GRVKTLHPNIHGGLLARRDADSHLQAAKDNNIELIDLVVVNLYPFKETILRPDVTYDLAV 120
```

```
            GRVKTLHPNIHGGLLARRDADSHLQAAKDNNIELIDLVVVNLYPFKETILRPD+TYDLAV
Sbjct: 61   GRVKTLHPNIHGGLLARRDADSHLQAAKDNNIELIDLVVVNLYPFKETILRPDITYDLAV 120

Query: 121 ENIDIGGPSMLRSAAKNHASVTVVVDSADYATVLGELADASQTTFKTRQRLAAKAFRHTA 180
            ENIDIGGPSMLRSAAKNHASVTVVVD ADYATVLGELADA QTTF+TRQRLAAK FRHTA
Sbjct: 121 ENIDIGGPSMLRSAAKNHASVTVVVDPADYATVLGELADAGQTTFETRQRLAAKVFRHTA 180

Query: 181 AYDALIAEYFTAQVGEAKPEKLTITYDLKQAMRYGENPQQDADFYQKALPTDYSIASAKQ 240
            AYDALIAEYFT QVGEAKPEKLTITYDLKQAMRYGENPQQDADFYQKALPTDYSIASAKQ
Sbjct: 181 AYDALIAEYFTTQVGEAKPEKLTITYDLKQAMRYGENPQQDADFYQKALPTDYSIASAKQ 240

Query: 241 LNGKELSFNNIRDADAAIRIIRDFKDSPTVVALKHMNPCGIGQADDIETAWDYAYEADPV 300
            LNGKELSFNNIRDADAAIRIIRDFKD PTVVALKHMNPCGIGQADDIETAWDY Y+ADPV
Sbjct: 241 LNGKELSFNNIRDADAAIRIIRDFKDRPTVVALKHMNPCGIGQADDIETAWDYTYKADPV 300

Query: 301 SIFGGIVVLNREVDAATAEKMHPIFLEIIIAPSYSEEALAILTNKKKNLRILELPFDAQA 360
            SIFGGI+VLNREVDAATA+KMHPIFLEIIIAPSYSEEALAILTNKKKNLRILELPFDAQA
Sbjct: 301 SIFGGIIVLNREVDAATAKKMHPIFLEIIIAPSYSEEALAILTNKKKNLRILELPFDAQA 360

Query: 361 ASEVEAEYTGVVGGLLVQNQDVVAENPSDWQVVTDRQPTEQEATALEFAWKAIKYVKSNG 420
            ASEVEAEYTGVVGGLLVQNQDVVAENPSDWQVVTDRQPTEQEATALEFAWKAIKYVKSNG
Sbjct: 361 ASEVEAEYTGVVGGLLVQNQDVVAENPSDWQVVTDRQPTEQEATALEFAWKAIKYVKSNG 420

Query: 421 IIITNDHMTLGLGAGQTNRVGSVKIAIEQAKDHLDGAVLASDAFFPFADNIEEIAAAGIK 480
            IIITNDHMTLGLGAGQTNRVGSVKIAIEQAKDHLDGAVLASDAFFPFADNIEEIAAAGIK
Sbjct: 421 IIITNDHMTLGLGAGQTNRVGSVKIAIEQAKDHLDGAVLASDAFFPFADNIEEIAAAGIK 480

Query: 481 AIIQPGGSVRDQESIDAANKHGLTMIFTGVRHFRH 515
            AIIQPGGSVRDQ+SIDAANKHGLTMIFTGVRHFRH
Sbjct: 481 AIIQPGGSVRDQDSIDAANKHGLTMIFTGVRHFRH 515
```

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

Example 298

A DNA sequence (GBSx0326) was identified in *S.agalactiae* <SEQ ID 961> which encodes the amino acid sequence <SEQ ID 962>. This protein is predicted to be similar to antibiotic resistance protein. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1842(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:CAB12342 GB:Z99106 similar to antibiotic resistance protein
            [Bacillus subtilis]
 Identities = 65/263 (24%), Positives = 117/263 (43%), Gaps = 34/263 (12%)

Query: 5    KNLEIVESIFGD-WDETIIWSCV-QGIMGEVFVDSLDQPKSSLAKLGRKSSFGFLAGQPT 62
             K    ++++F D +  T ++S + Q I G V+ D    PKS    +G +S   F+AG
Sbjct: 10   KKYSSLKTMFDDKYCPTFVYSILDQTIPGAVYADDQTFPKSFF--IGTESGIYFIAGDQG 67

Query: 63   ----------LFLLEVCSGEDIILVPQHKGWSDLIESTYGQNAHSFKRYATKKDTLFERS 112
                        +  +V S +   L      W  +++     + +  +R A         +
Sbjct: 68   NRDFHDFIAGYYEEQVKSSKRFTLFSSSDTWDSVLKPILKDDLNQMRRAAFSY-----QP 122

Query: 113  RLEKFVTQLPNGFELRAIDEKV------YNSCLEKEWSQDLVANYATYQYYKKQGIGYVV 166
             +  K   QLP G  L+ IDE +      +NS   +E+       + +   +G G+ V
Sbjct: 123  KSFKKTLQLPKGLVLKRIDEDIISHSTAFNSAYYEEY-------WNSVSQFASKGFGFAV 175
```

```
Query: 167 YYQGNIIAGASSYSTYKNGIEIEVDTHPDFRRRGLATIVAAQLILTCLDKGIYPSWDAH- 225
           +  ++++  +S     N  E+++ T  ++R  GLA  VA + I  C++ GI PSWD
Sbjct: 176 LHGNHVVSECTSIFLGHNRAEMDIYTLEEYRGLGLAYCVANRFIAFCMENGIVPSWDCDI 235

Query: 226 -TRTSLNLSEKLGYEFSHEYIAY 247
             +S+ L+ KLG++   EY  Y
Sbjct: 236 CNNSSIALAAKLGFKTVTEYTIY 258
```

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 299

A DNA sequence (GBSx0328) was identified in *S.agalactiae* <SEQ ID 963> which encodes the amino acid sequence <SEQ ID 964>. This protein is predicted to be phosphoribosylglycinamide formyltransferase homolog (purN). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0736(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 965> which encodes the amino acid sequence <SEQ ID 966>. Analysis of this protein sequence reveals the following:

```
     Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.53   Transmembrane   75 -  91 (  75 -  91)

----- Final Results -----
             bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAA04374 GB:AJ000883 purD [Lactococcus lactis]
 Identities = 236/419 (56%), Positives = 301/419 (71%), Gaps = 7/419 (1%)

Query: 50  LKLLVVGSGGREHAIAKKLLASKGVDQVFVAPGNDGMTLDGLDLVNIVVSEHSRLIAFAK 109
           +K+LV+GSGGREHA+AKK + S  V++VFVAPGN GM  DG+ +V+I    + +L+ FA+
Sbjct: 1   MKILVIGSGGREHALAKKFMESPQVEEVFVAPGNSGMEKDGIQIVHISELSNDKLVKFAQ 60

Query: 110 ENEISWAFIGPDDALAAGIVDDFNSAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA 169
              I   F+GP+ AL  G+VD F  A L  FGP K AAELE SKDFAK IM KY VPTA
Sbjct: 61  NQNIGLTFVGPETALMNGVVDAFIKAELPIFGPNKMAAELEGSKDFAKSIMKKYGVPTAD 120

Query: 170 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 229
           Y TF   E A AY++E+G P+V+KADGLA GKGV VA  +E A  A  ++      F  S
Sbjct: 121 YATFDSLEPALAYLDEKGVPLVIKADGLAAGKGVTVAFDIETAKSALADI-----FSGSQ 175

Query: 230 ARVVIEEFLDGEEFSLFAFANGDKFYIMPTAQDHKRAFDGDKGPNTGGMGAYAPVPHLPQ 289
            +VVIEEFLDGEEFSLF+F +  K Y MP AQDHKRAFD DKGPNTGGMGAY+PV H+ +
Sbjct: 176 GKVVIEEFLDGEEFSLFSFIHDGKIYPMPIAQDHKRAFDEDKGPNTGGMGAYSPVLHISK 235

Query: 290 SVVDTAVEMIVRPVLEGMVAEGRPYLGVLYVGLILTADGPKVIEFNSRFGDPETQIILPR 349
            VV+ A+E +V+P + GM+ EG+ + GVLY GLILT DG K IEFN+RFGDPETQ++LPR
Sbjct: 236 EVVNEALEKVVKPTVAGMIEEGKSFTGVLYAGLILTEDGVKTIEFNARFGDPETQVVLPR 295
```

```
Query: 350 LTSDFAQNIDDIMMGIEPYITWQKDGVTLGVVVASEGYPFDYEKGVPLPEKTDGDIITYY 409
           L SD AQ I DI+ G EP + W + GVTLGVVVA+EGYP   + G+ LPE  +G +  YY
Sbjct: 296 LKSDLAQAIIDILAGNEPTLEWLESGVTLGVVVAAEGYPSQAKLGLILPEIPEG-LNVYY 354

Query: 410 AGVKFSENSELLLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQQDTTGLFYRNDIGSKAI 468
           AGV  +EN++ L+S+GGRVY++  T + VK+ Q  +Y +L + +  G FYR+DIGS+AI
Sbjct: 355 AGVSKNENNQ-LISSGGRVYLVSETGEDVKSTQKLLYEKLDKLENDGFFYRHDIGSRAI 412
```

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

```
 Identities = 172/182 (94%), Positives = 176/182 (96%)

Query: 1   MKIAVFASGNGSNFQVIAEQFQVSFVFSDHRDAYVLERAQNLAIPSFAFELKEFENKAAY 60
           MKIAVFASGNGSNFQVIAEQF VSFVFSDHRDAYVLERAQNLAIPSFAFELKEFENK AY
Sbjct: 1   MKIAVFASGNGSNFQVIAEQFPVSFVFSDHRDAYVLERAQNLAIPSFAFELKEFENKVAY 60

Query: 61  EQAVVDLLDKHEIDLVCLAGYMKIVGETLLSAYEGRIINIHPTYLPEFPGAHGIKDAWEA 120
           EQA+VDLLDKHEIDLVCLAGYMKIVGETLL AYE RIINIHP YLPEFPGAHGI+DAWEA
Sbjct: 61  EQAIVDLLDKHEIDLVCLAGYMKIVGETLLLAYERRIINIHPAYLPEFPGAHGIEDAWEA 120

Query: 121 GVDQSGVTIHWVDSGVDTGQVIQQVHVPRLADDSLESFETRIHETEYQLYPAVLDSLGIK 180
           GVDQSGVTIHWVDSGVDTGQVIQQV VPRLADDSLESFETRIHETEYQLYPAVLDSLG++
Sbjct: 121 GVDQSGVTIHWVDSGVDTGQVIQQVRVPRLADDSLESFETRIHETEYQLYPAVLDSLGVE 180

Query: 181 RK 182
           RK
Sbjct: 181 RK 182
```

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

Example 300

A DNA sequence (GBSx0329) was identified in *S.agalactiae* <SEQ ID 967> which encodes the amino acid sequence <SEQ ID 968>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.59   Transmembrane  121 - 137 ( 121 - 137)

----- 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 GENPEPT database:

```
>GP:AAC16901 GB:AF016634 phosphoribosylformylglycinamide
           cyclo-ligase [Lactococcus lactis subsp. cremoris]
 Identities = 253/338 (74%), Positives = 288/338 (84%), Gaps = 4/338 (1%)

Query: 4   KNAYAQSGVDVEAGYEVVERIKKHVARTERAGVMGALGGFGGMFDLSQTGVKEPVLISGT 63
           +NAYA+SGVDVEAGYEVV RIKKHVA+TER GV+GALGGFGG FDLS   VKEPVLISGT
Sbjct: 5   ENAYAKSGVDVEAGYEVVSRIKKHVAKTERLGVLGALGGFGGSFDLSVLDVKEPVLISGT 64

Query: 64  DGVGTKLMLAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLYFLDYVATGKNEPAKLEQVVA 123
           DGVGTKLMLAI+ DKHDTIG DCVAMCVNDIIAAGAEPLYFLDY+ATGKN P KLEQVVA
Sbjct: 65  DGVGTKLMLAIRADKHDTIGIDCVAMCVNDIIAAGAEPLYFLDYIATGKNIPEKLEQVVA 124

Query: 124 GVAEGCVQASAALIGGETAEMPGMYGEDDYDLAGFAVGVAEKSQIIDGSK-VKEGDILLG 182
           GVAEGC+QA AALIGGETAEMPGMY EDDYDLAGFAVGVAEKSQ+IDG K V+ GD+LLG
Sbjct: 125 GVAEGCLQAGAALIGGETAEMPGMYDEDDYDLAGFAVGVAEKSQLIDGEKDVEAGDVLLG 184

Query: 183 LASSGIHSNGYSLVRRVFADYTGDEVLPELEGKQLKDVLLEPTRIYVKAALPLIKEELVN 242
           LASSGIHSNGYSLVR+VFAD+  +E LPEL+ + L D LL PT+IYVK  LPLIK+  +
```

```
Sbjct: 185 LASSGIHSNGYSLVRKVFADFDLNESLPELD-QSLIDTLLTPTKIYVKELLPLIKQNKIK 243

Query: 243 GIAHITGGGFIENVPRMFADDLAAEIDEDKVPVLPIFKALEKYGDIKHEEMFEIFNMGVG 302
           GIAHITGGGF EN+PRMF + L+AEI E    VLPIFKALEKYG IKHEEM+EIFNMG+G
Sbjct: 244 GIAHITGGGFHENLPRMFGNSLSAEIVEGSWDVLPIFKALEKYGSIKHEEMYEIFNMGIG 303

Query: 303 LMLDVNPENVDRVKELLDEPVYEIGRIIKKADDSVVIK 340
           +++ V PEN   +K+ L+   +EIG+++ + +  VVIK
Sbjct: 304 MVIAVAPENAAALKKELN--AFEIGQMVNRQEAPVVIK 339
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 969> which encodes the amino acid sequence <SEQ ID 970>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3236(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 = 321/340 (94%), Positives = 332/340 (97%)

Query: 1   MSEKNAYAQSGVDVEAGYEVVERIKKHVARTERAGVMGALGGFGGMFDLSQTGVKEPVLI 60
           MSEKNAYA+SGVDVEAGYEVVERIKKHVARTERAGVMGALGGFGGMFDLS+TGVKEPVL+
Sbjct: 1   MSEKNAYAKSGVDVEAGYEVVERIKKHVARTERAGVMGALGGFGGMFDLSKTGVKEPVLV 60

Query: 61  SGTDGVGTKLMLAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLYFLDYVATGKNEPAKLEQ 120
           SGTDGVGTKLMLAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLYFLDY+ATGKN P KLE+
Sbjct: 61  SGTDGVGTKLMLAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLYFLDYIATGKNNPVKLEE 120

Query: 121 VVAGVAEGCVQASAALIGGETAEMPGMYGEDDYDLAGFAVGVAEKSQIIDGSKVKEGDIL 180
           VV+GVAEGCVQA AALIGGETAEMPGMYG+DDYDLAGFAVGVAEKSQIIDGSKVKEGDIL
Sbjct: 121 VVSGVAEGCVQAGAALIGGETAEMPGMYGQDDYDLAGFAVGVAEKSQIIDGSKVKEGDIL 180

Query: 181 LGLASSGIHSNGYSLVRRVFADYTGDEVLPELEGKQLKDVLLEPTRIYVKAALPLIKEEL 240
           LGLASSGIHSNGYSLVRRVFADYTG E+LPELEGKQLKDVLLEPTRIYVKAALPLIKEEL
Sbjct: 181 LGLASSGIHSNGYSLVRRVFADYTGKELLPELEGKQLKDVLLEPTRIYVKAALPLIKEEL 240

Query: 241 VNGIAHITGGGFIENVPRMFADDLAAEIDEDKVPVLPIFKALEKYGDIKHEEMFEIFNMG 300
           V GI HITGGGFIEN+PRMFADDLAAEIDEDKVPVLPIFKALEKYGDIKHEEMFEIFNMG
Sbjct: 241 VKGIGHITGGGFIENIPRMFADDLAAEIDEDKVPVLPIFKALEKYGDIKHEEMFEIFNMG 300

Query: 301 VGLMLDVNPENVDRVKELLDEPVYEIGRIIKKADDSVVIK 340
           VGLML V+PENV+RVKELLDEPVYEIGRIIKKAD SVVIK
Sbjct: 301 VGLMLAVSPENVNRVKELLDEPVYEIGRIIKKADASVVIK 340
```

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

Example 301

A DNA sequence (GBSx0330) was identified in *S.agalactiae* <SEQ ID 971> which encodes the amino acid sequence <SEQ ID 972>. This protein is predicted to be phosphoribosylpyrophosphate amidotransferase (purF). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1112(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:AAD12627 GB:U64311 phosphoribosylpyrophosphate amidotransferase
           [Lactococcus lactis]
 Identities = 340/470 (72%), Positives = 404/470 (85%), Gaps = 6/470 (1%)

Query: 3   YEVKSLNEECGVFGIWGYPQAAQVTYFGLHSLQHRGQEGAGIISNDNGKLYGYRNVGLLS 62
           +E K+LNEECG+FG+WG+P AA++TYFGLH+LQHRGQEGAGI+ N+NGKL  +R +GL++
Sbjct: 37  FEAKTLNEECGLFGVWGHPDAARLTYFGLHALQHRGQEGAGILVNNNGKLNRHRGLGLVT 96

Query: 63  EVFKNQSELDNLTGNAAIGHVRYATAGSADIRNIQPFLYKFHDGQFALCHNGNLTNAISS 122
           EVF+++ +L+ LTG++AIGHVRYATAGSA+I NIQPF ++FHDG   L HNGNLTNA S
Sbjct: 97  EVFRHEKDLEELTGSSAIGHVRYATAGSANINNIQPFQFEFHDGSLGLAHNGNLTNAQSL 156

Query: 123 RKELEKQGAIFNASSDTEILMHLIRRSHNPSFMGKVKEALSTVKGGFAYLLMTEDKLIAA 182
           R ELEK GAIF+++SDTEILMHLIRRSH+P FMG+VKEAL+TVKGGFAYL+MTE+ ++AA
Sbjct: 157 RCELEKSGAIFSSNSDTEILMHLIRRSHHPEFMGRVKEALNTVKGGFAYLIMTENSIVAA 216

Query: 183 LDPNAFRPLSIGQMQNGAWVISSETCAFEVVGAKWVRDVEPGEVILIDDSGIQCDRYTDE 242
           LDPN FRPLSIG+M NGA V++SETCAF+VVGA W++DV+PGE+I I+D GI  D++TD
Sbjct: 217 LDPNGFRPLSIGKMSNGALVVASETCAFDVVGATWIQDVQPGEIIEINDDGIHVDQFTDS 276

Query: 243 TQLAICSMEYVYFARPDSTIHGVNVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSAAMG 302
           T + ICSMEY+YFARPDS I GVNVHTARK  GK LAQE K DADIVIGVPNSSLSAA G
Sbjct: 277 TNMTICSMEYIYFARPDSNIAGVNVHTARKRSGKILAQEAKIDADIVIGVPNSSLSAASG 336

Query: 303 FAEESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMKLSAVSGVVKGKRVVMIDDSIVR 362
           +AEESGLP EMGL+KNQY  RTFIQPTQELREQGVRMKLSAV GVV+GKRV+M+DDSIVR
Sbjct: 337 YAEESGLPYEMGLIKNQYVARTFIQPTQELREQGVRMKLSAVRGVVEGKRVIMVDDSIVR 396

Query: 363 GTTSRRIVGLLREAGATEVHVAIASPELKYPCFYGIDIQTRRELISANHAVDEVCDIIGA 422
           GTTSRRIV LL++AGA EVHVAIASP LKYPCFYGIDIQ R ELI+A H  DE+ + IGA
Sbjct: 397 GTTSRRIVKLLKDAGAAEVHVAIASPALKYPCFYGIDIQDRDELIAATHTTDEIREAIGA 456

Query: 423 DSLTYLSIDGLIKSIGLETKAPNGGLCVAYFDGHYPTPLYDYEEEYLRSL 472
           DSLTYLS  GL+++IG      +  LC++YFDG YPTPLYDYE +YL SL
Sbjct: 457 DSLTYLSQSGLVEAIG------HDKLCLSYFDGEYPTPLYDYEADYLESL 500
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 973> which encodes the amino acid sequence <SEQ ID 974>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0610(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 = 473/484 (97%), Positives = 481/484 (98%)

Query: 1   MTYEVKSLNEECGVFGIWGYPQAAQVTYFGLHSLQHRGQEGAGIISNDNGKLYGYRNVGL 60
           MTYEVKSLNEECGVFGIWG+PQAAQVTYFGLHSLQHRGQEGAGI+SNDNGKLYGYRNVGL
Sbjct: 20  MTYEVKSLNEECGVFGIWGHPQAAQVTYFGLHSLQHRGQEGAGIVSNDNGKLYGYRNVGL 79

Query: 61  LSEVFKNQSELDNLTGNAAIGHVRYATAGSADIRNIQPFLYKFHDGQFALCHNGNLTNAI 120
           LSEVFKNQSELDNLTGNAAIGHVRYATAGSADIRNIQPFLYKFHDGQFALCHNGNLTNAI
Sbjct: 80  LSEVFKNQSELDNLTGNAAIGHVRYATAGSADIRNIQPFLYKFHDGQFALCHNGNLTNAI 139

Query: 121 SSRKELEKQGAIFNASSDTEILMHLIRRSHNPSFMGKVKEALSTVKGGFAYLLMTEDKLI 180
           S RKELEKQGAIFNASSDTEILMHLIRRSHN SFMGKVKEAL+TVKGGFAYLLMTE+KLI
Sbjct: 140 SLRKELEKQGAIFNASSDTEILMHLIRRSHNSSFMGKVKEALNTVKGGFAYLLMTENKLI 199
```

```
Query: 181 AALDPNAFRPLSIGQMQNGAWVISSETCAFEVVGAKWVRDVEPGEVILIDDSGIQCDRYT 240
           AALDPNAFRPLSIGQMQNGAWVISSETCAFEVVGAKWVRDVEPGEVILIDD GIQCDRYT
Sbjct: 200 AALDPNAFRPLSIGQMQNGAWVISSETCAFEVVGAKWVRDVEPGEVILIDDRGIQCDRYT 259

Query: 241 DETQLAICSMEYVYFARPDSTIHGVNVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSAA 300
           DETQLAICSMEYVYFARPDSTIHGVNVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSAA
Sbjct: 260 DETQLAICSMEYVYFARPDSTIHGVNVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSAA 319

Query: 301 MGFAEESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMKLSAVSGVVKGKRVVMIDDSI 360
           MGFAEESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMKLSAVSGVVKGKRVVMIDDSI
Sbjct: 320 MGFAEESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMKLSAVSGVVKGKRVVMIDDSI 379

Query: 361 VRGTTSRRIVGLLREAGATEVHVAIASPELKYPCFYGIDIQTRRELISANHAVDEVCDII 420
           VRGTTSRRIVGLLREAGA+EVHVAIASPELKYPCFYGIDIQTRRELISANH+VDEVCDII
Sbjct: 380 VRGTTSRRIVGLLREAGASEVHVAIASPELKYPCFYGIDIQTRRELISANHSVDEVCDII 439

Query: 421 GADSLTYLSIDGLIKSIGLETKAPNGGLCVAYFDGHYPTPLYDYEEEYLRSLEEKTSFYI 480
           GADSLTYLS+DGLI+SIGLETKAPNGGLCVAYFDGHYPTPLYDYEEEYLRSLEEKTSFYI
Sbjct: 440 GADSLTYLSLDGLIESIGLETKAPNGGLCVAYFDGHYPTPLYDYEEEYLRSLEEKTSFYI 499

Query: 481 QKVK 484
           QKVK
Sbjct: 500 QKVK 503
```

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

Example 302

A DNA sequence (GBSx0331) was identified in *S.agalactiae* <SEQ ID 975> which encodes the amino acid sequence <SEQ ID 976>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4797(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 303

A DNA sequence (GBSx0332) was identified in *S.agalactiae* <SEQ ID 977> which encodes the amino acid sequence <SEQ ID 978>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3489(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 304

A DNA sequence (GBSx0333) was identified in *S.agalactiae* <SEQ ID 979> which encodes the amino acid sequence <SEQ ID 980>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1690(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:CAC12194 GB:AL445066 phosphoribosylformylglycinamidine synthase
          related protein [Thermoplasma acidophilum]
 Identities = 199/746 (26%), Positives = 329/746 (43%), Gaps = 103/746 (13%)

Query: 202 ADD--FAAYKAEQGLAMEVDDLLFIQDYFKSIGRVPTETELKVLDTYWSDHCRHTTFETE 259
           ADD    A     GLA+ +D++  ++ YF+ +GR P + E+  +   WS+HC + + +
Sbjct: 11  ADDARLKAISKRLGLALSLDEMKAVRSYFERLGRDPIDAEIHAVAQSWSEHCSYKSSKYY 70

Query: 260 LKNIDFSASKFQKQLQATYDKYIAMRDELGRSEKPQTLMDMATIFGRYERANGRLDDMEV 319
           LK       K+   L+  Y   +AM D+ G
Sbjct: 71  LK-------KYLGSLKTDYT-ILAMEDDAG---------------------------- 92

Query: 320 SDEINACSVEIEVDVDGVKEPWLLMFKNETHNHPTEIEPFGGAATCIGGAIRDPLSGRSY 379
                       VD DG    +  + K E+HNHP+ +EP+GGAAT IGG +RD L   +
Sbjct: 93  -----------VVDFDG---EYAYVLKMESHNHPSAVEPYGGAATGIGGIVRDVLCMGAQ 138

Query: 380 VYQAMRISGAGDITTPIAETRAGKLPQQVISKTAAHGYSSYGNQIGLATTYVREYFHPGF 439
              +     GD+++   E   G L  + I      G   YGN+IG+       YF   +
Sbjct: 139 PVALIDSLFLGDVSSDRYE---GLLSPRYIFGGVVGGIRDYGNRIGIPNVAGSLYFDKLY 195

Query: 440 VAKRMELGAVVGAAPKENVVREKP-EAGDVVVLLGGKTGRDGVGGATGSSKVQTVESVET 498
            +  +      VG   ++ +VR K  + GDV+VL+GGKTGRDG+ G   +S     + ++
Sbjct: 196 NSNPLVNAGCVGIVRRDRIVRSKSYKPGDVLVLMGGKTGRDGIHGVNFASTTLG-KVTKS 254

Query: 499 AGAEVQKGNAIEERKIQRLFRDGNVTRLIKKSNDFGAGGVCVAIGELAD----GLEIDLD 554
           +   +Q GN I E+ + +   + N   LI+   D G GG+  A  E+       G EI LD
Sbjct: 255 SRLAIQLGNPIVEQPMIKAVLEANDAGLIRAMKDLGGGGLSSAATEMVYAGGFGAEITLD 314

Query: 555 KVPLKYQGLNGTEIAISESQERMSVVVGPSDVDAFIAACNKENIDAVVVATVTEKPNLVM 614
            + LK   ++G EI ISESQERM +   P DV+       K N+D  V+  VT    + +
Sbjct: 315 DIKLKESNMSGWEIWISESQERMLMECYPEDVEKIRQIAEKWNLDFSVIGQVTADRRIRV 374

Query: 615 TWNGETIVDLERCFLDTNGV-RVVVDAKVVDKDLTVPEARTTSAETLEADMLKVLSDLNH 673
            +   I+D++  FLD + V +     K V+K +TVP+       E L + +   ++ LN
Sbjct: 375 YYKKRKIIDMDIEFLDDSPVYQRPYRIKEVEKSVTVPQ----EPEDLNSFVRDFMARLNT 430

Query: 674 ASQKGLQTIFDSSVGRSTVNHPIGGR-YQITPTESSVQKLPVQYGVTTTASVMAQGYNPY 732
            ++  +   +D +V  ST+  P  GR  + T  +++V K P++  +     V+  G  P
Sbjct: 431 CARFNVVRQYDHTVRGSTIVTPFVGRPNKETHADATVIK-PLENSM--RGLVLTSGSRPN 487

Query: 733 IAEWSPYHGAAYAVIEATARLVATGADWSRARFSYQEYFERMDKQAERFGQPVSALLGSI 792
           +   PY G   + EA   +++TG    R              ++ E  GQ V ++
Sbjct: 488 MVSVDPYAGTLLTLAEAYKNILSTG---GRPHSVVDALNFGNPEREEIMGQFVESVRAIG 544

Query: 793 EAQIQFGLPSIGGKDSMSGTFEELTVPPTLVAFGVTTADS-RKVLSPEFKAAGENIY--- 848
           +   + GLP + G  S    + +  + PT     V   D  R+  +  K +G  IY
Sbjct: 545 DFCRKMGLPVVAGNVSFYNEYRKTDIMPTPTIMMVGLIDDVRRSRTTYMKGSGNAIYLIG 604
```

```
Query: 849 ----------------YIPGQAISEDIDFDLIKANF--SQFEAIQAQHKITAASAVKYGG 890
                           Y  G   + D+D      +F  S+ + I + H +++      GG
Sbjct: 605 EPCDNLTGSEYSRMHGYTDGFLPAPDLDELTRIRDFLSSKADMILSSHDVSS------GG 658

Query: 891 VLESLALMTFGNRIGASVEIAELDSS 916
           +  +L+ M+FG+ IG  V+I+ + ++
Sbjct: 659 LFAALSEMSFGSGIGFHVDISNVSAA 684
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 981> which encodes the amino acid sequence <SEQ ID 982>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1415(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 = 1219/1256 (97%), Positives = 1226/1256 (97%)

Query: 11   SSYFRVAPLSDLVSYMNKRIFVEKKADFGIKSASLVKELTHNLQLASLKDLRIVQVYDVF 70
            SSYF VAPLSDLVSYMNKRIFVEKKADFGIKSASLVKELTHNLQL SLK LRIVQVYDVF
Sbjct: 2    SSYFPVAPLSDLVSYMNKRIFVEKKADFGIKSASLVKELTHNLQLTSLKALRIVQVYDVF 61

Query: 71   NLAEDLLARAEKHIFSEQVTDRLLTEAEITAELDKVAFFAIEALPGQFDQRAASSQEALL 130
            NLAEDLLARAEKHIFSEQVTD LLTE EITAELDKVAFFAIEALPGQFDQRAASSQEALL
Sbjct: 62   NLAEDLLARAEKHIFSEQVTDCLLTETEITAELDKVAFFAIEALPGQFDQRAASSQEALL 121

Query: 131  LLGSDSQVKVNTAQLYLVNKDIAEAELEAVKNYLLNPVDSRFKDITLPLEVQAFSVSDKT 190
            L GSDSQVKVNTAQLYLVNKDI EAELEAVKNYLLNPVDSRFKDITLPLE QAFSVSDKT
Sbjct: 122  LFGSDSQVKVNTAQLYLVNKDITEAELEAVKNYLLNPVDSRFKDITLPLEEQAFSVSDKT 181

Query: 191  ISNLDFFETYQADDFAAYKAEQGLAMEVDDLLFIQDYFKSIGRVPTETELKVLDTYWSDH 250
            I NLDFFETYQADDFA YKAEQGLAMEVDDLLFIQ+YFKSIG VPTETELKVLDTYWSDH
Sbjct: 182  IPNLDFFETYQADDFATYKAEQGLAMEVDDLLFIQNYFKSIGCVPTETELKVLDTYWSDH 241

Query: 251  CRHTTFPETELKNIDFSASKFQKQLQATYDKYIAMRDELGRSEKPQTLMDMATIFGRYERA 310
            CRHTTFPETELKNIDFSASKFQKQLQ TYDKYIAMRDELGRSEKPQTLMDMATIFGRYERA
Sbjct: 242  CRHTTFPETELKNIDFSASKFQKQLQTTYDKYIAMRDELGRSEKPQTLMDMATIFGRYERA 301

Query: 311  NGRLDDMEVSDEINACSVEIEVDVDGVKEPWLLMFKNETHNHPTEIEPFGGAATCIGGAI 370
            NGRLDDMEVSDEINACSVEIEVDVDGVKEPWLLMFKNETHNHPTEIEPFGGAATCIGGAI
Sbjct: 302  NGRLDDMEVSDEINACSVEIEVDVDGVKEPWLLMFKNETHNHPTEIEPFGGAATCIGGAI 361

Query: 371  RDPLSGRSYVYQAMRISGAGDITTPIAETRAGKLPQQVISKTAAHGYSSYGNQIGLATTY 430
            RDPLSGRSYVYQAMRISGAGDITTPIAETRAGKLPQQVISKTAAHGYSSYGNQIGLATTY
Sbjct: 362  RDPLSGRSYVYQAMRISGAGDITTPIAETRAGKLPQQVISKTAAHGYSSYGNQIGLATTY 421

Query: 431  VREYFHPGFVAKRMELGAVVGAAPKENVVREKPEAGDVVVLLGGKTGRDGVGGATGSSKV 490
            VREYFHPGFVAKRMELGAVVGAAPKENVVREKPEAGDVV+LLGGKTGRDGVGGATGSSKV
Sbjct: 422  VREYFHPGFVAKRMELGAVVGAAPKENVVREKPEAGDVVILLGGKTGRDGVGGATGSSKV 481

Query: 491  QTVESVETAGAEVQKGNAIEERKIQRLFRDGNVTRLIKKSNDFGAGGVCVAIGELADGLE 550
            QTVESVETAGAEVQKGNAIEERKIQRLFRDGNVTRLIKKSNDFGAGGVCVAIGELADGLE
Sbjct: 482  QTVESVETAGAEVQKGNAIEERKIQRLFRDGNVTRLIKKSNDFGAGGVCVAIGELADGLE 541

Query: 551  IDLDKVPLKYQGLNGTEIAISESQERMSVVVGPSDVDAFIAACNKENIDAVVVATVTEKP 610
            IDLDKVPLKYQGLNGTEIAISESQERMSVVV P+DVDAFIAACNKENIDAVVVATVTEKP
Sbjct: 542  IDLDKVPLKYQGLNGTEIAISESQERMSVVVRPNDVDAFIAACNKENIDAVVVATVTEKP 601

Query: 611  NLVMTWNGETIVDLERCFLDTNGVRVVVDAKVVDKDLTVPEARTTSAETLEADMLKVLSD 670
            NLVMTWNGE IVDLER FLDTNGVRVVVDAKVVDKDLTVPEARTTSAETLEAD LKVLSD
Sbjct: 602  NLVMTWNGEIIVDLERRFLDTNGVRVVVDAKVVDKDLTVPEARTTSAETLEADTLKVLSD 661
```

```
Query: 671  LNHASQKGLQTIFDSSVGRSTVNHPIGGRYQITPTESSVQKLPVQYGVTTTASVMAQGYN 730
            LNHASQKGLQTIFDSSVGRSTVNHPIGGRYQITPTESSVQKLPVQ+GVTTTASVMAQGYN
Sbjct: 662  LNHASQKGLQTIFDSSVGRSTVNHPIGGRYQITPTESSVQKLPVQHGVTTTASVMAQGYN 721

Query: 731  PYIAEWSPYHGAAYAVIEATARLVATGADWSRARFSYQEYFERMDKQAERFGQPVSALLG 790
            PYIAEWSPYHGAAYAVIEATARLVATGADWSRARFSYQEYFERMDKQAERFGQPVSALLG
Sbjct: 722  PYIAEWSPYHGAAYAVIEATARLVATGADWSRARFSYQEYFERMDKQAERFGQPVSALLG 781

Query: 791  SIEAQIQFGLPSIGGKDSMSGTFEELTVPPTLVAFGVTTADSRKVLSPEFKAAGENIYYI 850
            SIEAQIQ GLPSIGGKDSMSGTFE+LTVPPTLVAFGVTTADSRKVLSPEFKAAGENIYYI
Sbjct: 782  SIEAQIQLGLPSIGGKDSMSGTFEDLTVPPTLVAFGVTTADSRKVLSPEFKAAGENIYYI 841

Query: 851  PGQAISEDIDFDLIKANFSQFEAIQAQHKITAASAVKYGGVLESLALMTFGNRIGASVEI 910
            PGQAISEDIDFDLIK NFSQFEAIQAQHKITAASA KYGGVLESLALMTFGNRIGASVEI
Sbjct: 842  PGQAISEDIDFDLIKDNFSQFEAIQAQHKITAASAAKYGGVLESLALMTFGNRIGASVEI 901

Query: 911  AELDSSLTAQLGGFVFTSVEEIADVVKIGQTQADFTVTVNGNDLAGASLLSAFEGKLEEV 970
            AELDSSLTAQLGGFVFTS EEIAD VKIGQTQADFTVTVNGNDLAGASLL+AFEGKLEEV
Sbjct: 902  AELDSSLTAQLGGFVFTSAEEIADAVKIGQTQADFTVTVNGNDLAGASLLAAFEGKLEEV 961

Query: 971  YPTEFEQVDAIEEVPAVVSDVVIKAKEIIEKPVVYIPVFPGTNSEYDSAKAFEQVGASVN 1030
            YPTEFEQ D +EEVPAVVSD VIKAKE IEKPVVYIPVFPGTNSEYDSAKAFEQVGASVN
Sbjct: 962  YPTEFEQTDVLEEVPAVVSDTVIKAKETIEKPVVYIPVFPGTNSEYDSAKAFEQVGASVN 1021

Query: 1031 LVPFVTLNEAAIAESVDTMVANIAKANIIFFAGGFSAADEPDGSAKFIVNILLNEKVRAA 1090
            LVPFVTLNE AIAESVDTMVANIAKANIIFFAGGFSAADEPDGSAKFIVNILLNEKVRAA
Sbjct: 1022 LVPFVTLNEVAIAESVDTMVANIAKANIIFFAGGFSAADEPDGSAKFIVNILLNEKVRAA 1081

Query: 1091 IDSFIEKGGLIIGICNGFQALVKSGLLPYGNFEEAGETSPTLFYNDANQHVAKMVETRIA 1150
            IDSFIEKGGLIIGICNGFQALVKSGLLPYGNFEEAGETSPTLFYNDANQHVAKMVETRIA
Sbjct: 1082 IDSFIEKGGLIIGICNGFQALVKSGLLPYGNFEEAGETSPTLFYNDANQHVAKMVETRIA 1141

Query: 1151 NTNSPWLAGVEVGDIHVIPVSHGEGKFVVSASEFAELRDNGQIWSQYVDFDGQPSMDSKY 1210
            NTNSPWLAGVEVGDIH IPVSHGEGK VVSASEFAELRDNGQIWSQYVDFDGQPSMDSKY
Sbjct: 1142 NTNSPWLAGVEVGDIHAIPVSHGEGKLVVSASEFAELRDNGQIWSQYVDFDGQPSMDSKY 1201

Query: 1211 NPNGSVNAIEGITSKNGQIIGKMGHSERWEDGLFQNIPGNKDQKLFESAVKYFTGK 1266
            NPNGSVNAIEGITSKNGQIIGKMGHSERWEDGLFQNIPGNKDQ LF SAVKYFTGK
Sbjct: 1202 NPNGSVNAIEGITSKNGQIIGKMGHSERWEDGLFQNIPGNKDQILFASAVKYFTGK 1257
```

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

Example 305

A DNA sequence (GBSx0334) was identified in *S.agalactiae* <SEQ ID 983> which encodes the amino acid sequence <SEQ ID 984>. This protein is predicted to be phosphoribosylaminoimidazole-succinocarboxamide synthase (purC). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4783(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:AAA03540 GB:L15190 SAICAR synthetase [Streptococcus pneumoniae]
 Identities = 183/231 (79%), Positives = 203/231 (87%)

Query: 1   MTNQLIYTGKAKDIYSTKDENVIRTVYKDQATMLNGARKETIDGKGALNNQISSLIFEKL 60
            M+ QLIY+GKAKDIY+T+DEN+I + YKDQAT  NG +KE I GKG LNNQISS IFEKL
Sbjct: 1   MSKQLIYSGKAKDIYTTEDENLIISTYKDQATAFNGVKKEQIAGKGVLNNQISSFIFEKL 60
```

```
Query: 61  NMAGVVTHYIEQISKNEQLNKKVDIIPLEVVLRNVTAGSFSKRFGVEEGHVLETPIVEFY 120
           N AGV TH++E++S  EQLNKKV IIPLEVVLRN TAGSFSKRFGV+EG  LETPIVEFY
Sbjct: 61  NAAGVATHFVEKLSDTEQLNKKVKIIPLEVVLRNYTAGSFSKRFGVDEGIALETPIVEFY 120

Query: 121 YKNDNLNDPFINDEHVKFLGIVNDEEIAYLKGETRHINELLKDWFAQIGLNLIDFKLEFG 180
           YKND+L+DPFINDEHVKFL I +D++IAYLK E R INELLK WFA+IGL LIDFKLEFG
Sbjct: 121 YKNDDLDDPFINDEHVKFLQIADDQQIAYLKEEARRINELLKVWFAEIGLKLIDFKLEFG 180

Query: 181 FDKDGKIILADEFSPDNCRLWDADGNHMDKDVFRRDLGSLTDVYQVVLEKL 231
           FDKDGKIILADEFSPDNCRLWDADGNHMDKDVFRR LG LTDVY++V EKL
Sbjct: 181 FDKDGKIILADEFSPDNCRLWDADGNHMDKDVFRRGLGELTDVYEIVWEKL 231
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 985> which encodes the amino acid sequence <SEQ ID 986>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3935(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 221/234 (94%), Positives = 228/234 (96%)

Query: 1   MTNQLIYTGKAKDIYSTKDENVIRTVYKDQATMLNGARKETIDGKGALNNQISSLIFEKL 60
           +TNQLIY GKAKDIYSTKDENVIRTVYKDQATMLNGARKETIDGKGALNNQISSLIFEKL
Sbjct: 11  VTNQLIYKGKAKDIYSTKDENVIRTVYKDQATMLNGARKETIDGKGALNNQISSLIFEKL 70

Query: 61  NMAGVVTHYIEQISKNEQLNKKVDIIPLEVVLRNVTAGSFSKRFGVEEGHVLETPIVEFY 120
           N AGVVTHYIEQISKNEQLNKKVDIIPLEVVLRNVTAGSFSKRFGVEEGHVLETPIVEFY
Sbjct: 71  NKAGVVTHYIEQISKNEQLNKKVDIIPLEVVLRNVTAGSFSKRFGVEEGHVLETPIVEFY 130

Query: 121 YKNDNLNDPFINDEHVKFLGIVNDEEIAYLKGETRHINELLKDWFAQIGLNLIDFKLEFG 180
           YKND+L+DPFINDEHVKFLGIVNDEEIAYLKGETR INELLK WFAQIGLNLIDFKLEFG
Sbjct: 131 YKNDDLDDPFINDEHVKFLGIVNDEEIAYLKGETRRINELLKGWFAQIGLNLIDFKLEFG 190

Query: 181 FDKDGKIILADEFSPDNCRLWDADGNHMDKDVFRRDLGSLTDVYQVVLEKLIAL 234
           FD++G IILADEFSPDNCRLWD +GNHMDKDVFRRDLG+LTDVYQVVLEKLIAL
Sbjct: 191 FDQEGTIILADEFSPDNCRLWDKNGNHMDKDVFRRDLGNLTDVYQVVLEKLIAL 244
```

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

Example 306

A DNA sequence (GBSx0335) was identified in *S.agalactiae* <SEQ ID 987> which encodes the amino acid sequence <SEQ ID 988>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2779(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 9457> which encodes amino acid sequence <SEQ ID 9458> was also identified.

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

```
>GP:AAC35700 GB:AF041468 acyl carrier protein [Guillardia theta]
 Identities = 27/75 (36%), Positives = 52/75 (69%)

Query: 12 MSRDEVFEKMLELLRQQLGDPQLDITPESSLHDDLAIDSIALTEFIINLEDVFHLEIPDE 71
          M+  E+FEK+  ++ +QLG  +  +T +++  +DL  DS+   E ++ +E+ F++EIPD+
Sbjct: 1  MNEQEIFEKVQTIISEQLGVDKSQVTKDANFANDLGADSLDTVELVMAIEEAFNIEIPDD 60

Query: 72 AVEHMSSVQQLLDYI 86
          A E +S++QQ +D+I
Sbjct: 61 AAEQISNLQQAVDFI 75
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 989> which encodes the amino acid sequence <SEQ ID 990>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1917(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 = 36/77 (46%), Positives = 57/77 (73%)

Query: 12 MSRDEVFEKMLELLRQQLGDPQLDITPESSLHDDLAIDSIALTEFIINLEDVFHLEIPDE 71
          M+R E+FE+++ L+++Q     + IT ++ L +DLA+DSI L EFIIN+ED FH+ IPDE
Sbjct: 1  MTRQEIFERLINLIQKQRSYLSVAITEQTHLKNDLAVDSIELVEFIINVEDEFHIAIPDE 60

Query: 72 AVEHMSSVQQLLDYIIE 88
           VE M  ++ +LDY+++
Sbjct: 61 DVEDMVFMRDILDYLVQ 77
```

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

Example 307

A DNA sequence (GBSx0336) was identified in *S.agalactiae* <SEQ ID 991> which encodes the amino acid sequence <SEQ ID 992>. This protein is predicted to be fatty acid/phospholipid synthesis protein (plsX). Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.64   Transmembrane  101 - 117 ( 101 - 117)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9455> which encodes amino acid sequence <SEQ ID 9456> was also identified.

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

```
>GP:CAB13462 GB:Z99112 alternate gene name: ylpD [Bacillus subtilis]
 Identities = 174/329 (52%), Positives = 238/329 (71%), Gaps = 2/329 (0%)

Query: 8   KIAIDAMGGDYAPKAIVEGVNQAISDFSDIEVQLYGDQKKIEKYLTVT-ERVSIIHTEEK 66
          +IA+DAMGGD+APKA++GV + I  F D+ + L GD+  IE +LT T +R++++H +E
Sbjct: 2   RIAVDAMGGDHAPKAVIDGVIKGIEAFDDLHITLVGDKTTIESHLTTTSDRITVLHADEV 61
```

```
Query: 67  INSDDEPAKAVRRKKQSSMVLGAKAVKDGVAQAFISAGNTGALLAAGLFVVGRIKGVDRP 126
           I   DEP +AVRRKK SSMVL A+ V +  A A ISAGNTGAL+ AGLF+VGRIKG+DRP
Sbjct: 62  IEPTDEPVRAVRRKKNSSMVLMAQEVAENRADACISAGNTGALMTAGLFIVGRIKGIDRP 121

Query: 127 GLMSTMPTLDGVGFDMLDLGANAENTASHLHQYAILGSFYAKNVRGIEVPRVGLLNNGTE 186
            L  T+PT+ G GF +LD+GAN +     HL QYAI+GS Y++ VRG+  PRVGLLN GTE
Sbjct: 122 ALAPTLPTVSGDGFLLLDVGANVDAKPEHLVQYAIMGSVYSQQVRGVTSPRVGLLNVGTE 181

Query: 187 ETKGDSLHKEAYELLAAEPSINFIGNIEARDLMSSVADVVVTDGFTGNAVLKTMEGTAMS 246
           + KG+ L K+ +++L     +INFIGN+EARDL+  VADVVVTDGFTGN  LKT+EG+A+S
Sbjct: 182 DKKGNELTKQTFQILKETANINFIGNVEARDLLDDVADVVVTDGFTGNVTLKTLEGSALS 241

Query: 247 IMGSLKSSIKSGGVKAKLGALLLKDSLYQLKDSMDYSSAGGAVLFGLKAPIVKCHGSSDS 306
           I   ++  + +  +KL A +LK  L ++K  M+YS+ GGA LFGLKAP++K HGSSDS
Sbjct: 242 IFKMMR-DVMTSTLTSKLAAAVLKPKLKEMKMKMEYSNYGGASLFGLKAPVIKAHGSSDS 300

Query: 307 KAVYSTLKQVRTMLETQVVDQLVDAFTDE 335
            AV+  ++Q R M+   V   + +  +E
Sbjct: 301 NAVFHAIRQAREMVSQNVAALIQEEVKEE 329
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 993> which encodes the amino acid sequence <SEQ ID 994>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.07   Transmembrane  121 - 137 ( 120 - 138)

----- 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>
```

A related sequence was also identified in GAS <SEQ ID 9127> which encodes the amino acid sequence <SEQ ID 9128>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 16
>>> Seems to have no N-terminal signal sequence
     INTEGRAL    Likelihood = -2.07   Transmembrane   95 - 111 (  94 - 112)

----- Final Results -----
               bacterial membrane --- Certainty= 0.183(Affirmative) < succ>
                 bacterial outside --- Certainty= 0.000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

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

```
Identities = 254/330 (76%), Positives = 290/330 (86%)

Query: 6   MKKIAIDAMGGDYAPKAIVEGVNQAISDFSDIEVQLYGDQKKIEKYLTVTERVSIIHTEE 65
           MK+IAIDAMGGD APKAIVEGVNQAI  FSDIE+QLYGDQ KI  YL  ++RV+IIHT+E
Sbjct: 27  MKRIAIDAMGGDNAPKAIVEGVNQAIEAFSDIEIQLYGDQTKINSYLIQSDRVAIIHTDE 86

Query: 66  KINSDDEPAKAVRRKKQSSMVLGAKAVKDGVAQAFISAGNTGALLAAGLFVVGRIKGVDR 125
           KI SDDEPAKAVRRKK++SMVL AKAVK+G A A ISAGNTGALLA GLFVVGRIKGVDR
Sbjct: 87  KIMSDDEPAKAVRRKKKASMVLAAKAVKEGKADAIISAGNTGALLAVGLFVVGRIKGVDR 146

Query: 126 PGLMSTMPTLDGVGFDMLDLGANAENTASHLHQYAILGSFYAKNVRGIEVPRVGLLNNGT 185
           PGL+ST+PT+ G+GFDMLDLGANAENTA HLHQYAILGSFYAKNVRGI  PRVGLLNNGT
Sbjct: 147 PGLLSTIPTVTGLGFDMLDLGANAENTAKHLHQYAILGSFYAKNVRGIANPRVGLLNNGT 206

Query: 186 EETKGDSLHKEAYELLAAEPSINFIGNIEARDLMSSVADVVVTDGFTGNAVLKTMEGTAM 245
           EETKGD L K  YELL A+ +I+F+GN+EAR+LMS VADV+V+DGFTGNAVLK++EGTA+
Sbjct: 207 EETKGDPLRKATYELLTADNTISFVGNVEARELMSGVADVIVSDGFTGNAVLKSIEGTAI 266

Query: 246 SIMGSLKSSIKSGGVKAKLGALLLKDSLYQLKDSMDYSSAGGAVLFGLKAPIVKCHGSSD 305
```

```
           SIMG LK  I SGG+K K+GA LLK SLY++K ++DYSSAGGAVLFGLKAP+VK HGSSD
Sbjct: 267 SIMGQLKQIINSGGIKTKIGASLLKSSLYEMKKTLDYSSAGGAVLFGLKAPVVKSHGSSD 326

Query: 306 SKAVYSTLKQVRTMLETQVVDQLVDAFTDE 335
           KA++ST+KQVRTML+T VV QLV+ F  E
Sbjct: 327 VKAIFSTIKQVRTMLDTNVVGQLVEEFAKE 356
```

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

Example 308

A DNA sequence (GBSx0337) was identified in *S.agalactiae* <SEQ ID 995> which encodes the amino acid sequence <SEQ ID 996>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4668(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 309

A DNA sequence (GBSx0338) was identified in *S.agalactiae* <SEQ ID 997> which encodes the amino acid sequence <SEQ ID 998>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.84   Transmembrane   61 -  77 (  55 -  82)
   INTEGRAL    Likelihood =-10.14   Transmembrane   26 -  42 (  19 -  51)
   INTEGRAL    Likelihood = -9.77   Transmembrane  192 - 208 ( 186 - 211)
   INTEGRAL    Likelihood = -5.79   Transmembrane  267 - 283 ( 262 - 286)
   INTEGRAL    Likelihood = -3.77   Transmembrane  100 - 116 (  99 - 116)

----- Final Results -----
            bacterial membrane --- Certainty=0.6137(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 9453> which encodes amino acid sequence <SEQ ID 9454> was also identified.

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

```
>GP:CAA22372 GB:AL034446 putative transmembrane protein
          [Streptomyces coelicolor A3(2)]
 Identities = 47/154 (30%), Positives = 69/154 (44%), Gaps = 12/154 (7%)

Query: 120 SGFVEISSSNSFSFGPFFFLFLAYFIQSLTEEILFRGYVMTTVTKFKGSFAGVLCNSMLF 179
           SG+ E+    S          F+A    + TEE++FRG +   + +  G++  +    ++F
Sbjct: 118 SGYYEVDGLGSVQGAIGLVGFMA--AAAATEEVVFRGVLFRIIEEHIGTYLALGLTGLVF 175

Query: 180 SFIHFRN-----YGITAIALFNLFLLGIIFSILFNMTKNILFVTGVHTTWNFTMGCVLGN 234
             +H  N     +G  AIA+   F+L   ++     T+N+    GVH  WNF  G V
```

```
Sbjct: 176 GLMHLLNEDATLWGALAIAIEAGFMLAAAYAA----TRNLWLTIGVHFGWNFAAGGVFST 231

Query: 235 KVSGGDSPVSLFRITENSSFALWNGGDFGFEGGV 268
           VSG      L   T  S   L  GGDFG EG V
Sbjct: 232 VVSGNGDSEGLLDAT-MSGPKLLTGGDFGPEGSV 264
```

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

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

Example 310

A DNA sequence (GBSx0339) was identified in *S.agalactiae* <SEQ ID 999> which encodes the amino acid sequence <SEQ ID 1000>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2665(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 9451> which encodes amino acid sequence <SEQ ID 9452> was also identified.

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

```
>GP:BAB05088 GB:AP001511 unknown conserved protein [Bacillus halodurans]
 Identities = 81/242 (33%), Positives = 124/242 (50%), Gaps = 3/242 (1%)

Query: 8   GLVLYNRNYREDDKLVKIFTETEGKRMFFVKHAS--KSKFNAVLQPLTIAHFILKINDNG 65
           G+V+   +Y E +K+V +FT   GK     + A   KS+  AV Q  T    + + N  G
Sbjct: 7   GIVIRTVDYGESNKIVTVFTREYGKIALMARGAKRPKSRLTAVTQLFTYGMMMFQKNA-G 65

Query: 66  LSYIDDYKEVLAFQETNSDLFKLSYASYITSLADVAISDNVADAQLFIFLKKTLELIEDG 125
           L  +   + + +F+E  +DLF+ SY SY+T L +    D   +  LF  L +T+  + +G
Sbjct: 66  LGTLTQGEIIQSFREVRNDLFRASYVSYVTDLTNKLTEDEKRNPYLFELLYQTIHYMNEG 125

Query: 126 LDYEILTNIFEVQLLERFGVALNFHDCVFCHRVGLPFDFSHKYSGLLCPNHYYKDERRNH 185
           +D ++LT IFEV++    G+      CV C    +P  FS K +G LC     KD
Sbjct: 126 MDPDVLTRIFEVKMFTVAGIKPELDQCVSCRSTDVPVGFSIKEAGFLCKRCIEKDPHAYK 185

Query: 186 LDPNMLYLINRFQSIQFDDLQTISVKPEMKLKIRQFLDMIYDEYVGIHLKSKKFIDDLSSWG 247
           +   + L+  F      L TIS+KPE K  ++  +   YDEY G+HLKS++F+D L S G
Sbjct: 186 ITAQVAKLLRLFYHFDLQRLGTISLKPETKATLKTIIHQYYDEYSGLHLKSRRFLDQLESMG 247
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1001> which encodes the amino acid sequence <SEQ ID 1002>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1566(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
Identities = 159/251 (63%), Positives = 210/251 (83%)

Query: 1   MRVSQTYGLVLYNRNYREDDKLVKIFTETEGKRMFFVKHASKSKFNAVLQPLTIAHFILK 60
```

```
            M+++++ G+VL+NRNYREDDKLVKIFTE  GK+MFFVKH S+SK ++++QPLTIA FI K
Sbjct: 1    MQLTESLGIVLFNRNYREDDKLVKIFTEVAGKQMFFVKHISRSKMSSIIQPLTIADFIFK 60

Query: 61   INDNGLSYIDDYKEVLAFQETNSDLFKLSYASYITSLADVAISDNVADAQLFIFLKKTLE 120
            +ND GLSY+ DY  V  ++  N+D+F+L+YASY+ +LAD AI+DN +D+ LF FLKKTL+
Sbjct: 61   LNDTGLSYVVDYSNVNTYRYINNDIFRLAYASYVLALADAAIADNESDSHLFTFLKKTLD 120

Query: 121  LIEDGLDYEILTNIFEVQLLERFGVALNFHDCVFCHRVGLPFDFSHKYSGLLCPNHYYKD 180
            L+E+GLDYEILTNIFE+Q+L+RFG++LNFH+C  CHR  LP DFSH++S +LC  HYYKD
Sbjct: 121  LMEEGLDYEILTNIFEIQILDRFGISLNFHECAICHRTDLPLDFSHRFSAVLCSEHYYKD 180

Query: 181  ERRNHLDPNMLYLINRFQSIQFDDLQTISVKPEMKLKIRQFLDMIYDEYVGIHLKSKKFI 240
             RRNHLDPN++YL++RFQ I FDDL+TIS+  ++K K+RQF+D +Y +YVGI LKSK FI
Sbjct: 181  NRRNHLDPNVIYLLSRFQKITFDDLRTISLNKDIKKKLRQFIDELYHDYVGIKLKSKTFI 240

Query: 241  DDLSSWGSIMK 251
            D+L  WG IMK
Sbjct: 241  DNLVKWGDIMK 251
```

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

Example 311

A DNA sequence (GBSx0340) was identified in *S.agalactiae* <SEQ ID 1003> which encodes the amino acid sequence <SEQ ID 1004>. This protein is predicted to be aromatic amino acid aminotransferase (patA). Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.13   Transmembrane  141 - 157 ( 140 - 159)

----- Final Results -----
               bacterial membrane --- Certainty=0.2253(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 9449> which encodes amino acid sequence <SEQ ID 9450> was also identified.

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

```
>GP:AAF06954 GB:AF146529 aromatic amino acid aminotransferase
            [Lactococcus lactis subsp. cremoris]
 Identities = 261/391 (66%), Positives = 323/391 (81%)

Query: 38   MTLEKRFNKYLDRIEVSLIRQFDQSISDIPGMVKLTLGEPDFTTPDHVKEAAKSAIDANQ 97
            M L K+FN  LD+IE+SLIRQFDQ +S IP ++KLTLGEPDF TP+HVK+A  +AI+ NQ
Sbjct: 1    MDLLKKFNPNLDKIEISLIRQFDQQVSSIPDIIKLTLGEPDFYTPEHVKQAGIAAIENNQ 60

Query: 98   SYYTGMSGLLALRQAAADFAKDKYNLTYNPDCEILVTIGATEALSASLIAILEAGDVVLL 157
            S+YTGM+GLL LRQAA++F   KY L+Y  + EILVT+G TEA+S+ L++IL AGD VL+
Sbjct: 61   SHYTGMAGLLELRQAASEFLLKKYGLSYAAEDEILVTVGVTEAISSVLLSILVAGDEVLI 120

Query: 158  PAPAYPGYEPIVNLVGADIVEIDTRENDFRLTPEMLETAIIQQGEKLKAVLLNYPTNPTG 217
            PAPAYPGYEP++ L G  +VEIDTR NDF LTPEML+ AII++  K+KAV+LNYP NPTG
Sbjct: 121  PAPAYPGYEPLITLAGGSLVEIDTRANDFVLTPEMLDQAIIEREGKVKAVILNYPANPTG 180

Query: 218  ITYSRQEIAALAEVLKKYDIFVISDEVYSELTYTGQQHVSIAEYLPNQTILINGLSKSHA 277
            +TY+R++I  LAEVLKK+++FVI+DEVYSEL YT Q HVSIAEY P QTI++NGLSKSHA
Sbjct: 181  VTYNREQIKDLAEVLKKHEVFVIADEVYSELNYTDQPHVSIAEYAPEQTIVLNGLSKSHA 240

Query: 278  MTGWRVGLVYAPEAFIAQIIKSHQYMVTAASTISQFAGVEALSVGKNDTLPMRQGYIKRR 337
            MTGWR+GL++A     +AQIIK+HQY+VT+AST SQFA +EAL  G +D LPM++ Y+KRR
```

```
Sbjct: 241 MTGWRIGLIFAARELVAQIIKTHQYLVTSASTQSQFAAIEALKNGADDALPMKKEYLKRR 300

Query: 338 DYIIDKMSKLGFKIIKPSGAFYIFAKIPDSYPQDSFKFCQDFAYQQAVAIIPGVAFGKYG 397
           DYII+KMS LGFKII+P GAFYIFAKIP    QDSFKF  DFA + AVAIIPG+AFG+YG
Sbjct: 301 DYIIEKMSALGFKIIEPDGAFYIFAKIPADLEQDSFKFAVDFAKENAVAIIPGIAFGQYG 360

Query: 398 EGYIRLSYAASMEVIETAMARLKVFMESYEG 428
           EG++RLSYAASM+VIE AMARL  ++    G
Sbjct: 361 EGFVRLSYAASMDVIEQAMARLTDYVTKKRG 391
```

There is also homology to SEQ ID 1006.

SEQ ID 1004 (GBS332) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 3; MW 50.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 4; MW 76kDa).

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

Example 312

A DNA sequence (GBSx0341) was identified in *S.agalactiae* <SEQ ID 1007> which encodes the amino acid sequence <SEQ ID 1008>. This protein is predicted to be ribose-phosphate pyrophosphokinase (prsA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3118(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 9447> which encodes amino acid sequence <SEQ ID 9448> was also identified.

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

```
>GP:AAA62181 GB:M92842 prs [Listeria monocytogenes]
 Identities = 209/312 (66%), Positives = 266/312 (84%), Gaps = 3/312 (0%)

Query: 10  LKLFALSSNKELAKKVSQTIGIPLGQSTVRQFSDGEIQVNIEESIRGHHVFILQSTSSPV 69
           LK+F+L+SN+ELA+++++ +GI LG+S+V  FSDGEIQ+NIEESIRG HV+++QSTS+PV
Sbjct: 10  LKIFSLNSNRELAEEIAKEVGIELGKSSVTHFSDGEIQINIEESIRGCHVYVIQSTSNPV 69

Query: 70  NDNLMEILIMVDALKRASAESVSVVMPYYGYARQDRKARSREPITSKLVANMLEVAGVDR 129
           N NLME+LIM+DALKRASA ++++VMPYYGYARQDRKARSREPIT+KLVAN++E AG  R
Sbjct: 70  NQNLMELLIMIDALKRASAATINIVMPYYGYARQDRKARSREPITAKLVANLIETAGATR 129

Query: 130 LLTVDLHAAQIQGFFDIPVDHLMGAPLIADYFDRQGLVGDDVVVVSPDHGGVTRARKLAQ 189
           ++T+D+HA QIQGFFDIP+DHL    L++DYF  + L GDD+VVVSPDHGGVTRARK+A
Sbjct: 130 MITLDMHAPQIQGFFDIPIDHLNAVRLLSDYFSERHL-GDDLVVVSPDHGGVTRARKMAD 188

Query: 190 CLKTPIAIIDKRRSVTKMNTSEVMNIIGNIKGKKCILIDDMIDTAGTICHAADALAEAGA 249
            LK PIAIIDKRR   + N +EVMNI+GN++GK CI+IDD+IDTAGTI  AA AL EAGA
Sbjct: 189 RLKAPIAIIDKRR--PRPNVAEVMNIVGNVEGKVCIIIDDIIDTAGTITLAAKALREAGA 246

Query: 250 TAVYASCTHPVLSGPALDNIQNSAIEKLIVLDTIYLPEERLIDKIEQISIAELIGEAIIR 309
           T VYA C+HPVLSGPA+  I+ S IEKL+V ++I LPEE+ IDK+EQ+S+A L+GEAI+R
Sbjct: 247 TKVYACCSHPVLSGPAMKRIEESPIEKLVVTNSIALPEEKWIDKMEQLSVAALLGEAIVR 306

Query: 310 IHEKRPLSPLFE 321
           +HE   +S LFE
```

```
Sbjct: 307 VHENASVSSLFE 318
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1009> which encodes the amino acid sequence <SEQ ID 1010>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2685(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 298/322 (92%), Positives = 311/322 (96%)

Query: 1   MEEIMSYSNLKLFALSSNKELAKKVSQTIGIPLGQSTVRQFSDGEIQVNIEESIRGHHVF 60
           +EE MSYS+LKLFALSSNKELA+KV+  +GI LG+STVRQFSDGEIQVNIEESIRGHHVF
Sbjct: 1   LEEKMSYSDLKLFALSSNKELAEKVASAMGIQLGKSTVRQFSDGEIQVNIEESIRGHHVF 60

Query: 61  ILQSTSSPVNDNLMEILIMVDALKRASAESVSVVMPYYGYARQDRKARSREPITSKLVAN 120
           ILQSTSSPVNDNLMEILIMVDALKRASAE +SVVMPYYGYARQDRKARSREPITSKLVAN
Sbjct: 61  ILQSTSSPVNDNLMEILIMVDALKRASAEKISVVMPYYGYARQDRKARSREPITSKLVAN 120

Query: 121 MLEVAGVDRLLTVDLHAAQIQGFFDIPVDHLMGAPLIADYFDRQGLVGDDVVVVSPDHGG 180
           MLEVAGVDRLLTVDLHAAQIQGFFDIPVDHLMGAPLIADYFDR GLVG+DVVVVSPDHGG
Sbjct: 121 MLEVAGVDRLLTVDLHAAQIQGFFDIPVDHLMGAPLIADYFDRHGLVGEDVVVVSPDHGG 180

Query: 181 VTRARKLAQCLKTPIAIIDKRRSVTKMNTSEVMNIIGNIKGKKCILIDDMIDTAGTICHA 240
           VTRARKLAQ L+TPIAIIDKRRSV KMNTSEVMNIIGN+ GKKCILIDDMIDTAGTICHA
Sbjct: 181 VTRARKLAQFLQTPIAIIDKRRSVDKMNTSEVMNIIGNVSGKKCILIDDMIDTAGTICHA 240

Query: 241 ADALAEAGATAVYASCTHPVLSGPALDNIQNSAIEKLIVLDTIYLPEERLIDKIEQISIA 300
           ADALAEAGATAVYASCTHPVLSGPALDNIQ SAIEKLIVLDTIYLP+ERLIDKIEQISIA
Sbjct: 241 ADALAEAGATAVYASCTHPVLSGPALDNIQRSAIEKLIVLDTIYLPKERLIDKIEQISIA 300

Query: 301 ELIGEAIIRIHEKRPLSPLFEM 322
           +L+ EAIIRIHEKRPLSPLFEM
Sbjct: 301 DLVAEAIIRIHEKRPLSPLFEM 322
```

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

Example 313

A DNA sequence (GBSx0342) was identified in *S.agalactiae* <SEQ ID 1011> which encodes the amino acid sequence <SEQ ID 1012>. This protein is predicted to be a secreted protein. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3751(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 9277> which encodes amino acid sequence <SEQ ID 9278> was also identified.

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

```
>GP:AAD00288 GB:U78607 putative secreted protein [Streptococcus mutans]
 Identities = 111/157 (70%), Positives = 130/157 (82%), Gaps = 1/157 (0%)

Query: 1   MTAIKGQVGALESQQSELEAQNAQLEAVSQQLGQEIQTLSNKIVARNESLKKQVRSAQKG 60
           +  I+GQV AL++QQ+EL+A+N +LEA S  LGQ+IQTLS+KIVARNESLK+Q RSAQK
Sbjct: 55  LITIQGQVSALQTQQAELQAENQRLEAQSATLGQQIQTLSSKIVARNESLKQQARSAQKS 114

Query: 61  NL-TNYINTILNSKSVSDAVNRVVAIREVVSANEKMLAQQEADKAALEAKQIENQNAINT 119
           N  T+YIN I+NSKSVSDA+NRV AIREVVSANEKML QQE DKAA+E KQ ENQ AINT
Sbjct: 115 NAATSYINAIINSKSVSDAINRVSAIREVVSANEKMLQQQEQDKAAVEQKQQENQAAINT 174

Query: 120 VAANKQAIENNKAALATQRAQLEAAQLELSAQLTTVQ 156
           VAAN++ I  N  AL TQ+AQLEAAQL L A+LTT Q
Sbjct: 175 VAANQETIAQNTNALNTQQAQLEAAQLNLQAELTTAQ 211
```

There is also homology to SEQ ID 1014.

A related GBS gene <SEQ ID 8543> and protein <SEQ ID 8544> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:       8.29
GvH: Signal Score (-7.5): 0.8
     Possible site: 49
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   6.74 threshold:  0.0
   PERIPHERAL  Likelihood =  6.74     400
 modified ALOM score:  -1.85

*** 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)
```

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

```
32.8/56.3% over 439aa

                                                                    Lactococcus lactis
  GP|512521| usp 45 Insert characterized
  PIR|JN0097|JN0097 secreted 45K protein precursor - Insert characterized

ORF00094(301 - 1563 of 1941)
GP|512521|emb|CAA01320.1||A17083(1    -    440    of    461)    usp    45    (Lactococcus
lactis)PIR|JN0097|JN0097 secrete
d 45K protein precursor - Lactococcus lactis
%Match = 16.5
%Identity = 32.8  %Similarity = 56.3
Matches = 141  Mismatches = 178  Conservative Sub.s = 101

93       123       153       183       213       243       273       303
RKYYNFKSNYTLFLFLF*FHYGVIILIE*IEEGYRFLDLIMVHLEIVDFKYKCNNDVI*FREFFGKIFNVLS*RSSLIKM
                                                                          |
                                                                          M

333                387       417       447       477       507       537
KKRILSAVLVSGVTLGTAA--VTVNADDFDSKIAATDSVINTLSGQQAAAQNQVTAIKGQVGALESQQSELEAQNAQLEA
||:|:||:|:| | |  ||    | ||  :| ||  |: |::    :| || || ::: :| :|: :|:  :|| |::|:
KKKIISAILMSTVILSAAAPLSGVYAD-TNSDIAKQDATISSAQSAKAQAQAQVDSLQSKVDSLQQKQTSTKAQIAKIES
                  20        30        40        50        60        70        80

567       597       627       654       684       714       744       774
VSQQLGQEIQTLSNKIVARNESLKKQVRSAQ-KGNLTNYINTILNSKSVSDAVNRVVAIREVVSANEKMLAQQEADKAAL
 :: |  :| ||:  |  | ::|: | ||||   : |||:: ::||||::| : :| ||  | |||::|| ||| ::  |
EAKALNAQIATLNESIKERTKTLEAQARSAQVNSSATNYMDAVVNSKSLTDVIQKVTAIATVSSANKQMLEQQEKEQKEL
        90        100       110       120       130       140       150       160
```

```
804        834        864        894        924        954        984       1014
EAKQIENQNAINTVAANKQAIENNKAALATQRAQLEAAQLELSAQLTTVQNEKASLIQAKAQAEEAARKAAEAQAAAEAK
  |    :   |   :  |::::    | :|:|:|: | |   | : | |::| :|:  || ||:||::||: ||| ||:
SQKSETVKKNYNQFVSLSQSLDSQAQELTSQQAELKVATLNYQATIATAQDKKQALLDEKAAAEKAAQEAAKKQAAYEAQ
       170        180        190        200        210        220        230        240

1044       1065       1095       1125       1155       1185       1215
AQAEAKAQAESVA---KAQAAAQVESATAPTETVQTQPRTEIKPSNLTATSSATTVATTTATATNEPKVTQPSVVTKA--
 :  |:||| | |   ||  ||   :: : ::  |    |:   || :|::|: : : :::::::    :  |  : |
QKEAAQAQAASTAATAKAVEAATSSASASSSQAPQVSTSTDNTTSNASASNSSNSSSNSSSSSSSSSSSSSSSNSNAGG
       250        260        270        280        290        300        310        320

1266       1296       1326       1347       1374       1401                  1455
-VEAPKAVVSSTPRAVSKPVVRSYDSSNTYPMGQCT---WGA-KSMASWVGNYW-GNANQWGASARAAG--YSVGTTPRV
   :  :  ::         : |    | |  | ||   |    :   :: |   ||  || ::  | |  : ||  | |
NTNSGTSTGNTGGTTTGGSGINSSPIGNPYAGGGCTDYVWQYFAAQGIYIRNIMPGNGGQWASNGPAQGVLHVVGAAPGV
       330        340        350        360        370        380        390        400

1503       1533       1563       1593       1623       1653       1683
GAVAVWP----YDGGGYGHVAVVTSVANNSSIQVMESNYAGNMSIGNYRGSFNPSASGSVYYIYPN**ILRRSFVVSFLF
 | :       |    |||||:| || :: :| : |  |           | :
IASSFSADFVGYANSPYGHVAIVKSVNSDGTITIKEGGYGTTWWGHERTVSASGVTFLMPN
       410        420        430        440        450        460
```

SEQ ID 8544 (GBS65) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 6; MW 47.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 3; MW 72kDa) and in Figure 175 (lane 2 & 3; MW 72kDa).

The GBS65-GST fusion product was purified (Figure 102A; see also Figure 191, lane 4) and used to immunise mice (lane 1 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 102B), FACS, 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 314

A DNA sequence (GBSx0343) was identified in *S.agalactiae* <SEQ ID 1015> which encodes the amino acid sequence <SEQ ID 1016>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1184(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 315

A DNA sequence (GBSx0344) was identified in *S.agalactiae* <SEQ ID 1017> which encodes the amino acid sequence <SEQ ID 1018>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4736(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 316

A DNA sequence (GBSx0345) was identified in *S.agalactiae* <SEQ ID 1019> which encodes the amino acid sequence <SEQ ID 1020>. This protein is predicted to be elongation factor Tu (tufA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3012(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 9737> which encodes amino acid sequence <SEQ ID 9738> was also identified.

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

```
>GP:BAB03851 GB:AP001507 translation elongation factor Tu (EF-Tu)
          [Bacillus halodurans]
 Identities = 302/397 (76%), Positives = 350/397 (88%), Gaps = 2/397 (0%)

Query: 7   MAKEKYDRSKPHVNIGTIGHVDHGKTTLTAAITTVLARRLPTSVNQPKDYASIDAAPEER 66
           MAKEK+DRSK H NIGTIGHVDHGKTTLTAAITTVLA+R    V     Y +ID APEER
Sbjct: 1   MAKEKFDRSKTHANIGTIGHVDHGKTTLTAAITTVLAKRSGKGVAMA--YDAIDGAPEER 58

Query: 67  ERGITINTAHVEYETEKRHYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDGPMPQTR 126
           ERGITI+TAHVEYET+ RHYAH+D PGHADYVKNMITGAAQMDG ILVV++ DGPMPQTR
Sbjct: 59  ERGITISTAHVEYETDNRHYAHVDCPGHADYVKNMITGAAQMDGGILVVSAADGPMPQTR 118

Query: 127 EHILLSRQVGVKHLIVFMNKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPVIQGSALK 186
           EHILLSRQVGV +L+VF+NK D+VDDEELLELVEME+RDLLSEYDFPGDD+PVI+GSALK
Sbjct: 119 EHILLSRQVGVPYLVVFLNKCDMVDDEELLELVEMEVRDLLSEYDFPGDDVPVIRGSALK 178

Query: 187 ALEGDEKYEDIIMELMSTVDEYIPEPERDTDKPLLLPVEDVFSITGRGTVASGRIDRGTV 246
           ALEGD ++E+ I+ELM+ VD+YIP PERDT+KP ++PVEDVFSITGRGTVA+GR++RG +
Sbjct: 179 ALEGDAEWEEKIIELMAAVDDYIPTPERDTEKPFMMPVEDVFSITGRGTVATGRVERGQL 238

Query: 247 RVNDEVEIVGIKEDIQKAVVTGVEMFRKQLDEGLAGDNVGVLLRGVQRDEIERGQVLAKP 306
            V DEVEI+G++E+ +K  VTGVEMFRK LD   AGDN+G LLRGV R+E++RGQVLAKP
Sbjct: 239 NVGDEVEIIGLEEEAKKTTVTGVEMFRKLLDYAEAGDNIGALLRGVSREEVQRGQVLAKP 298
```

```
Query: 307 GSINPHTRFKGEVYILSKEEGGRHTPFFNNYRPQFYFRTTDVTGSIELPAGTEMVMPGDN 366
           G+I PHT FK EVY+LSKEEGGRHTPFF+NYRPQFYFRTTDVTG I+LP G EMVMPGDN
Sbjct: 299 GTITPHTNFKAEVYVLSKEEGGRHTPFFSNYRPQFYFRTTDVTGIIQLPDGVEMVMPGDN 358

Query: 367 VTIEVELIHPIAVEQGTTFSIREGGRTVGSGIVSEIE 403
           V + VELI PIA+E+GT FSIREGGRTVG+G+V+ I+
Sbjct: 359 VEMTVELIAPIAIEEGTKFSIREGGRTVGAGVVASIQ 395
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1021> which encodes the amino acid sequence <SEQ ID 1022>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1367(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 = 386/404 (95%), Positives = 396/404 (97%)

Query: 1   MEAFPKMAKEKYDRSKPHVNIGTIGHVDHGKTTLTAAITTVLARRLPTSVNQPKDYASID 60
           +EAFPKMAKEKYDRSKPHVNIGTIGHVDHGKTTLTAAITTVLARRLP+SVNQPKDYASID
Sbjct: 12  LEAFPKMAKEKYDRSKPHVNIGTIGHVDHGKTTLTAAITTVLARRLPSSVNQPKDYASID 71

Query: 61  AAPEERERGITINTAHVEYETEKRHYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDG 120
           AAPEERERGITINTAHVEYET  RHYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDG
Sbjct: 72  AAPEERERGITINTAHVEYETATRHYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDG 131

Query: 121 PMPQTREHILLSRQVGVKHLIVFMNKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPVI 180
           PMPQTREHILLSRQVGVKHLIVFMNKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPVI
Sbjct: 132 PMPQTREHILLSRQVGVKHLIVFMNKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPVI 191

Query: 181 QGSALKALEGDEKYEDIIMELMSTVDEYIPEPERDTDKPLLLPVEDVFSITGRGTVASGR 240
           QGSALKALEGD K+EDIIMELM TVD YIPEPERDTDKPLLLPVEDVFSITGRGTVASGR
Sbjct: 192 QGSALKALEGDTKFEDIIMELMDTVDSYIPEPERDTDKPLLLPVEDVFSITGRGTVASGR 251

Query: 241 IDRGTVRVNDEVEIVGIKEDIQKAVVTGVEMFRKQLDEGLAGDNVGVLLRGVQRDEIERG 300
           IDRGTVRVNDE+EIVGIKE+ +KAVVTGVEMFRKQLDEGLAGDNVG+LLRGVQRDEIERG
Sbjct: 252 IDRGTVRVNDEIEIVGIKEETKKAVVTGVEMFRKQLDEGLAGDNVGILLRGVQRDEIERG 311

Query: 301 QVLAKPGSINPHTRFKGEVYILSKEEGGRHTPFFNNYRPQFYFRTTDVTGSIELPAGTEM 360
           QV+AKP SINPHT+FKGEVYILSK+EGGRHTPFFNNYRPQFYFRTTDVTGSIELPAGTEM
Sbjct: 312 QVIAKPSSINPHTKFKGEVYILSKDEGGRHTPFFNNYRPQFYFRTTDVTGSIELPAGTEM 371

Query: 361 VMPGDNVTIEVELIHPIAVEQGTTFSIREGGRTVGSGIVSEIEA 404
           VMPGDNVTI VELIHPIAVEQGTTFSIREGGRTVGSGIVSEIEA
Sbjct: 372 VMPGDNVTINVELIHPIAVEQGTTFSIREGGRTVGSGIVSEIEA 415
```

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

Example 317

A DNA sequence (GBSx0346) was identified in *S.agalactiae* <SEQ ID 1023> which encodes the amino acid sequence <SEQ ID 1024>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -0.64   Transmembrane   90 - 106 (  90 - 106)

----- Final Results -----
               bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
```

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

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

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

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

Example 318

A DNA sequence (GBSx0347) was identified in *S.agalactiae* <SEQ ID 1025> which encodes the amino acid sequence <SEQ ID 1026>. This protein is predicted to be ftsW. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.15   Transmembrane   44 -  60 (  35 -  70)
   INTEGRAL    Likelihood = -4.73   Transmembrane   76 -  92 (  74 -  98)
   INTEGRAL    Likelihood = -3.88   Transmembrane  117 - 133 ( 113 - 134)

----- 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 GENPEPT database:

```
>GP:AAB39929 GB:U58049 putative cell division protein ftsW
          [Enterococcus hirae]
 Identities = 78/159 (49%), Positives = 107/159 (67%), Gaps = 4/159 (2%)

Query: 1   MANSXYAMSNGGWFGRGLGNSIEKLGYLPEATTDFVFSIVIEELGVIGAGFILALVFFLI 60
           M+NS YA+ NGG FGRG+GNSI K GYLPE+ TDF+FS++ EE G+IGA  +L L+F L
Sbjct: 240 MSNSYYALYNGGLFGRGMGNSITKKGYLPESETDFIFSVIAEEFGLIGALLVLFLLFLLC 299

Query: 61  LRIMHVGIKAKDPFNSMIALGIGAMLLMQVFVNIGGISGLIPSTGVTFPFLSQGGNSLLV 120
           +RI     K K+   ++I +G+G  +L+Q  +NIG I GLIP TGV  PF+S GG S L+
Sbjct: 300 MRIFQKSTKQKNQQANLILIGVGTWILVQTSINIGSILGLIPMTGVPLPFVSYGGTSYLI 359

Query: 121 LSVAIGFVLNIDANEKKELIMKEAEEQYKPQEKNEKIIN 159
           LS AIG  LNI + + KE    + ++  + Q K  K++N
Sbjct: 360 LSFAIGLALNISSRQVKE----KNKQVERLQLKKPKLLN 394
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1027> which encodes the amino acid sequence <SEQ ID 1028>. Analysis of this protein sequence reveals the following:

```
     Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.93   Transmembrane  312 - 328 ( 303 - 338)
   INTEGRAL    Likelihood = -8.23   Transmembrane   22 -  38 (  17 -  47)
   INTEGRAL    Likelihood = -6.85   Transmembrane  192 - 208 ( 187 - 211)
   INTEGRAL    Likelihood = -5.10   Transmembrane  218 - 234 ( 212 - 236)
   INTEGRAL    Likelihood = -4.83   Transmembrane   86 - 102 (  85 - 107)
   INTEGRAL    Likelihood = -3.72   Transmembrane  385 - 401 ( 383 - 402)
   INTEGRAL    Likelihood = -3.45   Transmembrane   61 -  77 (  61 -  79)
   INTEGRAL    Likelihood = -2.39   Transmembrane  344 - 360 ( 344 - 360)

----- Final Results -----
               bacterial membrane --- Certainty=0.5373(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB59721 GB:AJ250603 FtsW protein [Enterococcus faecium]
 Identities = 131/397 (32%), Positives = 223/397 (55%), Gaps = 23/397 (5%)

Query: 15  KRHLLNYSILLPYLILSVIGLIMVYSTTSVSLIQAHANPFKSVINQGVFWIISLVAITFI 74
           KR  +++ IL PYL LS+IGL+ VYS +S  L+QA  N    ++ Q +F  +S   I
Sbjct: 3   KRKKIDWWILGPYLTLSMIGLLEVYSASSYRLLQADENTKSLLLRQLIFIFLSWGVIFLA 62

Query: 75  YKLKLNFLTNTRVLTVVMLGEAFLLIIAR--FFTTAIKGAHGWIVIGPVSFQPAEYLKII 132
             +KL++L + ++    +    F LI+ R   F   + GA  WI +  + FQP+E   +
Sbjct: 63  RSIKLHYLLHPKIAGYGLALSIFFLILVRVGIFGVTVNGAQRWISLFGIQFQPSELANLF 122

Query: 133 MVWYLALTFAKIQKNISLYDYQALTRRKWWPTQWNDLRDWRVYSLLMVLLVAAQPDLGNA 192
           +++YL+  F                    P +  +L+   + ++ + LL+  QP +  A
Sbjct: 123 LIFYLSWFFRDGNN----------------PPK--NLKKPFLITVSITLLILFQPKIAGA 164

Query: 193 SIIVLTAIIMFSISGIGYRWPSAILVMITGLSTVFLGTIAVIGVERVAKIP-VFGYVAKR 251
            +I+  A ++F  + + ++    ++V  + L     G +  +G +    +P +F  +A R
Sbjct: 165 LMILSIAWVIFWAAAVPFKKGIYLIVTFSALLIGAAGGVLYLGNK--GWLPQMFNHAYER 222

Query: 252 FSAFFNPFHDLTDSGHQLANSYYAMSNGGWFGQGLGNSIEKRGYLPEAQTDFVFSVVIEE 311
            +   +PF D   +G+Q+ +S+YA+ NGG +G+GLGNSI K+GYLPE +TDF+FS++ EE
Sbjct: 223 IATLRDPFIDSHGAGYQMTHSFYALYNGGIWGRGLGNSITKKGYLPETETDFIFSIITEE 282

Query: 312 LGLIGAGFILALVFFLILRIMNVGIKAKNPFNAMMALGVGGMMLMQVFVNIGGISGLIPS 371
           LGLIGA  +L L+F L +RI  +  + KN    +  LG G ++ +Q  +N+G I+GL+P
Sbjct: 283 LGLIGALCVLFLLFSLCMRIFCLSSRCKNQQAGLFLLGFGTLLFVQTIMNVGSIAGLMPM 342

Query: 372 TGVTFPPFLSQGGNSLLVLSVAVGFVLNIDASEKRDDI 408
           TGV  PF+S GG S L+LS+ +G  LNI +  + +++
Sbjct: 343 TGVPLPFVSYGGTSYLILSLGIGITLNISSKIQAEEL 379
```

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

```
 Identities = 130/166 (78%), Positives = 152/166 (91%), Gaps = 2/166 (1%)

Query: 1   MANSXYAMSNGGWFGRGLGNSIEKLGYLPEATTDFVFSIVIEELGVIGAGFILALVFFLI 60
           +ANS YAMSNGGWFG+GLGNSIEK GYLPEA TDFVFS+VIEELG+IGAGFILALVFFLI
Sbjct: 269 LANSYYAMSNGGWFGQGLGNSIEKRGYLPEAQTDFVFSVVIEELGLIGAGFILALVFFLI 328

Query: 61  LRIMHVGIKAKDPFNSMIALGIGAMLLMQVFVNIGGISGLIPSTGVTFPFLSQGGNSLLV 120
           LRIM+VGIKAK+PFN+M+ALG+G M+LMQVFVNIGGISGLIPSTGVTFPFLSQGGNSLLV
Sbjct: 329 LRIMNVGIKAKNPFNAMMALGVGGMMLMQVFVNIGGISGLIPSTGVTFPFLSQGGNSLLV 388

Query: 121 LSVAIGFVLNIDANEKKELIMKEAEEQYK--PQEKNEKIINLDAFK 164
           LSVA+GFVLNIDA+EK++ I KEAE  Y+   +++N K++N+  F+
Sbjct: 389 LSVAVGFVLNIDASEKRDDIFKEAELSYRKDTRKENSKVVNIKQFQ 434
```

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

Example 319

A DNA sequence (GBSx0348) was identified in *S.agalactiae* <SEQ ID 1029> which encodes the amino acid sequence <SEQ ID 1030>. This protein is predicted to be probable cell division protein ftsw (ftsW). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.77   Transmembrane   12 -  28 (   7 -  37)
   INTEGRAL    Likelihood = -7.22   Transmembrane   76 -  92 (  74 -  97)
   INTEGRAL    Likelihood = -6.53   Transmembrane  182 - 198 ( 178 - 201)
   INTEGRAL    Likelihood = -4.62   Transmembrane   51 -  67 (  46 -  69)
   INTEGRAL    Likelihood = -2.87   Transmembrane  202 - 218 ( 202 - 218)
```

```
----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9327> which encodes amino acid sequence <SEQ ID 9328> was also identified.

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

```
>GP:CAA44490 GB:X62621 ORF2 N-terminal [Lactococcus lactis]
 Identities = 82/199 (41%), Positives = 122/199 (61%), Gaps = 9/199 (4%)

Query: 1   MKIDKRHLLNYSILIPYLILSILGLIVIYSTTSATLIQLGANPFRSVINQGVFWAVSLVA 60
           M ++K + LNYSILIPYLIL+ +G+++I+STT    +Q G NP++ VINQ  F  +S++
Sbjct: 1   MNLNKNNFLNYSILIPYLILAGIGIVMIFSTTVPDQLQKGLNPYKLVINQTAFVLLSIIM 60

Query: 61  IIFIYKLKLNFLKNSKVLTMAVLVEVFLLLIARF------FTQEVNGAHGWIVIGPI-SF 113
           I  IY+LKL  LKN K++ + +++ +  L+  R        T  VNGA GWI I  I +
Sbjct: 61  IAVIYRLKLRALKNRKMIGIIMVILILSLIFCRIMPSSFALTAPVNGARGWIHIPGIGTV 120

Query: 114 QPAEYLKVIIVWYLAFTFARRQKKIEIYDYQALTKGRWLPRSLSDLKDWRFYSLFMIGLV 173
           QPAE+ KV I+WYLA  F+ +Q++IE  D   + KG+ L + L     WR   + ++ +
Sbjct: 121 QPAEFAKVFIIWYLASVFSTKQEEIEKNDINEIFKGKTLTQKL--FGGWRLPVVAILLVD 178

Query: 174 IAQPDLGNGSIIVLTVIIM 192
           +  PDLGN  II    +IM
Sbjct: 179 LIMPDLGNTMIIGAVALIM 197
```

There is also homology to SEQ ID 1028.

A related GBS gene <SEQ ID 8545> and protein <SEQ ID 8546> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:     15.18
GvH: Signal Score (-7.5): -3.58
     Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 5 value:  -9.77 threshold:  0.0
   INTEGRAL    Likelihood = -9.77   Transmembrane   12 -  28 (   7 -  37)
   INTEGRAL    Likelihood = -7.22   Transmembrane   76 -  92 (  74 -  97)
   INTEGRAL    Likelihood = -6.69   Transmembrane  210 - 226 ( 201 - 227)
   INTEGRAL    Likelihood = -6.53   Transmembrane  182 - 198 ( 178 - 201)
   INTEGRAL    Likelihood = -4.62   Transmembrane   51 -  67 (  46 -  69)
   PERIPHERAL  Likelihood =  1.32     116
 modified ALOM score:   2.45

*** Reasoning Step: 3

----- 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:

```
ORF02700(301 - 876 of 1377)
EGAD|8615|8419(1  -  197  of  198)  hypothetical  protein  in  rpmg  3'region  ,  fragment
(Lactococcus lactis) SP|P27174|YRG2_LACLA HYPOTHETICAL PROTEIN IN RPMG 3'REGION (ORF2)
(FRAGMENT).   GP|44069|emb|CAA44490.1||X62621   ORF2   N-terminal   (Lactococcus   lactis)
PIR|PC1134|PC1134 hypothetical protein 198 (rmpG 3' region) - Lactococcus lactis (fragment)
%Match = 15.1
%Identity = 42.3  %Similarity = 64.9
Matches = 82  Mismatches = 64  Conservative Sub.s = 44
```

```
87        117       147       177       207       237       267       297
KA*I*Y*I**L*LVILFLLPFFINFL*IYLTGLND*NVPSNISN*SFIFVISIVGGYXX*LIXXXIMHNGNFLKY*RK*Y

327       357       387       417       447       477       507       537
NMKIDKRHLLNYSILIPYLILSILGLIVIYSTTSATLIQLGANPFRSVINQGVFWAVSLVAIIFIYKLKLNFLKNSKVLT
 | ::| ::|||||||||||||: :|:::|:|||    :| | ||:: |||| |  :|:: |  ||:||| ||| |::
 MNLNKNNFLNYSILIPYLILAGIGIVMIFSTTVPDQLQKGLNPYKLVINQTAFVLLSIIMIAVIYRLKLRALKNRKMIG
          10        20        30        40        50        60        70

567       585       609       636       666       696       726       756
MAVLVEVFLLLIARF----FT--QEVNGAHGWIVIGPI-SFQPAEYLKVIIVWYLAFTFARRQKKIEIYDYQALTKGRWL
: ::: :: |:  |     |     |||| ||| |  | : ||||: || |:|||| |: :|::||  |   : ||: |
IIMVILILSLIFCRIMPSSFALTAPVNGARGWIHIPGIGTVQPAEFAKVFIIWYLASVFSTKQEEIEKNDINEIFKGKTL
          90       100       110       120       130       140       150

786       816       846       876       906       936       966       996
PRSLSDLKDWRFYSLFMIGLVIAQPDLGNGSIIVLTVIIMYCISGIGYRWFSALLGLIVVGSTLFIGTIAVVGVETMAKV
 : |  :  ||:  : :: : :  |||||  ||    :||
TQKL--FGGWRLPVVAILLVDLIMPDLGNTMIIGAVALIMI
          170       180       190
```

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

Example 320

A DNA sequence (GBSx0349) was identified in *S.agalactiae* <SEQ ID 1031> which encodes the amino acid sequence <SEQ ID 1032>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3665(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 1033> which encodes the amino acid sequence <SEQ ID 1034>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2373(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 = 35/41 (85%), Positives = 37/41 (89%)

Query: 1  MEKEAKQIIDLKRNLFKIDVRAQKDEEKVFMRTACCYSPFY 41
          +EKEAKQ+IDLKRNLFKIDVRAQKDEEKVFMRTAC  S  Y
Sbjct: 1  LEKEAKQMIDLKRNLFKIDVRAQKDEEKVFMRTACRQSRVY 41
```

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

Example 321

A DNA sequence (GBSx0351) was identified in *S.agalactiae* <SEQ ID 1037> which encodes the amino acid sequence <SEQ ID 1038>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.65   Transmembrane   78 -  94 ( 78 -  95)
   INTEGRAL    Likelihood = -1.33   Transmembrane  421 - 437 ( 420 - 437)

----- Final Results -----
               bacterial membrane --- Certainty=0.1659(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAA00827 GB:A09073 phosphoenol pyruvate carboxylase
          [Corynebacterium glutamicum]
 Identities = 335/958 (34%), Positives = 539/958 (55%), Gaps = 80/958 (8%)

Query: 22  EIITEEVGLLKQLLDEATQKLIGSESFDKIE--KIVSLSLTD---DYTGLKETISALSNE 76
           + +.+++  L Q+L E   +  G E ++ +E  ++ S  +     +   L +    ++
Sbjct: 3   DFLRDDIRFLGQILGEVIAEQEGQEVYELVEQARLTSFDIAKGNAEMDSLVQVFDGITPA 62

Query: 77  EMVIVSRYFSILPLLINISEDVDLAYEINYKNNLNQDYLGKLST----TIDVV------- 125
           +   ++R FS   LL N++ED+       Y   L +  L    T    T+D
Sbjct: 63  KATPIARAFSHFALLANLAEDL-------YDEELREQALDAGDTPPDSTLDATWLKLNEG 115

Query: 126 -AGHENAKDILEHVNVVPVLTAHPTQVQRKTVLELTSKIHDLLRKYRDVKAGIVNQ---- 180
             G E   D+L +  V PVLTAHPT+ +R+TV +    I   +R+   +++
Sbjct: 116 NVGAEAVADVLRNAEVAPVLTAHPTETRRRTVFDAQKWITTHMRERHALQSAEPTARTQS 175

Query: 181 --EKWYADLRRYIGIIMQTDTIREKKLKVKNEITNVMEYYNRSLIKAVTKLTAEYKALAA 238
             ++   ++RR I I+ QT  IR  + ++++EI   + YY  SL++ + ++  +
Sbjct: 176 KLDEIEKNIRRRITILWQTALIRVARPRIEDEIEVGLRYYKLSLLEEIPRINRDVAVELR 235

Query: 239 KK---GIHLENPKPLTM-GMWIGGDRDGNPFVTAETLRLSAMVQSEVIINHYIEQLNELY 294
           ++   G+ L   KP+   G WIGGD DGNP+VTAET+  S    +E ++ +Y  QL+ L
Sbjct: 236 ERFGEGVPL---KPVVKPGSWIGGDHDGNPYVTAETVEYSTHRAAETVLKYYARQLHSLE 292

Query: 295 RNMSLSINLTEVSPELVTLANQSQDNSVYRENEPYRKAFNFIQDKLVQTLLNLKVGSSPK 354
            +SLS  + +V+P+L+ LA+   ++  R +EPYR+A + ++ +++ T
Sbjct: 293 HELSLSDRMNKVTPQLLALADAGHNDVPSRVDEPYRRAVHGVRGRILAT---------- 341

Query: 355 EKFVSRQESSDIVGRYIKSHIAQVASDIQTEELPAYATAEEFKQDLLLVKQSLVQYGQDS 414
                +++++G          +  +    YA+ EEF  D L +  SL +
Sbjct: 342 --------TAELIGE-------DAVEGVWFKVFTPYASPEEFLNDALTIDHSLRESKDVL 386

Query: 415 LVDGELACLIQAVDIFGFYLATIDMRQDSSINEACVAELLKSANIVDDYSSLSEEEKCQL 474
           + D  L+ LI A++ FGF L  +D+RQ+S   E  + EL + A +  +Y  LSE EK ++
Sbjct: 387 IADDRLSVLISAIESFGFNLYALDLRQNSESYEDVLTELFERAQVTANYRELSEAEKLEV 446

Query: 475 LLKELTEDPRTLSSTHAPKSELLQKELAIFQTARELKDQLGEDIINQHIISHTESVSDMF 534
           LLKEL      +      SE+  +EL IF+TA E   + G  ++   IIS   SV+D+
Sbjct: 447 LLKELRSPRPLIPHGSDRYSEVTDRELGIFRTASEAVKKFGPRMVPHCIISMASSVTDVL 506

Query: 535 ELAIMLKEVGLIDAN----QARIQIVPLFETIEDLDNSRDIMTQYLHYELVKKWIATNNN 590
           E  ++LKE GLI AN    +  + ++PLFETIEDL      I+ +    +L + ++   +N
Sbjct: 507 EPMVLLKEFGLIAANGDNPRGTVDVIPLFETIEDLQAGAGILDELWKIDLYRNYLLQRDN 566

Query: 591 YQEIMLGYSDSNKDGGYLSSGWTLYKAQNELTKIGEENGIKITFFHGRGGTVGRGGGPSY 650
            QE+MLGYSDSNKDGGY S+ W LY A+ +L ++    G+K+  FHGRGGTVGRGGGPSY
Sbjct: 567 VQEVMLGYSDSNKDGGYFSANWALYDAELQLVELCRSAGVKLRLFHGRGGTVGRGGGPSY 626

Query: 651 EAITSQPFGSIKDRIRLTEQGEIIENKYGNQDAAYYNLEMLISASIDRMVTRMITNPNEI 710
           +AI +QP G+++  +R+TEQGEII  KYGN + A  NLE L+SA+++       + + +E+
```

```
Sbjct: 627 DAILAQPRGAVQGSVRITEQGEIISAKYGNPETARRNLEALVSATLE----ASLLDVSEL 682

Query: 711 DNFRETMDGIVSESNAV----YRNLVFDNPYFYDYFFEASPIKEVSSLNIGSRPAARKTI 766
           + +   D I+SE + +    Y +LV ++  F DYF +++P++E+ SLNIGSRP++RK
Sbjct: 683 TDHQRAYD-IMSEISELSLKKYASLVHEDQGFIDYFTQSTPLQEIGSLNIGSRPSSRKQT 741

Query: 767 TEISGLRAIPWVFSWSQNRIMFPGWYGVGSAFKHFI---EQDEANLAKLQTMYQKWPFFN 823
           + +  LRAIPWV SWSQ+R+M PGW+GVG+A + +I   EQ    +A+LQT+ + WPFF
Sbjct: 742 SSVEDLRAIPWVLSWSQSRVMLPGWFGVGTALEQWIGEGEQATQRIAELQTLNESWPFFT 801

Query: 824 SLLSNVDMVLSKSNMNIALQYAQLAGSKEVRD-VFNIILNEWQLTKDMILAIEQHDNLLE 882
           S+L N+  V+SK+ + +A  YA L    EV + V+++I  E+ LTK M   I   D+LL+
Sbjct: 802 SVLDNMAQVMSKAELRLAKLYADLIPDTEVAERVYSVIREEYFLTKKMFCVITGSDDLLD 861

Query: 883 ENPMLHASLDYRLPYFNVLNYVQIELIKRLRSNQLDEDYEKLIHITINGIATGLRNSG 940
           +NP+L  S+  R PY   LN +Q+E+++R R     E   + I +T+NG++T LRNSG
Sbjct: 862 DNPLLARSVQRRYPYLLPLNVIQVEMMRRYRKGDQSEQVSRNIQLTMNGLSTALRNSG 919
```

A related GBS nucleic acid sequence <SEQ ID 10961> which encodes amino acid sequence <SEQ ID 10962> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1039> which encodes the amino acid sequence <SEQ ID 1040>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1613(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 = 659/927 (71%), Positives = 779/927 (83%), Gaps = 11/927 (1%)

Query: 14  KLESSSNKEIITEEVGLLKQLLDEATQKLIGSESFDKIEKIVSLSLTDDYTGLKETISAL 73
           KLESS+N++II EEV LLK++L+  T+++IG ++F  IE I+ LS   DY  L++ ++ +
Sbjct: 5   KLESSNNQDIIAEEVALLKEMLENITRRMIGDDAFTVIESIMVLSEKQDYIELEKVVANI 64

Query: 74  SNEEMVIVSRYFSILPLLINISEDVDLAYEINYKNNLNQDYLGKLSTTIDVVAGHENAKD 133
           SN+EM ++SRYFSILPLLINISEDVDLAYEINY+NN + DYLGKL+ TI  +AG +N KD
Sbjct: 65  SNQEMEVISRYFSILPLLINISEDVDLAYEINYQNNTDTDYLGKLALTIKDLAGKDNGKD 124

Query: 134 ILEHVNVVPVLTAHPTQVQRKTVLELTSKIHDLLRKYRDVKAGIVNQEKWYADLRRYIGI 193
           ILE VNVVPVLTAHPTQVQRKT+LELT+ IH LLRKYRD KAG++N EKW  +L RYI +
Sbjct: 125 ILEQVNVVPVLTAHPTQVQRKTILELTTHIHKLLRKYRDAKAGVINLEKWRQELYRYIEM 184

Query: 194 IMQTDTIREKKLKVKNEITNVMEYYNRSLIKAVTKLTAEYKALAAKKGIHLENPKPLTMG 253
           IMQTD IREKKL+VKNEI NVM+YY+ SLI+AVTKLT EYK LA K G+ L+NPKP+TMG
Sbjct: 185 IMQTDIIREKKLQVKNEIKNVMQYYDGSLIQAVTKLTTEYKNLAQKHGLELDNPKPITMG 244

Query: 254 MWIGGDRDGNPFVTAETLRLSAMVQSEVIINHYIEQLNELYRNMSLSINLTEVSPELVTL 313
           MWIGGDRDGNPFVTAETL LSA VQSEVI+N+YI++L  LYR  SLS  L + + E+  L
Sbjct: 245 MWIGGDRDGNPFVTAETLCLSATVQSEVILNYYIDELAALYRTFSLSSTLVQPNSEVERL 304

Query: 314 ANQSQDNSVYRENEPYRKAFNFIQDKLVQTLLNLKVGSSPKEKFVSRQESSDIVGRYIKS 373
           A+ SQD S+YR NEPYR+AF++IQ +L QT + L        +  +   SS +     S
Sbjct: 305 ASLSQDQSIYRGNEPYRRAFHYIQSRLKQTQIQLT------NQPAASMSSSVGLNTSAWS 358

Query: 374 HIAQVASDIQTEELPAYATAEEFKQDLLLVKQSLVQYGQDSLVDGELACLIQAVDIFGFY 433
             A + + I      AY +  +FK DL  ++QSL+  G  +L++G+L  ++QAVDIFGF+
Sbjct: 359 SPASLENPIL-----AYDSPVDFKADLKAIEQSLLDNGNSALIEGDLREVMQAVDIFGFF 413

Query: 434 LATIDMRQDSSINEACVAELLKSANIVDDYSSLSEEEKCQLLLKELTEDPRTLSSTHAPK 493
           LA+IDMRQDSS+ EACVAELLK ANIVDDYSSLSE EKC +LL++L E+PRTLSS    K
```

```
Sbjct: 414 LASIDMRQDSSVQEACVAELLKGANIVDDYSSLSETEKCDVLLQQLMEEPRTLSSAAVAK 473

Query: 494 SELLQKELAIFQTARELKDQLGEDIINQHIISHTESVSDMFELAIMLKEVGLIDANQARI 553
           S+LL+KELAI+ TARELKD+LGE++I QHIISHTESVSDMFELAIMLKEVGL+D  +AR+
Sbjct: 474 SDLLEKELAIYTTARELKDKLGEEVIKQHIISHTESVSDMFELAIMLKEVGLVDQQRARV 533

Query: 554 QIVPLFETIEDLDNSRDIMTQYLHYELVKKWIATNNNYQEIMLGYSDSNKDGGYLSSGWT 613
           QIVPLFETIEDLDN+RDIM  YL +++VK WIATN NYQEIMLGYSDSNKDGGYL+SGWT
Sbjct: 534 QIVPLFETIEDLDNARDIMAAYLSHDIVKSWIATNRNYQEIMLGYSDSNKDGGYLASGWT 593

Query: 614 LYKAQNELTKIGEENGIKITFFHGRGGTVGRGGGPSYEAITSQPFGSIKDRIRLTEQGEI 673
           LYKAQNELT IGEE+G+KITFFHGRGGTVGRGGGPSY+AITSQPFGSIKDRIRLTEQGEI
Sbjct: 594 LYKAQNELTAIGEEHGVKITFFHGRGGTVGRGGGPSYDAITSQPFGSIKDRIRLTEQGEI 653

Query: 674 IENKYGNQDAAYYNLEMLISASIDRMVTRMITNPNEIDNFRETMDGIVSESNAVYRNLVF 733
           IENKYGN+D AYY+LEMLISASI+RMVT+MIT+PNEID+FRE MD IV++SN +YR LVF
Sbjct: 654 IENKYGNKDVAYYHLEMLISASINRMVTQMITDPNEIDSFREIMDSIVADSNIIYRKLVF 713

Query: 734 DNPYFYDYFFEASPIKEVSSLNIGSRPAARKTITEISGLRAIPWVFSWSQNRIMFPGWYG 793
           DNP+FYDYFFEASPIKEVSSLNIGSRPAARKTITEI+GLRAIPWVFSWSQNRIMFPGWYG
Sbjct: 714 DNPHFYDYFFEASPIKEVSSLNIGSRPAARKTITEITGLRAIPWVFSWSQNRIMFPGWYG 773

Query: 794 VGSAFKHFIEQDEANLAKLQTMYQKWPFFNSLLSNVDMVLSKSNMNIALQYAQLAGSKEV 853
           VGSAFK +I++ + NL +LQ MYQ WPFF+SLLSNVDMVLSKSNMNIA QYAQLA  ++V
Sbjct: 774 VGSAFKRYIDRAQGNLERLQHMYQTWPFFHSLLSNVDMVLSKSNMNIAFQYAQLAERQDV 833

Query: 854 RDVFNIILNEWQLTKDMILAIEQHDNLLEENPMLHASLDYRLPYFNVLNYVQIELIKRLR 913
           RDVF  IL+EWQLTK++ILAI+ HD+LLE+NP L  SL  RLPYFNVLNY+QIELIKR R
Sbjct: 834 RDVFYEILDEWQLTKNVILAIQDHDDLLEDNPSLKHSLKSRLPYFNVLNYIQIELIKRWR 893

Query: 914 SNQLDEDYEKLIHITINGIATGLRNSG 940
           +NQLDE+ EKLIH TINGIATGLRNSG
Sbjct: 894 NNQLDENDEKLIHTTINGIATGLRNSG 920
```

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

Example 322

A DNA sequence (GBSx0352) was identified in *S.agalactiae* <SEQ ID 1041> which encodes the amino acid sequence <SEQ ID 1042>. This protein is predicted to be Bacillus licheniformis Pz-peptidase homologue (pepF). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3012(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 1043> which encodes the amino acid sequence <SEQ ID 1044>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3137(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 = 512/593 (86%), Positives = 564/593 (94%)

Query: 1   MKLKKRSEFPENELWDLTALYKDRQDFLLAIEKALEDIKVFKKNYEGKLNCVEDFTSALM 60
           M+LKKRSEFPENELWDLTALYKDRQDFLLAIEKAL+DI +FK+NYEG+L  V+DFT AL+
Sbjct: 26  MELKKRSEFPENELWDLTALYKDRQDFLLAIEKALQDIDLFKRNYEGRLTSVDDFTQALI 85

Query: 61  EIEHIYIQMSHIDTYAFMPQTTDFSNEEFAQISQAGSDFATKANVLLSFFNTALANADIK 120
           EIEHIYIQMSHI TYAFMPQTTDFS+E FAQI+QAG DF TKA+V LSFF+TALANAD+
Sbjct: 86  EIEHIYIQMSHIGTYAFMPQTTDFSDESFAQIAQAGDDFMTKASVALSFFDTALANADLD 145

Query: 121 ILDSLENNPHFKATIRQAKIQKQHLLSPEVEKALTNLNEVLNTPYDIYTKMRAGDFDMED 180
           +LD+LE NP+F A IR AKIQK+HLLSP+VEKAL NL EV+N PYDIYTKMRAGDFDM+D
Sbjct: 146 VLDTLEKNPYFSAAIRMAKIQKEHLLSPDVEKALANLREVINAPYDIYTKMRAGDFDMDD 205

Query: 181 FEVDGKTYKNSFVTYENYFQNHENAEIREKSFRSFSKGLRKHQNAAAAAYLAKVKSEKLI 240
           FEVDGKTYKNSFV+YEN++QNHENAEIREK+FRSFSKGLRKHQN AAAAYLAKVKSEKL+
Sbjct: 206 FEVDGKTYKNSFVSYENFYQNHENAEIREKAFRSFSKGLRKHQNTAAAAYLAKVKSEKLL 265

Query: 241 ADMRGYDSVFDYLLSEQEVDRSMFDRQIDLIMDEFGPVAQRFLKHIADVNGIEKMTFADW 300
           ADM+GY SVFDYLL+EQEVDRS+FDRQIDLIM EFGPVAQ+FLKH+A VNG+EKMTFADW
Sbjct: 266 ADMKGYASVFDYLLAEQEVDRSLFDRQIDLIMTEFGPVAQKFLKHVAQVNGLEKMTFADW 325

Query: 301 KLDIDNELNPEVSINDAYDLVMKSVAPLGKEYSQEVERYQKERWVDFAANANKDSGGYAA 360
           KLDIDN+LNPEVSI+ AYDLVMKS+APLG+EY++E+ERYQ ERWVDFAANANKDSGGYAA
Sbjct: 326 KLDIDNDLNPEVSIDGAYDLVMKSLAPLGQEYTKEIERYQTERWVDFAANANKDSGGYAA 385

Query: 361 DPYKVHPYVLMSWTGRMSDVYTLIHEIGHSGQFIFSDNHQSFFNTHMSTYYVEAPSTFNE 420
           DPYKVHPYVLMSWTGRMSDVYTLIHEIGHSGQFIFSDNHQS+FNTHMSTYYVEAPSTFNE
Sbjct: 386 DPYKVHPYVLMSWTGRMSDVYTLIHEIGHSGQFIFSDNHQSYFNTHMSTYYVEAPSTFNE 445

Query: 421 LLLSDYLENQFDTARQKRFALAHRLTDTYFHNFITHLLEAAFQRKVYTLIEEGGTFGAEQ 480
           L+LSDYLE+QFD  RQKRFALAHRLTDTYFHNFITHLLEAAFQRKVYTLIEEGGTFGA+Q
Sbjct: 446 LMLSDYLEHQFDDPRQKRFALAHRLTDTYFHNFITHLLEAAFQRKVYTLIEEGGTFGADQ 505

Query: 481 LNAIMKEVLTQFWGDAIEIDDDAALTWMRQAHYYMGLYSYTYSAGLVISTAGYLNLKNNP 540
           LNA+MKEVLT FWGDA++IDDDAALTWMRQAHYYMGLYSYTYSAGLVISTAGYLNLK+NP
Sbjct: 506 LNAMMKEVLTDFWGDAVDIDDDAALTWMRQAHYYMGLYSYTYSAGLVISTAGYLNLKHNP 565

Query: 541 NGAKEWLAFLKSGGSRTPLETALLISADISTDKPLRDTINFLSNTVDQIINYS 593
           NGAKEWL FLKSGGSRTPL+TA+LI ADI+T+KPLRDTI FLS+TVDQII+Y+
Sbjct: 566 NGAKEWLDFLKSGGSRTPLDTAMLIGADIATEKPLRDTIQFLSDTVDQIISYT 618
```

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

Example 323

A DNA sequence (GBSx0353) was identified in *S.agalactiae* <SEQ ID 1045> which encodes the amino acid sequence <SEQ ID 1046>. 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 no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1047> which encodes the amino acid sequence <SEQ ID 1048>. 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 no significant homology with any sequences in the GENPEPT database.

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

```
Identities = 72/127 (56%), Positives = 85/127 (66%)

Query: 1   MKKYIKLFLLTVFATTLVACGQPSTSNKTTTSSTLEVGKVELVVKEDTNVLSEKVVYHKG 60
           + K  K   L + A  LVAC Q +   +TT S       V LVVKEDTN + EKV + KG
Sbjct: 1   VNKRFKTGFLALVAMLLVACSQGTKQIQTTPSVPKADHHVRLVVKEDTNTVDEKVSFGKG 60

Query: 61  DTVLDVLKANYKVKEKDGFITSIDGISQDETKGLYWMFKVNNKLAPKAANQIKVKKNDKI 120
           DTVL+VLK NY+VKEKDGFIT+IDGI QD     YW+FKVN K+A K A+QI VK  D I
Sbjct: 61  DTVLEVLKDNYEVKEKDGFITAIDGIEQDTKANKYWLFKVNGKMADKGADQITVKDGDSI 120

Query: 121 EFYQEVY 127
           EFYQEV+
Sbjct: 121 EFYQEVF 127
```

SEQ ID 1046 (GBS185) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 28 (lane 6; MW 15.7kDa).

GBS185-His was purified as shown in Figure 199, lane 8.

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

Example 324

A DNA sequence (GBSx0354) was identified in *S.agalactiae* <SEQ ID 1049> which encodes the amino acid sequence <SEQ ID 1050>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -4.46   Transmembrane   75 -  91 (  67 -  94)
   INTEGRAL    Likelihood = -4.41   Transmembrane   33 -  49 (  30 -  49)
   INTEGRAL    Likelihood = -2.60   Transmembrane   53 -  69 (  52 -  70)
   INTEGRAL    Likelihood = -1.38   Transmembrane  108 - 124 ( 106 - 124)
   INTEGRAL    Likelihood = -0.06   Transmembrane  149 - 165 ( 149 - 165)

----- Final Results -----
              bacterial membrane --- Certainty=0.2784(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 9731> which encodes amino acid sequence <SEQ ID 9732> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10929> which encodes amino acid sequence <SEQ ID 10930> 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 1051> which encodes the amino acid sequence <SEQ ID 1052>. Analysis of this protein sequence reveals the following:

```
     Possible site: 48
>>> Seems to have a cleavable N-term signal seq.
```

```
   INTEGRAL    Likelihood = -7.96   Transmembrane   50 -  66 (  49 -  71)
   INTEGRAL    Likelihood = -5.73   Transmembrane  101 - 117 (  99 - 124)
   INTEGRAL    Likelihood = -4.41   Transmembrane  141 - 157 ( 139 - 159)
   INTEGRAL    Likelihood = -4.25   Transmembrane   73 -  89 (  67 -  92)

----- Final Results -----
               bacterial membrane --- Certainty=0.4185(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
 Identities = 82/163 (50%), Positives = 120/163 (73%), Gaps = 3/163 (1%)

Query: 10  LTRVAILSALCVVLRYAFAPLPNIQPITAIFLITVVLFDLKEGVATVTITMLVSSFLMGF 69
           ++R+AI+SALCVVLR  F+ LPN+QP+TA  L  ++ F L E V  + + + +S+FL+GF
Sbjct: 6   MSRIAIMSALCVVLRMVFSSLPNVQPVTAFLLSYLLYFGLAEAVLVMMLCLFLSAFLLGF 65

Query: 70  GPWVFLQIISFTLILCLWKFLIYPLTKAVCFGKITEVVLQTFFAGGLGVVYGVIIDTCFA 129
           GPWVF Q+  F L+L LW+F++YPL++    F K  ++  Q F      G++YGV+IDTCFA
Sbjct: 66  GPWVFWQVTCFVLVLLLWRFVLYPLSQQ--FPKY-QLGCQAFLVALCGLLYGVLIDTCFA 122

Query: 130 WLYHMPWWTYVLAGLSFNMAHALSTCLFYPLLLPILRRFRNEK 172
           +LY MPWW+YVLAG+ FN+AHALST +F+P+++ + RR   E+
Sbjct: 123 YLYSMPWWSYVLAGMPFNIAHALSTLVFFPVVMMLFRRLIGEQ 165
```

A related GBS gene <SEQ ID 8549> and protein <SEQ ID 8550> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      6.79
GvH: Signal Score (-7.5): -0.91
     Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 3 value:  -4.46 threshold:  0.0
   INTEGRAL    Likelihood = -4.46   Transmembrane   35 -  51 (  29 -  54)
   INTEGRAL    Likelihood = -1.38   Transmembrane   68 -  84 (  66 -  84)
   INTEGRAL    Likelihood = -0.06   Transmembrane  109 - 125 ( 109 - 125)
   PERIPHERAL  Likelihood =  7.53       88
 modified ALOM score:   1.39

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2784(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:

```
ORF01220(421 - 552 of 1002)
GP|9950155|gb|AAG07353.1|AE004814_8|AE004814(16   -   56   of   69)   hypothetical   protein
(Pseudomonas aeruginosa)
%Match = 3.2
%Identity = 39.5  %Similarity = 60.5
Matches = 17  Mismatches = 15  Conservative Sub.s = 9

222       252       282       312       342       372       402       432
STLTKLTRVAILSALCVVLRYAFAPLPNIQPITAIFLITVVLFDLKEGVATVTITMLVSSFLMGFGPWVFLQIISFTLIL
                                                                        |:::
                                                                    MDPELFEEWMMTGLVTVLI
                                                                           10

462       492       522       552       582       612       642       672
CLWKFLIYPLTKAVCFGKITEVVLQTFFAGGLGVVYGVIIDTCFAWLYHMPWWTYVLAGLSFNMAHALSTCLFYPLLLPI
```

```
 :  |::: | |    ||   ::|  ||| ||||  | ||
LFMAFIVWDLAKKSKAGKFGTLIL--FFALGLGV-LGFIIKGLVIGSLEGAGM
         30        40          50         60
```

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

Example 325

A DNA sequence (GBSx0355) was identified in *S.agalactiae* <SEQ ID 1053> which encodes the amino acid sequence <SEQ ID 1054>. This protein is predicted to be endolysin. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have a cleavable N-term signal seq.

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

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

```
>GP:CAA72266 GB:Y11477 endolysin [Bacteriophage Bastille]
 Identities = 64/210 (30%), Positives = 95/210 (44%), Gaps = 15/210 (7%)

Query: 66  KPIIDVSGWQLPKEIDYDTLSKNISGVVIRVFGGSKISKTNNAAYTTGIDKSFKTHIKEF 125
           K I+D+S      +ID+DT    +S  + R   G + +  +N      +D+ +KT +
Sbjct: 12  KTIVDISHHNA--DIDFDTAKNYVSMFIARTGDGHRYN--SNGELQGVVDRKYKTFVANM 67

Query: 126 QKRNIPVAVYSYALGSSVKEMKEEAQIFYKNAAPYKPTFYWIDVEEETMSNMNKGVQAFR 185
           + R IP   Y +   S V   K+EA+ F+ N      T +  D E  T  NM + +Q F
Sbjct: 68  KARGIPFGNYMFNRFSGVASAKQEAEFFW-NYGDKDATVWVCDAEVSTAPNMKECIQVFI 126

Query: 186 KELKRLGAKNVGIYIGTYFMTEQGISVKGFDAVWIPTYGSDSGYYEAAPQTELKYDLHQY 245
             LK LGAK VG+YIG +   E G      D  WIP YG+   +           DL Q+
Sbjct: 127 DRLKELGAKKVGLYIGHHKYQEFGGKDVNCDFTWIPRYGNKPAF---------ACDLWQW 177

Query: 246 TSQGYLPGFNQPLDLNQIAVNKDKKKTYEK 275
           T  G + G  +  D+N +  +K      EK
Sbjct: 178 TEYGNIAGIGK-CDINVLYGDKPMSFFTEK 206
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1055> which encodes the amino acid sequence <SEQ ID 1056>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-16.98   Transmembrane    8 -  24 (   3 -  28)

----- 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>
```

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

```
 Identities = 198/278 (71%), Positives = 235/278 (84%)

Query: 1   MRRRIKPIVVAVFFSLFGLLLIIGHLHSTNTLKKELVEAKKTIPSVKASKVPQKSTSSKD 60
           MRR+IKPIVV VFF L  ++LIIG   + +  +KE+ +AK  IP   ++    K+++S+
Sbjct: 1   MRRKIKPIVVLVFFILLAMVLIIGKRQANHAKQKEVEDAKSHIPIATSNPGKAKTSTSET 60

Query: 61  KEFVLKPIIDVSGWQLPKEIDYDTLSKNISGVVIRVFGGSKISKTNNAAYTTGIDKSFKT 120
           ++F+L PI+DVSGWQLP+EIDYDTLS++ISG ++RV+GGS+I+  NNAA+TTGIDKSFKT
```

```
Sbjct: 61  EDFILNPIVDVSGWQLPEEIDYDTLSRHISGAIVRVYGGSQITAHNNAAFTTGIDKSFKT 120

Query: 121 HIKEFQKRNIPVAVYSYALGSSVKEMKEEAQIFYKNAAPYKPTFYWIDVEEETMSNMNKG 180
           HIKEFQKRN+PVAVYSYALG S KEMKEEA+ FYKNAAPY PT+YWIDVEE TM +MNKG
Sbjct: 121 HIKEFQKRNVPVAVYSYALGRSTKEMKEEARAFYKNAAPYNPTYYWIDVEEATMKDMNKG 180

Query: 181 VQAFRKELKRLGAKNVGIYIGTYFMTEQGISVKGFDAVWIPTYGSDSGYYEAAPQTELKY 240
           V AFR+ELK+LGA+NVG+YIGTYFM EQ IS KGFD+VWIPTYGSDSGYYEAAP T L Y
Sbjct: 181 VTAFREELKKLGAENVGLYIGTYFMAEQDISTKGFDSVWIPTYGSDSGYYEAAPNTTLDY 240

Query: 241 DLHQYTSQGYLPGFNQPLDLNQIAVNKDKKKTYEKLFG 278
           DLHQYTSQGYL GFN  LDLNQIAV KD KKT+EKLFG
Sbjct: 241 DLHQYTSQGYLSGFNNALDLNQIAVTKDTKKTFEKLFG 278
```

A related GBS gene <SEQ ID 8551> and protein <SEQ ID 8552> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 5
McG: Discrim Score:     13.20
GvH: Signal Score (-7.5): -0.72
     Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   7.05 threshold:  0.0
   PERIPHERAL  Likelihood =  7.05     196
 modified ALOM score:  -1.91

*** 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:

```
32.4/47.3% over 194aa
                                                                  Bacteriophage
Bastille
  GP|1865711| endolysin Insert characterized

ORF01218(496 - 1125 of 1446)
GP|1865711|emb|CAA72266.1||Y11477(12 - 206 of 364) endolysin (Bacteriophage Bastille)
%Match = 7.9
%Identity = 32.3  %Similarity = 47.3
Matches = 65  Mismatches = 100  Conservative Sub.s = 30

315       345       375       405       435       465       495       525
VTISIMRRRIKPIVVAVFFSLFGLLLIIGHLHSTNTLKKELVEAKKTIPSVKASKVPQKSTSSKDKEFVLKPIIDVSGWQ
                                                          :|    | |:|:|  :
                                                        MALEANKYPKEKTIVDIS--H
                                                              10

555       585       615       645       675       705       735       765
LPKEIDYDTLSKNISGVVIRVFGGSKISKTNNAAYTTGIDKSFKTHIKEFQKRNIPVAVYSYALGSSVKEMKEEAQIFYK
   :||:|| :||   :  |   |       :|       :|: :|| :   : | ||   | :   | |   |:||: |:
HNADIDFDT-AKNYVSMFIARTGDGHRYNSN-GELQGVVDRKYKTFVANMKARGIPFGNYMFNRFSGVASAKQEAEFFW-
           30        40        50        60        70        80        90

795       825       855       885       915       945       975      1005
NAAPYKPTFYWIDVEEETMSNMNKGVQAFRKELKRLGAKNVGIYIGTYFMTEQGISVKGFDAVWIPTYGSDSGYYEAAPQ
|      | :  | |  |  || : :| |   || |||| ||:||| :   | |       |  ||| ||         :
NYGDKDATVWVCDAEVSTAPNMKECIQVFIDRLKELGAKKVGLYIGHHKYQEFGGKDVNCDFTWIPRYG---------NK
            110       120       130       140       150       160

1035      1065      1095      1125      1155      1185      1215      1245
TELKYDLHQYTSQGYLPGXNQPLDLNQIAVNKDKKKTYEKLFGKVKE*KLLLTVAFLINYLLFNSSIERIFWVGFFLSVV
  :  || |:|  | : |  :  |:| :  :|      ||   |      | |       :   |:  :    : : :
```

```
PAFACDLWQWTEYGNIAGIGK-CDINVLYGDKPMSFFTEKEGAKETLVPALNKVVTYEVGTNLIPEIQDKLAFLGYEARI
          180        190        200        210        220        230        240
```

SEQ ID 8552 (GBS206) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 6; MW 31.7kDa).

GBS206-His was purified as shown in Figure 206, lane 6.

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

Example 326

A DNA sequence (GBSx0356) was identified in *S.agalactiae* <SEQ ID 1057> which encodes the amino acid sequence <SEQ ID 1058>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.44   Transmembrane  183 - 199 ( 183 - 200)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9729> which encodes amino acid sequence <SEQ ID 9730> was also identified.

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

```
>GP:AAG20117 GB:AE005090 NADH dehydrogenase/oxidoreductase-like
          protein; NolA [Halobacterium sp. NRC-1]
 Identities = 38/156 (24%), Positives = 83/156 (52%), Gaps = 13/156 (8%)

Query: 19  TMEILIAGGSGFLGKQIIKAALTKGHKVAYLSRHEGKGDIFKDPRLTYIRGDITEADKIH 78
           +M++L+ GG+GF+G  + +     +GH V    +R       +  D  +T I GD+T  + +
Sbjct: 8   SMDVLVTGGTGFIGTHLCRELDDRGHDVTAFAREPADAALPAD--VTRIVGDVTVKETVA 65

Query: 79  LEDRTFDILIDCIGA---IKPNQLD----ELNVKATQKAVALCHKNQIPKLVYISA---- 127
                D +++ +       KP+  D    ++++  T+  VA    +  +++ ++ +SA
Sbjct: 66  NAIDGHDAVVNLVALSPLFKPSGGDSRHLDVHLGGTENVVAAASEAGVEYILQLSALDAD 125

Query: 128 NSGYSAYIKSKRKAEQIIKASGLDYLFVRPGLMYGE 163
            +G +AY+++K +AE+ +++S L +  VRP +++G+
Sbjct: 126 PTGPTAYLRAKGRAEEAVRSSDLHHTIVRPSVVFGD 161
```

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

A related GBS gene <SEQ ID 8553> and protein <SEQ ID 8554> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 5
McG: Discrim Score:     -7.99
GvH: Signal Score (-7.5): -6.34
     Possible site: 41
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -1.44 threshold:  0.0
   INTEGRAL    Likelihood = -1.44   Transmembrane  183 - 199 ( 183 - 200)
   PERIPHERAL  Likelihood =  4.29       20
 modified ALOM score:   0.79

*** Reasoning Step: 3
```

```
----- 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>

RGD motif 68-70
```

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

```
32.5/54.4% over 274aa
                                                                    Schizosaccharomyces
pombe
  GP|3395590| hypothetical protein Insert characterized
  PIR|T41177|T41177 hypothetical protein SPCC1840.09 - fission yeast Insert characterized

ORF01216(358 - 990 of 1272)
GP|3395590|emb|CAA20132.1||AL031179(1    -    275    of    276)    hypothetical    protein
(Schizosaccharomyces pombe) PIR|T41177|T41177 hypothetical protein SPCC1840.09 - fission
yeast (Schizosaccharomyces pombe)
%Match = 7.3
%Identity = 32.4  %Similarity = 54.3
Matches = 71  Mismatches = 88  Conservative Sub.s = 48

144       174       204       234       264       294       324       354
*L**ISTDS*K*A*IPFQGIMIINIATVLFGMLN*KFYK*LNMKCPDVMT*NHTVVRY*TITLTRHIKISILNLQNEGEG

384       414       444       474       504       534       564
TMEILIAGGSGFLGKQIIKAALTKGHKVAYLSRHEGKGDIFKDPRLTYIRGDITEADKIHLEDRTFDILIDCIGAI----
 |:|:: ||||||  | | |: ||::|  :||    |   |:| :  :  :  :| |    :  : :| |    :
 MKIVVLGGSGFLGHNICKLAIAKGYEVVSVSRRGAGGLHNKEPWMDDVEWETLDAQK--DPNSLLPVLRDASAVVNSVG
         10        20        30        40        50        60        70

585       615                 648       678
-----------------------------------KPNQLDELNVKATQKAV---------ALCHKNQIPKLVYIS
                                   || :  :   | | :|:          :  |  :|   |:|
ILMENNYKKILQNPRGPVSHLINSLSSNMFKTGQNPLAPKPEEAKQSKNKVTFEAINRDLAIETAKIAAKANVPVYCYVS
           90       100       110       120       130       140       150

699       726       753       783       810       840       846       876
ANS---GYSA-YIKSKRKAE-QIIKASGLDYLFVRPGLMYG-EERPLSIFQAKCIKLFSHL--------PFLGIVVQKVF
|::   |    |||:||:|| :| | | |  :|:|||:||   :||::   |    : | :          |||    : :
AHAAAPGLDPRYIKTKREAEREISKISNLRSIFLRPGFMYNFNDRPFTGALASLFTVSSSINRATSGALNFLGTASAEPL
           170       180       190       200       210       220       230

930       960       990      1020      1050      1080      1110
PTK-VVIVA-EAIVTTLRKKPTQKILSIEELNNK*FIKKATVNSSFYSFTFPKSFS*VFFLSLLTAI*FKSSG*LXPGR*
|:: | : | |||     | | : |  :: : :| | :|:
PSEEVALAALEAISDPSVKGPVE-ISELKSMAHK-FKQKSL
           250       260       270
```

SEQ ID 8554 (GBS303) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 5; MW 28.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 5; MW 53.2kDa).

The GBS303-GST fusion product was purified (Figure 207, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 275), 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 327

A DNA sequence (GBSx0357) was identified in *S.agalactiae* <SEQ ID 1059> which encodes the amino acid sequence <SEQ ID 1060>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2850(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:AAC36853 GB:L23802 regulatory protein [Enterococcus faecalis]
 Identities = 61/164 (37%), Positives = 96/164 (58%), Gaps = 13/164 (7%)

Query: 1   MSKKNKIKKTLVDQILDKAKIEH---------DSLQLDALQGDLPNGIQKQDIFKTLALI 51
           M+KK   +KT   +++++ K+ +          D L  +++   L  GI+K  IFKTL  +
Sbjct: 1   MAKKKTQQKTNAMRMVEQHKVPYKEYEFAWSEDHLSAESVAESL--GIEKGRIFKTLVTV 58

Query: 52  GDKTGPIIGILPLTEHLSEKKLAKISGNKKVQMIPQKDLQKITGYIHGANNPIGIRQKHN 111
           G+KTGP++ ++P  + L  KKLAK SGNKKV+M+  KDL+  TGYI G  +P G+  K
Sbjct: 59  GNKTGPVVAVIPGNQELDLKKLAKASGNKKVEMLHLKDLEATTGYIRGGCSPTGM--KKQ 116

Query: 112 YPIFIDTIALEKQELIVSAGEIGRSIRINSEVLADFVNAKFADI 155
           +P ++   A +   +IVSAG+ G  I +  E +    N +FA+I
Sbjct: 117 FPTYLAEEAQQYSAIIVSAGKRGMQIELAPEAILSLTNGQFAEI 160
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1061> which encodes the amino acid sequence <SEQ ID 1062>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2651(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 = 114/157 (72%), Positives = 139/157 (87%)

Query: 1   MSKKNKIKKTLVDQILDKAKIEHDSLQLDALQGDLPNGIQKQDIFKTLALIGDKTGPIIG 60
           M+KK K+KKTLV+QILDKA I H  L+L+AL+GD P+ +Q  DI+KTLAL GD+TGP+IG
Sbjct: 1   MAKKTKLKKTLVEQILDKANIAHQGLKLNALEGDFPDDLQPSDIYKTLALTGDQTGPLIG 60

Query: 61  ILPLTEHLSEKKLAKISGNKKVQMIPQKDLQKITGYIHGANNPIGIRQKHNYPIFIDTIA 120
           I+PLTEHLSEK+LAK+SGNKKV M+PQKDLQK TGYIHGANNP+GIRQKH+YPIFID  A
Sbjct: 61  IIPLTEHLSEKQLAKVSGNKKVSMVPQKDLQKTTGYIHGANNPVGIRQKHSYPIFIDQTA 120

Query: 121 LEKQELIVSAGEIGRSIRINSEVLADFVNAKFADIKE 157
           LEK ++IVSAGE+GRSI+I+S+ LADFV A FAD+K+
Sbjct: 121 LEKGQIIVSAGEVGRSIKISSQALADFVGASFADLKK 157
```

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

Example 328

A DNA sequence (GBSx0358) was identified in *S.agalactiae* <SEQ ID 1063> which encodes the amino acid sequence <SEQ ID 1064>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4719(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 8555> which encodes amino acid sequence <SEQ ID 8556> was also identified. This protein belongs to the glycolysis/gluconeogenesis pathway, and such proteins have been experimentally detected as surface-exposed in *Streptococci*. The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD36444 GB:AE001791 phosphoglycerate mutase [Thermotoga maritima]
 Identities = 65/191 (34%), Positives = 93/191 (48%), Gaps = 13/191 (6%)

Query: 5   MKFYLVRHGKTQWNLEGRFQGANGDSPLLEEAIEELEELGQYLSSIHFDAVYSSDLGRAR 64
           MK YL+RHG+T WN +G +QG   D PL E   E+  +L   L  +  DA+YSS L R+
Sbjct: 1   MKLYLIRHGETIWNEKGLWQGVT-DVPLNERGREQARKLANSLKRV--DAIYSSPLKRSL 57

Query: 65  DTVNILNDANSCPKEIHYTPQLREWALGTLEGCKIATMQAIYPRQMTAFYQNPLQFKHDM 124
           +T   + A    KEI     LRE +   G +      YP +   + +P      M
Sbjct: 58  ETAEEI--ARRFEKEIIVEEDLRECEISLWNGLTVEEAIREYPVEFKKWSSDP---NFGM 112

Query: 125 FGAESLYQTTHRVESFLRSLASK----NYDKVLIVGHGANLTASIRSLLGYQYGSLHYKD 180
            G ES+    +RV   +  + S+      + V+IV H  +L A I  +LG     LH
Sbjct: 113 EGLESMRNVQNRVVKAIMKIVSQEKLNGSENVVIVSHSLSLRAFICWILGLPL-YLHRNF 171

Query: 181 KLDNASLTIIE 191
           KLDNASL+++E
Sbjct: 172 KLDNASLSVVE 182
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1065> which encodes the amino acid sequence <SEQ ID 1066>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3628(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 = 127/205 (61%), Positives = 152/205 (73%)

Query: 5   MKFYLVRHGKTQWNLEGRFQGANGDSPLLEEAIEELEELGQYLSSIHFDAVYSSDLGRAR 64
           MK Y VRHGKT WNLEGRFQGA GDSPLLEEA +E+  LG+ L+ + FDAVY+SDL RA
Sbjct: 1   MKLYFVRHGKTLWNLEGRFQGAGGDSPLLEEAKDEIHLLGKELAKVAFDAVYTSDLQRAM 60

Query: 65  DTVNILNDANSCPKEIHYTPQLREWALGTLEGCKIATMQAIYPRQMTAFYQNPLQFKHDM 124
            T  I+ DA      ++++T QLREW LG LEG KIATM AIYP+QM AF +N  QFK D
Sbjct: 61  ATAAIILDAFDQQPKLYHTDQLREWRLGKLEGAKIATMAAIYPQQMLAFRENLAQFKPDQ 120

Query: 125 FGAESLYQTTHRVESFLRSLASKNYDKVLIVGHGANLTASIRSLLGYQYGSLHYKDKLDN 184
           F AES+YQTT RV   ++S   K+Y  VLIVGHGANLTA+IRSLLG++   L  K  LDN
Sbjct: 121 FEAESIYQTTQRVCHLIQSFKDKHYQNVLIVGHGANLTATIRSLLGFEPALLLAKGGLDN 180

Query: 185 ASLTIIETHDFKDFNCLTWNDKSYL 209
           ASLTI+ET D+  ++CL WNDKS+L
Sbjct: 181 ASLTILETKDYLTYDCLIWNDKSFL 205
```

SEQ ID 8556 (GBS314) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 4; MW 27.2kDa), in Figure 169 (lane 15-17; MW 41.6kDa) and in Figure 239 (lane 4; MW 41.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 4; MW 52.1kDa).

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

Example 329

A DNA sequence (GBSx0359) was identified in *S.agalactiae* <SEQ ID 1067> which encodes the amino acid sequence <SEQ ID 1068>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3014(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:CAB12562 GB:Z99108 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 69/232 (29%), Positives = 108/232 (45%), Gaps = 9/232 (3%)

Query: 4   SIVFDVDDTIYDQQAPYRIAVEKCFPDFDMSAINQAYIRFRHYSDIGFPRVMAGEWTTEY 63
           +++FDVDDTI D QA   +A+   F D ++   N    +++  +    +     G+ T +
Sbjct: 6   TLLFDVDDTILDFQAAEALALRLLFEDQNIPLTNDMKAQYKTINQGLWRAFEEGKMTRDE 65

Query: 64  FRFWRCKETLLEFGYREIDEATGIYFQEIYEHELENITMLDEMRMTLDFLKSKNVPMGII 123
               R    L E+GY    EA G   ++ Y   LE    L +    L     +   + I+
Sbjct: 66  VVNTRFSALLKEYGY----EADGALLEQKYRRFLEEGHQLIDGAFDLISNLQQQFDLYIV 121

Query: 124 TNGPTEHQLKKVKKLGLYDYVDPKRVIVSQATGFQKPEKEIFNLAAEQF-DMNPSTTLYV 182
           TNG + Q K+++  GL+ +   K + VS+ TGFQKP KE FN   E+    +   TL +
Sbjct: 122 TNGVSHTQYKRLRDSGLFPFF--KDIFVSEDTGFQKPMKEYFNYVFERIPQFSAEHTLII 179

Query: 183 GDSYDNDIMGAFNGGWHSMWFNHRGRSLKPGIKPVYDVAIDNFEQLFGAVKV 234
           GDS   DI G   G  + W N   +   P I P Y+  I   E+L+  + +
Sbjct: 180 GDSLTADIKGGQLAGLDTCWMNPDMKPNVPEIIPTYE--IRKLEELYHILNI 229
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1069> which encodes the amino acid sequence <SEQ ID 1070>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3216(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 = 276/300 (92%), Positives = 292/300 (97%)

Query: 1   MITSIVFDVDDTIYDQQAPYRIAVEKCFPDFDMSAINQAYIRFRHYSDIGFPRVMAGEWT 60
           MIT+IVFDVDDTIYDQQAPYRIA+EKCFPDFDMS +NQAYIRFRHYSD+GFPRVMAGEWT
Sbjct: 1   MITAIVFDVDDTIYDQQAPYRIAMEKCFPDFDMSVMNQAYIRFRHYSDVGFPRVMAGEWT 60

Query: 61  TEYFRFWRCKETLLEFGYREIDEATGIYFQEIYEHELENITMLDEMRMTLDFLKSKNVPM 120
           TEYFRFWRCKETLLEFGYREIDEA G++FQE+YEHELENITMLDEMRMTLDFLKSKNVPM
```

```
Sbjct: 61  TEYFRFWRCKETLLEFGYREIDEAAGVHFQEVYEHELENITMLDEMRMTLDFLKSKNVPM 120

Query: 121 GIITNGPTEHQLKKVKKLGLYDYVDPKRVIVSQATGFQKPEKEIFNLAAEQFDMNPSTTL 180
           GIITNGPTEHQLKKV+KLGLYDY+D KRVIVSQATGFQKPEKEIFNLAAEQFDMNP TTL
Sbjct: 121 GIITNGPTEHQLKKVRKLGLYDYIDAKRVIVSQATGFQKPEKEIFNLAAEQFDMNPQTTL 180

Query: 181 YVGDSYDNDIMGAFNGGWHSMWFNHRGRSLKPGIKPVYDVAIDNFEQLFGAVKVLFDLPD 240
           YVGDSYDNDIMGAFNGGWHSMWFNHRGR LKPG KPVYDVAIDNFEQLFGAVKVLFDLPD
Sbjct: 181 YVGDSYDNDIMGAFNGGWHSMWFNHRGRQLKPGTKPVYDVAIDNFEQLFGAVKVLFDLPD 240

Query: 241 NKFIFDINDKSNPVLEMGLNNGLMMAAERLLESNMSVDKVVILLRLTAKQEKVLRMKYAR 300
           NKFIFD+NDK NP+L+MG+NNGLMMAAERLLESNMS+DKVVILLRLT +QEKVLR+KYAR
Sbjct: 241 NKFIFDVNDKKNPILQMGINNGLMMAAERLLESNMSIDKVVILLRLTKQQEKVLRLKYAR 300
```

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

Example 330

A DNA sequence (GBSx0360) was identified in *S.agalactiae* <SEQ ID 1071> which encodes the amino acid sequence <SEQ ID 1072>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2451(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 9727> which encodes amino acid sequence <SEQ ID 9728> was also identified.

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

```
>GP:CAB11858 GB:Z99104 lysyl-tRNA synthetase [Bacillus subtilis]
 Identities = 318/490 (64%), Positives = 390/490 (78%), Gaps = 1/490 (0%)

Query: 44  EELNDQQIVRREKMAALTEQGIDPFGKRFERTATSGQLNEKYADKSKEDLHDIEETATIA 103
           EELNDQ  VRR+KM  L + GIDPFG RFERT  S ++   Y D +KE+L +     TIA
Sbjct: 9   EELNDQLQVRRDKMNQLRDNGIDPFGARFERTHQSQEVISAYQDLTKEELEEKAIEVTIA 68

Query: 104 GRLMTKRGKGKVGFAHIQDREGQIQIYVRKDSVGEENYEIFKKADLGDFLGVEGQVMRTD 163
           GR+MTKRGKGK GFAH+QD EGQIQIYVRKDSVG++ YEIFK +DLGD +GV G+V +T+
Sbjct: 69  GRMMTKRGKGKAGFAHLQDLEGQIQIYVRKDSVGDDQYEIFKSSDLGDLIGVTGKVFKTN 128

Query: 164 MGELSIKATHITHLSKALRPLPEKFHGLTDIETIYRKRHLDLISNRDSFDRFVTRSKIIS 223
           +GELS+KAT     L+KALRPLP+K+HGL D+E  YR+R+LDLI N DS   F+TRSKII
Sbjct: 129 VGELSVKATSFELLTKALRPLPDKYHGLKDVEQRYRQRYLDLIVNPDSKHTFITRSKIIQ 188

Query: 224 EIRRFMDSNGFLEVETPVLHNEAGGASARPFITHHNAQDIDMVLRIATELHLKRLIVGGM 283
            +RR++D +G+LEVETP +H+  GGASARPFITHHNA DI + +RIA ELHLKRLIVGG+
Sbjct: 189 AMRRYLDDHGYLEVETPTMHSIPGGASARPFITHHNALDIPLYMRIAIELHLKRLIVGGL 248

Query: 284 ERVYEIGRIFRNEGMDATHNPEFTSIEAYQAYADYQDIMDLTEGIIQHVTKTVKGDGPIN 343
           E+VYEIGR+FRNEG+   HNPEFT IE Y+AYADY+DIM LTE ++ H+ + V G   I
Sbjct: 249 EKVYEIGRVFRNEGVSTRHNPEFTMIELYEAYADYKDIMSLTENLVAHIAQEVLGTTTIQ 308

Query: 344 YQGTEIKINEPFKRVHMVDAVKEITGIDFWKEMTLEEAQALAQEKNVPLEKHFTTVGHII 403
           Y   +I +   +KR+HMVDAVKE TG+DFW+E+T+E+A+  A+E  V + K   TVGHII
Sbjct: 309 YGEEQIDLKPEWKRIHMVDAVKEATGVDFWEEVTVEQAREYAKEHEVEI-KDSMTVGHII 367

Query: 404 NAFFEEFVEDTLIQPTFVFGHPVEVSPLAKKNDTDPRFTDRFELFIMTKEYANAFTELND 463
           N FFE+ +E+TLIQPTF++GHPVE+SPLAKKN  DPRFTDRFELFI+ +E+ANAFTELND
Sbjct: 368 NEFFEQKIEETLIQPTFIYGHPVEISPLAKKNPEDPRFTDRFELFIVGREHANAFTELND 427
```

```
Query: 464 PIDQLSRFEAQASAKELGDDEATGVDYDYVEALEYGMPPTGGLGIGIDRLCMLLTDTTTI 523
           PIDQ  RFEAQ   +E G+DEA  +D D+VEALEYGMPPTGGLGIGIDRL MLLT+  +I
Sbjct: 428 PIDQRERFEAQLKEREAGNDEAHLMDEDFVEALEYGMPPTGGLGIGIDRLVMLLTNAPSI 487

Query: 524 RDVLLFPTMK 533
           RDVLLFP M+
Sbjct: 488 RDVLLFPQMR 497
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1073> which encodes the amino acid sequence <SEQ ID 1074>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4694(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 = 439/500 (87%), Positives = 474/500 (94%)

Query: 34  LEEIMSNQHIEELNDQQIVRREKMAALTEQGIDPFGKRFERTATSGQLNEKYADKSKEDL 93
           LEE MSNQHIEELNDQQIVRREKM AL EQGIDPFGKRF+RTA S +L EKYADK+KE+L
Sbjct: 1   LEENMSNQHIEELNDQQIVRREKMTALAEQGIDPFGKRFDRTANSAELKEKYADKTKEEL 60

Query: 94  HDIEETATIAGRLMTKRGKGKVGFAHIQDREGQIQIYVRKDSVGEENYEIFKKADLGDFL 153
           H++ ETA +AGRLMTKRGKGKVGFAH+QDREGQIQ+YVRKDSVGE+NYEIFKKADLGDF+
Sbjct: 61  HELNETAIVAGRLMTKRGKGKVGFAHLQDREGQIQLYVRKDSVGEDNYEIFKKADLGDFI 120

Query: 154 GVEGQVMRTDMGELSIKATHITHLSKALRPLPEKFHGLTDIETIYRKRHLDLISNRDSFD 213
           GVEG+VMRTDMGELSIKAT +THLSK+LRPLPEKFHGLTDIETIYRKRHLDLISNR+SFD
Sbjct: 121 GVEGEVMRTDMGELSIKATKLTHLSKSLRPLPEKFHGLTDIETIYRKRHLDLISNRESFD 180

Query: 214 RFVTRSKIISEIRRFMDSNGFLEVETPVLHNEAGGASARPFITHHNAQDIDMVLRIATEL 273
           RFVTRSK+ISEIRR++D    FLEVETPVLHNEAGGA+ARPF+THHNAQ+IDMVLRIATEL
Sbjct: 181 RFVTRSKMISEIRRYLDGLDFLEVETPVLHNEAGGAAARPFVTHHNAQNIDMVLRIATEL 240

Query: 274 HLKRLIVGGMERVYEIGRIFRNEGMDATHNPEFTSIEAYQAYADYQDIMDLTEGIIQHVT 333
           HLKRLIVGGMERVYEIGRIFRNEGMDATHNPEFTSIE YQAYADY DIM+LTEGIIQH
Sbjct: 241 HLKRLIVGGMERVYEIGRIFRNEGMDATHNPEFTSIEVYQAYADYLDIMNLTEGIIQHAA 300

Query: 334 KTVKGDGPINYQGTEIKINEPFKRVHMVDAVKEITGIDFWKEMTLEEAQALAQEKNVPLE 393
           K V+GDGPI+YQGTEI+INEPFKRVHMVDA+KE+TG DFW EMT+EEA ALA+EK VPLE
Sbjct: 301 KAVRGDGPIDYQGTEIRINEPFKRVHMVDAIKEVTGADFWPEMTVEEAIALAKEKQVPLE 360

Query: 394 KHFTTVGHIINAFFEEFVEDTLIQPTFVFGHPVEVSPLAKKNDTDPRFTDRFELFIMTKE 453
           KHF +VGHIINAFFEEFVE+TL+QPTFVFGHPVEVSPLAKKN  D RFTDRFELFIMTKE
Sbjct: 361 KHFISVGHIINAFFEEFVEETLVQPTFVFGHPVEVSPLAKKNPEDTRFTDRFELFIMTKE 420

Query: 454 YANAFTELNDPIDQLSRFEAQASAKELGDDEATGVDYDYVEALEYGMPPTGGLGIGIDRL 513
           YANAFTELNDPIDQLSRFEAQA AKELGDDEATG+DYD+VEALEYGMPPTGGLGIGIDRL
Sbjct: 421 YANAFTELNDPIDQLSRFEAQAQAKELGDDEATGIDYDFVEALEYGMPPTGGLGIGIDRL 480

Query: 514 CMLLTDTTTIRDVLLFPTMK 533
           CMLLT+TTTIRDVLLFPTMK
Sbjct: 481 CMLLTNTTTIRDVLLFPTMK 500
```

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

Example 331

A DNA sequence (GBSx0361) was identified in *S.agalactiae* <SEQ ID 1075> which encodes the amino acid sequence <SEQ ID 1076>. This protein is predicted to be 6,7-dimethyl-8-ribityllumazine synthase (ribH). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1042(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:CAB14257 GB:Z99116 riboflavin synthase (beta subunit) [Bacillus subtilis]
 Identities = 103/151 (68%), Positives = 120/151 (79%)

Query: 1   MTIIEGQLVANEMKIGIVVSRFNELITSKLLSGAVDGLLRHGVSEEDIDIVWVPGAFEIP 60
           M II+G LV   +KIGIVV RFN+ ITSKLLSGA D LLRHGV   DID+ WVPGAFEIP
Sbjct: 1   MNIIQGNLVGTGLKIGIVVGRFNDFITSKLLSGAEDALLRHGVDTNDIDVAWVPGAFEIP 60

Query: 61  YMARKMALYKDYDAIICLGVVIKGSTDHYDYVCNEVTKGIGHLNSQSDIPHIFGVLTTDN 120
           + A+KMA  K YDAII LG VI+G+T HYDYVCNE  KGI    + + +P IFG++TT+N
Sbjct: 61  FAAKKMAETKKYDAIITLGTVIRGATTHYDYVCNEAAKGIAQAANTTGVPVIFGIVTTEN 120

Query: 121 IEQAIERAGTKAGNKGYDCALSAIEMVNLDK 151
           IEQAIERAGTKAGNKG DCA+SAIEM NL++
Sbjct: 121 IEQAIERAGTKAGNKGVDCAVSAIEMANLNR 151
```

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 332

A DNA sequence (GBSx0362) was identified in *S.agalactiae* <SEQ ID 1077> which encodes the amino acid sequence <SEQ ID 1078>. This protein is predicted to be GTP cyclohydrolase ii (ribA/B). Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1918(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 9725> which encodes amino acid sequence <SEQ ID 9726> was also identified.

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

```
>GP:AAA86524 GB:U27202 GTP cyclohydrase II/
          3,4-dihydroxy-2-butanone-4-phosphate synthase
          [Actinobacillus pleuropneumoniae]
 Identities = 230/395 (58%), Positives = 307/395 (77%)

Query: 19  FSPIKKLLQDIKSGKMVVLMDDENRENEGDLICAAEMVTKESINFMAKFGKGLICLPLSN 78
           FS ++  ++ I+ GK++++ DDE+RENEGD ICAAE  T E+INFMA +GKGLIC P+S
Sbjct: 6   FSKVEDAIEAIRQGKIILVTDDEDRENEGDFICAAEFATPENINFMATYGKGLICTPIST 65
```

```
Query: 79   YYAEKLELAQMASHNTDNHETAFTISIDHLSTSTGISAEDRALTAKMVANDSSKAKDFRR 138
              A+KL     M + N DNHETAFT+S+DH+ T TGISA +R++TA  + +D++KA DFRR
Sbjct: 66   EIAKKLNFHPMVAVNQDNHETAFTVSVDHIDTGTGISAFERSITAMKIVDDNAKATDFRR 125

Query: 139 PGHLFPLLAKEGGVLARNGHTEATVDLCRLAGLKECGLCCEIMAEDGSMMRKDELLAFAQ 198
           PGH+FPL+AKEGGVL RNGHTEATVDL RLAGLK  GLCCEIMA+DG+MM   +L  FA
Sbjct: 126 PGHMFPLIAKEGGVLVRNGHTEATVDLARLAGLKHAGLCCEIMADDGTMMTMPDLQKFAV 185

Query: 199 KHDLAIATIKQLQDYRRQEEGGVVREIEIQLPTQFGHFTAYGYSEVVANKEHVALVKGDI 258
           +H++   TI+QLQ+YRR+ +  V +   +++PT++G F A+ + EV++ KEHVALVKGD+
Sbjct: 186 EHNMPFITIQQLQEYRRKHDSLVKQISVVKMPTKYGEFMAHSFVEVISGKEHVALVKGDL 245

Query: 259 SSGEDVLCRLHSECLTGDVFHSLRCDCGEQLANALQQIEAEGRGVLLYMRQEGRGIGLIN 318
           + GE VL R+HSECLTGD F S RCDCG+Q A A+ QIE EGRGV+LY+RQEGRGIGLIN
Sbjct: 246 TDGEQVLARIHSECLTGDAFGSQRCDCGQQFAAAMTQIEQEGRGVILYLRQEGRGIGLIN 305

Query: 319 KLKAYHLQEEGLDTLEANLALGFEGDERDYGVSAQLLKDLGINSINLLTNNPDKIQQLEA 378
           KL+AY LQ++G+DT+EAN+ALGF+ DER+Y + AQ+ + LG+ SI LLTNNP KI+ L+
Sbjct: 306 KLRAYELQDKGMDTVEANVALGFKEDEREYYIGAQMFQQLGVKSIRLLTNNPAKIEGLKE 365

Query: 379 EGICVKNRVPLQVAVTAYDLNYLKTKKEKMGHLLD 413
           +G+ +  R P+ V      D++YLK K+ KMGH+ +
Sbjct: 366 QGLNIVAREPIIVEPNKNDIDYLKVKQIKMGHMFN 400
```

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

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

Example 333

A DNA sequence (GBSx0363) was identified in *S.agalactiae* <SEQ ID 1079> which encodes the amino acid sequence <SEQ ID 1080>. This protein is predicted to be riboflavin synthase alpha chain (ribE). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3517(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 9723> which encodes amino acid sequence <SEQ ID 9724> was also identified.

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

```
>GP:BAB05274 GB:AP001512 riboflavin synthase alpha subunit [Bacillus halodurans]
 Identities = 98/216 (45%), Positives = 147/216 (67%), Gaps = 2/216 (0%)

Query: 1   MFTGIIEEMGQVSRIRNGIKSQQLSIDAPKLVPLLRKGDSVAVNGVCLTVLDKSETAFIA 60
           MFTGIIE++G +  I+   ++  ++I + K+V  ++ GDS+AVNGVCLTV   ++T F
Sbjct: 1   MFTGIIEDVGTIDAIQQTGEAIVMTITSKKIVSDVQLGDSIAVNGVCLTVTSFTDTQFTV 60

Query: 61  DVMPESMMRTSLAALRLHSKVNLELALRSDSRLGGHFVLGHVDGVGKIEKIQKDDIAVRF 120
           D+MPE++  TSL  L   S+VNLE A+ ++ R GGH V GHVDG+G I K ++ D AV +
Sbjct: 61  DLMPETVRATSLRLLSKGSRVNLERAMVANGRFGGHIVSGHVDGIGTIRKKERKDNAVYY 120

Query: 121 SIDAPPSIMSYIIEKGSVALDGISLTVVSFTEHSFEVSVIPHTMAQTNLSLKKVGDLLNI 180
           +I+   S+  Y+I KGSVA+DG SLT+   ++ +F +S+IPHTM +T + LKK GD++NI
Sbjct: 121 TIEVSSSLRRYMIHKGSVAVDGTSLTIFDVSDKTFTISIIPHTMEETIIGLKKAGDIVNI 180
```

```
Query: 181 EVDVLGKYAEKFLAPTNRTNHTSSVMDWSFLSENGY 216
           E D++GKY E+F+      N     +  +FL+E+GY
Sbjct: 181 ECDLIGKYIEQFVQQGKPVNEGG--LTKAFLTEHGY 214
```

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

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

Example 334

A DNA sequence (GBSx0364) was identified in *S.agalactiae* <SEQ ID 1081> which encodes the amino acid sequence <SEQ ID 1082>. This protein is predicted to be riboflavin-specific deaminase (ribD). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.01   Transmembrane  307 - 323 ( 307 - 323)

----- Final Results -----
               bacterial membrane --- Certainty=0.1404(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAA86522 GB:U27202 riboflavin-specific deaminase [Actinobacillus
           pleuropneumoniae]
 Identities = 182/353 (51%), Positives = 259/353 (72%)

Query: 6   DYMALALKEAEKGMGFVAPNPLVGAVIVKDDRIISKGYHKRFGDLHAERQAIKNADEDIS 65
           DYM  A+  A++G+G+  PNPLVG VIVK+  I+++GYH++ G  HAER A+ +  ED+S
Sbjct: 51  DYMRRAIALAKQGLGWTNPNPLVGCVIVKNGEIVAEGYHEKIGGWHAERNAVLHCKEDLS 110

Query: 66  GSTLYVTLEPCCHVGKQPPCTEALIKSGIKKVVVGSLDPNPLVSGKGIALLRKEGLNVEV 125
           G+T YVTLEPCCH G+ PPC++ LI+ GIKKV +GS DPNPLV+G+G   LR+ G+ V
Sbjct: 111 GATAYVTLEPCCHHGRTPPCSDLLIERGIKKVFIGSSDPNPLVAGRGANQLRQAGVEVVE 170

Query: 126 GILREECDALNERFIFHMTYKQPFVYLKYAMTLDGKIATKTGDSKWISNEHSRQSVQKLR 185
           G+L+EECDALN  F  ++  K+P+V +KYAMT DGKIAT +G+SKWI+ E +R  VQ+ R
Sbjct: 171 GLLKEECDALNPIFFHYIQTKRPYVLMKYAMTADGKIATGSGESKWITGESARARVQQTR 230

Query: 186 QKCSAIMVGINTVLADNPRLTCRIPKGEALVRIVCDSQLKIPLDSYLVKSAKTIPTWIAT 245
            + SAIMVG++TVLADNP L  R+P  +  VRIVCDSQL+ PLD  LV++AK   T IAT
Sbjct: 231 HQYSAIMVGVDTVLADNPMLNSRMPNAKQPVRIVCDSQLRTPLDCQLVQTAKEYRTVIAT 290

Query: 246 CSDNLAQQQTLKEMGCRLIKVPRKDGKLDLKVLMTILGQEGIDSLLIEGGSSLHFSALKA 305
            SD+L + +  + +G  ++     ++ ++DL+ L+  LG+  IDSLL+EGGSSL+FSAL++
Sbjct: 291 VSDDLQKIEQFRPLGVDVLVCKARNKRVDLQDLLQKLGEMQIDSLLLEGGSSLNFSALES 350

Query: 306 GIVNRLIVFIAPKIIGGLKAKTAISGEGLDWLNQAFRVKDIELSRMDSDVVIE 358
           GIVNR+  +IAPK++GG +AKT I GEG+  ++QA ++K      +  D++++
Sbjct: 351 GIVNRVHCYIAPKLVGGKQAKTPIGGEGIQQIDQAVKLKLKSTELIGEDILLD 403
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1083> which encodes the amino acid sequence <SEQ ID 1084>. Analysis of this protein sequence reveals the following:

```
      Possible site: 25

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.17   Transmembrane   88 - 104 (  88 - 105)

----- Final Results -----
               bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
```

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

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

```
>GP:CAB11794 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 71/161 (44%), Positives = 109/161 (67%)

Query: 13  LEEQTYFMQEALKEAEKSLQKAEIPIGCVIVKDGEIIGRGHNAREESNQAIMHAEMMAIN 72
           + +   +M+EA+KEA+K+ +K E+PIG V+V +GEII R HN RE   ++I HAEM+ I+
Sbjct: 1   MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVID 60

Query: 73  EANAHEGNWRLLDTTLFVTIEPCVMCSGAIGLARIPHVIYGASNQKFGGVDSLYQILTDE 132
           EA    G WRL   TL+VT+EPC MC+GA+ L+R+  V++GA + K G   +L  +L +E
Sbjct: 61  EACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEE 120

Query: 133 RLNHRVQVERGLLAADCANIMQTFFRQGRERKKIAKHLIKE 173
           R NH+ +V  G+L  +C  ++  FFR+ R++KK A+  + E
Sbjct: 121 RFNHQAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE 161
```

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

```
 Identities = 48/146 (32%), Positives = 71/146 (47%), Gaps = 21/146 (14%)

Query: 7   YMALALKEAEKGMGFVAPNPLVGAVIVKDDRIISKGYHKRFGD----LHAERQAIKNADE 62
           +M  ALKEAEK +   A  P +G VIVKD  II +G++ R       +HAE  AI  A+
Sbjct: 19  FMQEALKEAEKSLQ-KAEIP-IGCVIVKDGEIIGRGHNAREESNQAIMHAEMMAINEANA 76

Query: 63  D-----ISGSTLYVTLEPCCHVGKQPPCTEALIKSGIKKVVVGSLDPNPLVSGKGIALLR 117
                +  +TL+VT+EPC        C+ A+  + I  V+ G+ +            +L
Sbjct: 77  HEGNWRLLDTTLFVTIEPCV------MCSGAIGLARIPHVIYGASNQKFGGVDSLYQILT 130

Query: 118 KEGLN----VEVGILREECDALNERF 139
            E LN    VE G+L  +C  + + F
Sbjct: 131 DERLNHRVQVERGLLAADCANIMQTF 156
```

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

Example 335

A DNA sequence (GBSx0365) was identified in *S.agalactiae* <SEQ ID 1085> which encodes the amino acid sequence <SEQ ID 1086>. This protein is predicted to be Nramp metal ion transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.89   Transmembrane  169 - 185 ( 160 - 191)
   INTEGRAL    Likelihood =-11.09   Transmembrane  140 - 156 ( 128 - 165)
   INTEGRAL    Likelihood = -6.85   Transmembrane  359 - 375 ( 354 - 379)
   INTEGRAL    Likelihood = -6.48   Transmembrane  269 - 285 ( 263 - 287)
   INTEGRAL    Likelihood = -6.16   Transmembrane  426 - 442 ( 423 - 445)
   INTEGRAL    Likelihood = -5.57   Transmembrane   62 -  78 (  58 -  80)
   INTEGRAL    Likelihood = -4.94   Transmembrane  107 - 123 ( 103 - 127)
   INTEGRAL    Likelihood = -4.46   Transmembrane  391 - 407 ( 389 - 408)
   INTEGRAL    Likelihood = -4.35   Transmembrane  310 - 326 ( 307 - 328)

----- 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 GENPEPT database:

```
>GP:AAF83825 GB:AE003939 manganese transport protein [Xylella
```

```
           fastidiosa]
 Identities = 192/436 (44%), Positives = 274/436 (62%), Gaps = 14/436 (3%)

Query: 10  SLSEVNQSVEVPHNSSFWNTLRAFLGPGALVAVGYMDPGNWITSVIGGATYRYLLLFVVL 69
           SL E++ SV V      +    L AFLGPG +V+VGYMDPGNW T + GG+ + Y+LL V+L
Sbjct: 39  SLGEMHASVAVSRRGHWGFRLLAFLGPGYMVSVGYMDPGNWATGLAGGSRFGYMLLSVIL 98

Query: 70  VSSLMAMQLQQMAGKLGIVTRQDLAQATASRLPKPLRYLLFIIIELALIATDLAEVIGSA 129
           +S++MA+ LQ +A +LGI +  DLAQA  +R  +      L+++ ELA+IA DLAEVIG+A
Sbjct: 99  LSNVMAIVLQALAARLGIASDMDLAQACRARYSRGTTLALWVVCELAIIACDLAEVIGTA 158

Query: 130 IALHLLFGWPLLLSIMITILDVFLLLLLMKLGVQKIEAFVSVLILTILIIFTYLVVLSQP 189
           IAL+LL G P++  ++IT +DV L+LLLM  G + +EAFV  L+L I   F   +VL+ P
Sbjct: 159 IALNLLLGVPIIWGVVITAVDVVLVLLLMHRGFRALEAFVIALLLVIFGCFVVQIVLAAP 218

Query: 190 DLDAMFKGFLPHHELFNISHEGKNSPLTLALGIIGATVMPHNLYLHSSLSQTRRVDYHNK 249
            L  + GF+P  ++            L LA+GI+GATVMPHNLYLHSS+ QTR
Sbjct: 219 PLQEVLGGFVPRWQVV-----ADPQALYLAIGIVGATVMPHNLYLHSSIVQTRAYP-RTP 272

Query: 250 SSIKKAVRFMTLDSNIQLSLAFVVNSLLLVLGASLFYG-HANDISAFSQMYLALSDKTIT 308
              + A+R+   DS + L LA  +N+ +L+L A++F+  H  D+    Q Y  L+
Sbjct: 273 VGRRSALRWAVADSTLALMLALFINASILILAAAVFHAQHHFDVEEIEQAYQLLAPVLGV 332

Query: 309 GAVASSFLSTLFAVALLASGQNSTITGTLTGQIVMEGFLHFKLPQWLIRLCTRLLTLLPI 368
           G  A     TLFA ALLASG NST+T TL GQIVMEGFL  +L  WL R+ TR L ++P+
Sbjct: 333 GVAA-----TLFATALLASGINSTVTATLAGQIVMEGFLRLRLRPWLRRVLTRGLAIVPV 387

Query: 369 FVIALLVGGEENTLDQLIVYSQVFLSLALPFSIFPLIYFTSQKSIMGEHANAKWNTYLAY 428
           V+  L G  E    +L++ SQV LS+ LPF++ PL+   + + +MG      +W   +A+
Sbjct: 388 IVVVALYG--EQGTGRLLLLSQVILSMQLPFAVIPLLRCVADRKVMGALVAPRWLMVVAW 445

Query: 429 LVAIILTLLNLKLIMD 444
           L+A ++ +LN+KL+ D
Sbjct: 446 LIAGVIVVLNVKLLGD 461
```

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 336

A DNA sequence (GBSx0366) was identified in *S.agalactiae* <SEQ ID 1087> which encodes the amino acid sequence <SEQ ID 1088>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-14.12   Transmembrane  113 - 129 (  98 - 132)
   INTEGRAL    Likelihood =-12.15   Transmembrane  228 - 244 ( 220 - 249)
   INTEGRAL    Likelihood =-10.83   Transmembrane  175 - 191 ( 167 - 195)
   INTEGRAL    Likelihood = -5.04   Transmembrane   57 -  73 (  55 -  75)
   INTEGRAL    Likelihood = -3.93   Transmembrane  146 - 162 ( 142 - 166)
   INTEGRAL    Likelihood = -1.38   Transmembrane  199 - 215 ( 199 - 215)
   INTEGRAL    Likelihood = -0.32   Transmembrane   82 -  98 (  82 -  98)

----- Final Results -----
              bacterial membrane --- Certainty=0.6647(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAF11325 GB:AE002018 hypothetical protein [Deinococcus radiodurans]
 Identities = 63/215 (29%), Positives = 108/215 (49%), Gaps = 13/215 (6%)

Query: 11  LLLVFILTIIVNYLSATGFLTGNSQKSLSDRYQTLLTPAPLAFSIWSVIYL-LTFLVILR 69
```

```
           LL   +LT++VNYLS    L GNS   +SDR     TPA L F++W  I+L L    + +
Sbjct: 10  LLAATVLTLVVNYLSNALPLFGNSNAEVSDRLPNAFTPAGLTFTVWGPIFLGLLVFAVYQ 69

Query: 70  AIFSKSQSYQDNFASIFPYFLGLLLVNNIWTVFFTSNLIGLSTIIIFAYCILLV-IIIKI 128
           A+ ++  +  D     +P+ LG LL N  W + F S  IGLS +I+ A   +LV + + +
Sbjct: 70  ALPAQRGARLDRL--FWPFLLGNLL-NVAWLLAFQSLNIGLSVVIMLALLAVLVRLYLSV 126

Query: 129 LS---KNKSKLLLRITFGIHAGWLLVASLVNLAVYLVKI----DFNYPLPKVYIAIIALI 181
            S   +   + L++   ++  W+ VA++ N+  +LV       F      V+ A++ ++
Sbjct: 127 RSLPPQGAERWTLQLPVSLYLAWISVATIANITAFLVSAGVTQSFLGIAGPVWSALLLVV 186

Query: 182 FITVLSLYLARVLQNAYLILSVFWAWLMVFKAHLE 216
              +   +L R    A+  + + WA+  V+ A  E
Sbjct: 187 AAAIGVFFLWRFRDYAFAAV-LLWAFYGVYVARPE 220
```

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 337

A DNA sequence (GBSx0367) was identified in *S.agalactiae* <SEQ ID 1089> which encodes the amino acid sequence <SEQ ID 1090>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3401(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:AAC65352 GB:AE001215 T. pallidum predicted coding region
         TP0352 [Treponema pallidum]
 Identities = 28/64 (43%), Positives = 41/64 (63%)

Query: 3   EFTFEIVEKLLVLSENEKGWTKELNRVSFNGAPAKFDLRTWSPDHTKMGKGITLSNEEFK 62
           +F +E+      LS +  GW+ EL  +S+NG P K+D+R WSPD +KMGKG+TL+  E
Sbjct: 12  DFHYEVTRNWGTLSTSGNGWSLELKSISWNGRPEKYDIRAWSPDKSKMGKGVTLTRAEIV 71

Query: 63  VILD 66
            + D
Sbjct: 72  ALRD 75
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1091> which encodes the amino acid sequence <SEQ ID 1092>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4021(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 = 59/70 (84%), Positives = 64/70 (91%)

Query: 1   MSEFTFEIVEKLLVLSENEKGWTKELNRVSFNGAPAKFDLRTWSPDHTKMGKGITLSNEE 60
           M+EFTF I E LL LSEN+KGWTKELNRVSFNGA AK+D+RTWSPDHTKMGKGITL+NEE
Sbjct: 1   MAEFTFNIEEHLLTLSENDKGWTKELNRVSFNGAEAKWDIRTWSPDHTKMGKGITLTNEE 60
```

```
Query: 61 FKVILDAFRK 70
          FK ILDAFRK
Sbjct: 61 FKTILDAFRK 70
```

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

Example 338

A DNA sequence (GBSx0368) was identified in *S.agalactiae* <SEQ ID 1093> which encodes the amino acid sequence <SEQ ID 1094>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.66   Transmembrane   92 - 108 (  92 - 110)

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

```
>GP:CAB14676 GB:Z99117 similar to protease [Bacillus subtilis]
 Identities = 201/407 (49%), Positives = 277/407 (67%), Gaps = 2/407 (0%)

Query: 4   VKKRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGINYAHARD 63
           + K+PE+L+PAG LEKLK+A+ YGADAVF+GGQ YGLRS A NF++EE+ EG+ +A
Sbjct: 18  ITKKPELLAPAGNLEKLKIAVHYGADAVFIGGQEYGLRSNADNFTIEEIAEGVEFAKKYG 77

Query: 64  AKVYVAANMVTHEGNELGAGPWFRELRDMGLDAVIVSDPALIVICATEAPGLEIHLSTQA 123
           AK+YV  N+  H  N  G   + + L D  +  +IV+DP +I  C   AP +E+HLSTQ
Sbjct: 78  AKIYVTTNIFAHNENMDGLEDYLKALGDANVAGIIVADPLIIETCRRVAPNVEVHLSTQQ 137

Query: 124 SSTNYETFEFWKEMGLTRVVLAREVTMAELAEIRKRTDVEIEAFVHGAMCISYSGRCVLS 183
           S +N++  +FWKE GL RVVLARE +  E+ E++++ D+EIE+F+HGAMCI+YSGRCVLS
Sbjct: 138 SLSNWKAVQFWKEEGLDRVVLARETSALEIREMKEKVDIEIESFIHGAMCIAYSGRCVLS 197

Query: 184 NHMSHRDANRGGCSQSCRWKYDLYDMPFGQERQSLKGEIPEPFSMSAVDMCMIEHIPDMI 243
           NHM+ RD+NRGGC QSCRW YDLY    G    +L GE   PF+MS  D+ +IE IP MI
Sbjct: 198 NHMTARDSNRGGCCQSCRWDYDLYQTD-GANAVALYGEEDAPFAMSPKDLKLIESIPKMI 256

Query: 244 ENGVDSLKIEGRMKSIHYVSTVTNCYKAAVDAYMESPEAFEAIKEDLIDELWKVAQRELA 303
           E G+DSLKIEGRMKSIHYV+TV + Y+  +DAY   PE F  I+++ ++EL K A R+ A
Sbjct: 257 EMGIDSLKIEGRMKSIHYVATVVSVYRKVIDAYCADPENF-VIQKEWLEELDKCANRDTA 315

Query: 304 TGFYYHTPTENEQLFGARRKIPQYKFVGEVVSFDNAKMEATIRQRNVIMEGDRVEFYGPG 363
           T F+  TP   EQ+FG   K   Y FVG V+++D      T++QRN   +GD VEF+GP
Sbjct: 316 TAFFEGTPGYEEQMFGEHAKKTTYDFVGLVLNYDEDTQMVTLQQRNFFKKGDEVEFFGPE 375

Query: 364 FRHFECFIDGLRDAEGNKIDRAPNPMELLTITLPNPVKKGDMIRACK 410
             +F   I+ + D +GN++D A +P++++   L   +   +M+R  K
Sbjct: 376 IENFTHTIETIWDEDGNELDAARHPLQIVKFKLDKKIYPSNMMRKGK 422
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1095> which encodes the amino acid sequence <SEQ ID 1096>. Analysis of this protein sequence reveals the following:

```
     Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.66   Transmembrane   92 - 108 (  92 - 110)

----- 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 homology with the following sequences in the databases:

```
>GP:BAB04993 GB:AP001511 protease [Bacillus halodurans]
 Identities = 201/403 (49%), Positives = 280/403 (68%), Gaps = 4/403 (0%)

Query: 6   KRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGIDYAHARGAK 65
           K+PE+L+PAG+LEKLKVAI YGADAV++GGQ +GLRS A NFS+EE++EG+++A+  GAK
Sbjct: 17  KKPELLAPAGSLEKLKVAIHYGADAVYIGGQEFGLRSNADNFSIEEMREGVEFANKYGAK 76

Query: 66  VYVAANMVTHEGNEIGAGEWFRQLRDMGLDAVIVSDPALIVICSTEAPGLEIHLSTQASS 125
           VYV  N+  H  N  G  E+   L+++G+  +IV+DP +I  C   AP +E+HLSTQ S
Sbjct: 77  VYVTTNIYAHNENMDGLEEYLSALQEVGVTGIIVADPLIIETCKRVAPKVEVHLSTQQSL 136

Query: 126 TNYETFEFWKAMGLTRVVLAREVNMAELAEIRKRTDVEIEAFVHGAMCISYSGRCVLSNH 185
           +N+   +FWK  GL RVVLAREV + E+ E++K  D+EIE FVHGAMCISYSGRCVLSNH
Sbjct: 137 SNWLAVKFWKEEGLHRVVLAREVGLEEMLEMKKHVDIEIETFVHGAMCISYSGRCVLSNH 196

Query: 186 MSHRDANRGGCSQSCRWKYDLYDMPFGGE-RRSLKGEIPEDYSMSSVDMCMIDHIPDLIE 244
           M+ RD+NRGGC QSCRW YDLY+     E     +G++P  Y+MS  D+ +I  IP LIE
Sbjct: 197 MTARDSNRGGCCQSCRWDYDLYEQQDSAEIPLFAEGDVP--YTMSPKDLNLIQAIPQLIE 254

Query: 245 NGVDSLKIEGRMKSIHYVSTVTNCYKAAVGAYMESPEAFYAIKEELIDELWKVAQRELAT 304
            G+DSLK+EGRMKSIHYV+TVT+ Y+  + AY   P+ F  IK E ++EL K A R+ A
Sbjct: 255 AGIDSLKVEGRMKSIHYVATVTSVYRKVIDAYCSDPDNF-KIKREWLEELEKCANRDFAP 313

Query: 305 GFYYGIPTENEQLFGARRKIPQYKFVGEVVAFDSASMTATIRQRNVIMEGDRIECYGPGF 364
            F+ G PT  EQ++G   K  +Y FVG V+ ++  +   T++QRN   +GD +E +GP
Sbjct: 314 QFFEGTPTYKEQMYGIHPKRTKYDFVGLVLDYNEKTGIVTLQQRNHFKQGDEVEFFGPEI 373

Query: 365 RHFETVVKDLHDADGQKIDRAPNPMELLTISLPREVKPGDMIR 407
             F   V+ + D DG ++D A +P++++   + ++V P +M+R
Sbjct: 374 NRFTQTVEKIWDEDGNELDAARHPLQIVKFKVDQKVYPQNMMR 416
```

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

```
 Identities = 386/427 (90%), Positives = 404/427 (94%)

Query: 1   MSNVKKRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGINYAH 60
           MS++KKRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGI+YAH
Sbjct: 1   MSHMKKRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGIDYAH 60

Query: 61  ARDAKVYVAANMVTHEGNELGAGPWFRELRDMGLDAVIVSDPALIVICATEAPGLEIHLS 120
           AR AKVYVAANMVTHEGNE+GAG WFR+LRDMGLDAVIVSDPALIVIC+TEAPGLEIHLS
Sbjct: 61  ARGAKVYVAANMVTHEGNEIGAGEWFRQLRDMGLDAVIVSDPALIVICSTEAPGLEIHLS 120

Query: 121 TQASSTNYETFEFWKEMGLTRVVLAREVTMAELAEIRKRTDVEIEAFVHGAMCISYSGRC 180
           TQASSTNYETFEFWK MGLTRVVLAREV MAELAEIRKRTDVEIEAFVHGAMCISYSGRC
Sbjct: 121 TQASSTNYETFEFWKAMGLTRVVLAREVNMAELAEIRKRTDVEIEAFVHGAMCISYSGRC 180

Query: 181 VLSNHMSHRDANRGGCSQSCRWKYDLYDMPFGQERQSLKGEIPEPFSMSAVDMCMIEHIP 240
           VLSNHMSHRDANRGGCSQSCRWKYDLYDMPFG ER+SLKGEIPE +SMS+VDMCMI+HIP
Sbjct: 181 VLSNHMSHRDANRGGCSQSCRWKYDLYDMPFGGERRSLKGEIPEDYSMSSVDMCMIDHIP 240

Query: 241 DMIENGVDSLKIEGRMKSIHYVSTVTNCYKAAVDAYMESPEAFEAIKEDLIDELWKVAQR 300
           D+IENGVDSLKIEGRMKSIHYVSTVTNCYKAAV AYMESPEAF AIKE+LIDELWKVAQR
Sbjct: 241 DLIENGVDSLKIEGRMKSIHYVSTVTNCYKAAVGAYMESPEAFYAIKEELIDELWKVAQR 300

Query: 301 ELATGFYYHTPTENEQLFGARRKIPQYKFVGEVVSFDNAKMEATIRQRNVIMEGDRVEFY 360
           ELATGFYY  PTENEQLFGARRKIPQYKFVGEVV+FD+A M ATIRQRNVIMEGDR+E Y
Sbjct: 301 ELATGFYYGIPTENEQLFGARRKIPQYKFVGEVVAFDSASMTATIRQRNVIMEGDRIECY 360

Query: 361 GPGFRHFECFIDGLRDAEGNKIDRAPNPMELLTITLPNPVKKGDMIRACKEGLVNLYQND 420
           GPGFRHFE  + L DA+G KIDRAPNPMELLTI+LP  VK GDMIRACKEGLVNLYQ D
Sbjct: 361 GPGFRHFETVVKDLHDADGQKIDRAPNPMELLTISLPREVKPGDMIRACKEGLVNLYQKD 420

Query: 421 GTSKTVR 427
```

```
          GTSKTVR
Sbjct: 421 GTSKTVR 427
```

SEQ ID 1094 (GBS385) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 3; MW 50kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 7; MW 75.7kDa).

The GBS385-GST fusion product was purified (Figure 213, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 312), 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 339

A DNA sequence (GBSx0369) was identified in *S.agalactiae* <SEQ ID 1097> which encodes the amino acid sequence <SEQ ID 1098>. This protein is predicted to be collagenase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2208(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:CAB14677 GB:Z99117 similar to protease [Bacillus subtilis]
 Identities = 92/304 (30%), Positives = 161/304 (52%), Gaps = 5/304 (1%)

Query: 1   MEKIILTATAESIEQVKQLLAIGIDRIYVGEENYGLRLPHSFSDDELREIAKLVHDAGKE 60
           M+K  L  T  S   +  L+  G      VGE+ YGLRL   FS +++ +  ++ H  G +
Sbjct: 1   MKKPELLVTPTSTADILPLIQAGATAFLVGEQRYGLRLAGEFSREDVTKAVEIAHKEGAK 60

Query: 61  LTVACNALMHQEMMDNIKPFLELMKEINVDYLVVGDAGVFYINKRDGYNFKLIYDTSVFV 120
           + VA NA+ H + +  +  +L  + E  VD  V GD  V    +    + KL + T
Sbjct: 61  VYVAVNAIFHNDKVGELGEYLAFLAEAGVDAAVFGDPAVLMAARESAPDLKLHWSTETTG 120

Query: 121 TSSRQVNFWGQHGAVETVLAREIPSEELFKMSENLEFPAEILVYGASVIHHSKRPLLQNY 180
           T+    N+WG+ GA  +VLARE+  + + ++ EN E   EI V+G + +  SKR L+ NY
Sbjct: 121 TNYYTCNYWGRKGAARSVLARELNMDSIVEIKENAEVEIEIQVHGMTCMFQSKRSLIGNY 180

Query: 181 YNF---THITDEKTRERGLFLAEPGDPESHYSIYEDKHGTHIFINNDINMMTKVTELVEH 237
           + +       + K +E G+FL +  + ++ Y I+ED++GTHI   ND+ ++ ++ EL++
Sbjct: 181 FEYQGKVMDIERKKKESGMFLHDK-ERDNKYPIFEDENGTHIMSPNDVCIIDELEELIDA 239

Query: 238 HFTHWKLDGIYCPGDNFVAIAEIFVETARL-IENGTFTQDQAFLFDERIRKLHPKGRGLD 296
               +K+DG+    + + + +++ E   L +EN    + +   + ERI  + P  R +D
Sbjct: 240 GIDSFKIDGVLKMPEYLIEVTKMYREAIDLCVENRDEYEAKKEDWIERIESIQPVNRKID 299

Query: 297 TGFY 300
           TGF+
Sbjct: 300 TGFF 303
```

A related GBS nucleic acid sequence <SEQ ID 10949> which encodes amino acid sequence <SEQ ID 10950> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1099> which encodes the amino acid sequence <SEQ ID 1100>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1716(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 = 245/308 (79%), Positives = 273/308 (88%)

Query: 1   MEKIILTATAESIEQVKQLLAIGIDRIYVGEENYGLRLPHSFSDDELREIAKLVHDAGKE 60
           MEKII+TATAESIEQVK LLA G+DRIYVGE NYGLRLPH+FS DELR+IAKLVHDAGKE
Sbjct: 1   MEKIIITATAESIEQVKALLAAGVDRIYVGEANYGLRLPHNFSYDELRQIAKLVHDAGKE 60

Query: 61  LTVACNALMHQEMMDNIKPFLELMKEINVDYLVVGDAGVFYINKRDGYNFKLIYDTSVFV 120
           LTVACNALMHQ+MMD IKPFL+LM EI VDYLVVGDAGVFY+NKRDGYNFKLIYDTSVFV
Sbjct: 61  LTVACNALMHQDMMDQIKPFLDLMIEIAVDYLVVGDAGVFYVNKRDGYNFKLIYDTSVFV 120

Query: 121 TSSRQVNFWGQHGAVETVLAREIPSEELFKMSENLEFPAEILVYGASVIHHSKRPLLQNY 180
           TSSRQVNFWGQHGAVE+VLAREIPS ELF ++ENLEFPAE+LVYGASVIHHSKRPLL+NY
Sbjct: 121 TSSRQVNFWGQHGAVESVLAREIPSAELFTLAENLEFPAEVLVYGASVIHHSKRPLLENY 180

Query: 181 YNFTHITDEKTRERGLFLAEPGDPESHYSIYEDKHGTHIFINNDINMMTKVTELVEHHFT 240
           Y+FT I DE +RERGLFLAEPGD  SHYSIYED HGTHIFINNDI+MM+K+ EL  H  T
Sbjct: 181 YHFTKIDDEVSRERGLFLAEPGDASSHYSIYEDNHGTHIFINNDIDMMSKLGELYAHGLT 240

Query: 241 HWKLDGIYCPGDNFVAIAEIFVETARLIENGTFTQDQAFLFDERIRKLHPKGRGLDTGFY 300
           HWKLDGIYCPGD+FVAI ++F++   L+E G FTQ++A   D+ +   HP GRGLDTGFY
Sbjct: 241 HWKLDGIYCPGDDFVAITKLFIQAKTLLEAGQFTQEEAEKLDQAVHAHHPAGRGLDTGFY 300

Query: 301 DFDPSTVK 308
           +FDP TVK
Sbjct: 301 EFDPKTVK 308
```

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

Example 340

A DNA sequence (GBSx0371) was identified in *S.agalactiae* <SEQ ID 1101> which encodes the amino acid sequence <SEQ ID 1102>. This protein is predicted to be cDNA EST yk542c12.5 comes from this gene. 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:AAD15622 GB:U75480 unknown [Streptococcus mutans]
 Identities = 69/152 (45%), Positives = 101/152 (66%), Gaps = 12/152 (7%)

Query: 1   MSKLFKTLVISAASGAAAAYFLTTKKGKELRKNAEKFYGEYKENPEEYHQIAKDKASEYS 60
           MSK  KT +I A +GAAAAYFL+T KGK+ +K   + + +YKENP+EYHQ A DK +EY
Sbjct: 1   MSKFLKTAIIGAGTGAAAAYFLSTDKGKQFKKKIHQTFTDYKENPKEYHQYAADKVNEYK 60
```

```
Query: 61  NLAVDTFKDYKGKFESGELTTEDIVSAVKEKSGEVVDFANDFVNQAKSKFSDEDTAKKED 120
           ++AV +FKDYK KFE+GELT ++I+S+VKEK+ +   FAN  ++Q K   +   T +K +
Sbjct: 61  DVAVHSFKDYKDKFETGELTKDNIISSVKEKASQAGKFANSKLSQVKDHLA--QTVEKAE 118

Query: 121 KAP----------ETKVEDIVIDYKENTEDKE 142
             +          + +V+DIVIDY+   + K+
Sbjct: 119 ASTNDAGIPLGEMKAQVDDIVIDYQAEEKTKK 150
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1103> which encodes the amino acid sequence <SEQ ID 1104>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.81   Transmembrane   15 -  31 (  14 -  31)

----- 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>
```

A related sequence was also identified in GAS <SEQ ID 9117> which encodes the amino acid sequence <SEQ ID 9118>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 19
>>> 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>
```

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

```
Identities = 69/140 (49%), Positives = 91/140 (64%), Gaps = 8/140 (5%)

Query: 1   MSKLFKTLVISAASGAAAAYFLTTKKGKELRKNAEKFYGEYKENPEEYHQIAKDKASEYS 60
           M+K FK LVI A SG AAAYFL+T+KGK L+  AEK Y  YKE+P++YHQ AK+K SEYS
Sbjct: 8   MNKSFKNLVIGAVSGVAAAYFLSTEKGKALKNRAEKAYQAYKESPDDYHQFAKEKGSEYS 67

Query: 61  NLAVDTFKDYKGKFESGELTTEDIVSAVKEKSGEVVDFANDFVNQAKSKFSD-EDTAKKE 119
           +LA DTF D K K  SG+LT ED++  +K+K+          FV + K    ++ E   K++
Sbjct: 68  HLARDTFYDVKDKLASGDLTKEDMLDLLKDKT-------TAFVQKTKETLAEVEAKEKQD 120

Query: 120 DKAPETKVEDIVIDYKENTE 139
           D   +   EDI+IDY E  E
Sbjct: 121 DVIIDLNEEDIIIDYTEQDE 140
```

SEQ ID 1102 (GBS164) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 30 (lane 4; MW 17.4kDa).

The GBS164-His fusion product was purified (Figure 115A; see also Figure 200, lane 4) and used to immunise mice (lane 1+2+3 product; 20μg/mouse). The resulting antiserum was used for Western blot, FACS (Figure 115B), 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 341

A DNA sequence (GBSx0372) was identified in *S.agalactiae* <SEQ ID 1105> which encodes the amino acid sequence <SEQ ID 1106>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood =-16.93   Transmembrane    6 -  22 (   1 -  31)

----- Final Results -----
               bacterial membrane --- Certainty=0.7771(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:AAD15621 GB:U75480 unknown [Streptococcus mutans]
 Identities = 88/129 (68%), Positives = 112/129 (86%)

Query: 1   MIEIAVLIIAIAFVVLVLGILFVLKKVSETIEETKQTIKVLTSDVNVTLYQTNEILAKAN 60
           M EIA+LI+AIAF VLV+ ++ +L+K+S+T++E++QT+K+LTSDVNVTLYQTNE+LAKAN
Sbjct: 1   MWEIALLIVAIAFAVLVIYLILLLRKISDTVDESRQTLKILTSDVNVTLYQTNELLAKAN 60

Query: 61  VLVDDVNGKVSTIDPLFVAIADLSESVSDLNLQARHIGQKASSATSSVTKAGSALAIGKA 120
           VLV+DVNGKV TIDPLF AIADLS SVSDLN QAR+ G+K   +T++V KAG+A   GK
Sbjct: 61  VLVEDVNGKVETIDPLFTAIADLSVSVSDLNRQARYFGKKTRKSTANVGKAGAAYTFGKV 120

Query: 121 ASKIFRKKG 129
           ASK+FRKKG
Sbjct: 121 ASKLFRKKG 129
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1107> which encodes the amino acid sequence <SEQ ID 1108>. Analysis of this protein sequence reveals the following:

```
     Possible site: 16
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL     Likelihood = -0.85   Transmembrane   18 -  34 (  17 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.1341(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:AAD15621 GB:U75480 unknown [Streptococcus mutans]
 Identities = 83/128 (64%), Positives = 110/128 (85%)

Query: 6   ISLMIIALAFVALVIFLIIVLKKVSETIDEAKKTISVLTSDVNVTLHQTNDILAKANILV 65
           I+L+I+A+AF  LVI+LI++L+K+S+T+DE+++T+ +LTSDVNVTL+QTN++LAKAN+LV
Sbjct: 4   IALLIVAIAFAVLVIYLILLLRKISDTVDESRQTLKILTSDVNVTLYQTNELLAKANVLV 63

Query: 66  EDVNGKVATIDPLFVAIADLSESLSDLNSQARHFGQKATNATGNVSKAGKLALVGKVASK 125
           EDVNGKV TIDPLF AIADLS S+SDLN QAR+FG+K   +T NV KAG     GKVASK
Sbjct: 64  EDVNGKVETIDPLFTAIADLSVSVSDLNRQARYFGKKTRKSTANVGKAGAAYTFGKVASK 123

Query: 126 VFGKKGEK 133
           +F KKG++
Sbjct: 124 LFRKKGKQ 131
```

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

```
 Identities = 92/131 (70%), Positives = 116/131 (88%)

Query: 1   MIEIAVLIIAIAFVVLVLGILFVLKKVSETIEETKQTIKVLTSDVNVTLYQTNEILAKAN 60
           ++ I+++IIA+AFV LV+ ++ VLKKVSETI+E K+TI VLTSDVNVTL+QTN+ILAKAN
```

```
Sbjct: 3   LVGISLMIIALAFVALVIFLIIVLKKVSETIDEAKKTISVLTSDVNVTLHQTNDILAKAN 62

Query: 61  VLVDDVNGKVSTIDPLFVAIADLSESVSDLNLQARHIGQKASSATSSVTKAGSALAIGKA 120
           +LV+DVNGKV+TIDPLFVAIADLSES+SDLN QARH GQKA++AT +V+KAG     +GK
Sbjct: 63  ILVEDVNGKVATIDPLFVAIADLSESLSDLNSQARHFGQKATNATGNVSKAGKLALVGKV 122

Query: 121 ASKIFRKKGDK 131
           ASK+F KKG+K
Sbjct: 123 ASKVFGKKGEK 133
```

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

Example 342

A DNA sequence (GBSx0373) was identified in *S.agalactiae* <SEQ ID 1109> which encodes the amino acid sequence <SEQ ID 1110>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0462(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 343

A DNA sequence (GBSx0374) was identified in *S.agalactiae* <SEQ ID 1111> which encodes the amino acid sequence <SEQ ID 1112>. This protein is predicted to be prolipoprotein diacylglyceryl transferase (lgt). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.39   Transmembrane  231 - 247 ( 225 - 251)
   INTEGRAL    Likelihood = -7.64   Transmembrane   89 - 105 (  87 - 107)
   INTEGRAL    Likelihood = -5.20   Transmembrane   18 -  34 (  13 -  36)
   INTEGRAL    Likelihood = -1.86   Transmembrane   46 -  62 (  46 -  64)

----- Final Results -----
               bacterial membrane --- Certainty=0.4354(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 9721> which encodes amino acid sequence <SEQ ID 9722> was also identified.

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

```
>GP:AAC80171 GB:U75480 putative prolipoprotein diacylglycerol
          transferase [Streptococcus mutans] (ver 3)
 Identities = 184/257 (71%), Positives = 226/257 (87%)

Query: 2   MINPVAIRLGPFSIRWYAICIVSGMLLAVYLAMKEAPRKNIKSDDILDFILMAFPLSIVG 61
           MINP+AI+LGP +IRWY+ICIV+G++LAVYL ++EAP+KNIKSDD+LDFIL+AFPL+IVG
```

```
Sbjct: 1   MINPIAIKLGPLTIRWYSICIVTGLILAVYLTIREAPKKNIKSDDVLDFILIAFPLAIVG 60

Query: 62  ARIYYVIFEWAYYSKHPVEIIAIWNGGIAIYGGLITGAILLVIFSYRRLINPIDFLDIAA 121
           AR+YYVIF+W YY K+P EI  IW+GGIAIYGGL+TGA++L IFSY R+I PIDFLD+AA
Sbjct: 61  ARLYYVIFDWDYYLKNPSEIPVIWHGGIAIYGGLLTGALVLFIFSYYRMIKPIDFLDVAA 120

Query: 122 PGVMIAQAIGRWGNFINQEAYGRAVKNLNYVPNFIKNQMYIDGAYRVPTFLYESLWNFLG 181
           PGVM+AQ+IGRWGNF+NQEAYG+ V  LNY+P+FI+ QMYIDG YR PTFLYESLWN LG
Sbjct: 121 PGVMLAQSIGRWGNFVNQEAYGKTVTQLNYLPDFIRKQMYIDGHYRTPTFLYESLWNLLG 180

Query: 182 FVIIMSIRHRPRTLKQGEVACFYLVWYGCGRFIIEGMRTDSLYLAGLRVSQWLSVILVII 241
           F+IIM +R RP  LK+GEVA FYL+WYG GRF+IEGMRTDSL  A LRVSQWLSV+LV++
Sbjct: 181 FIIIMILRRRPNLLKEGEVAFFYLIWYGSGRFVIEGMRTDSLMFASLRVSQWLSVLLVVV 240

Query: 242 GIVMIIYRRREQHISYY 258
           G+++++ RRR   I YY
Sbjct: 241 GVILMVIRRRNHAIPYY 257
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1113> which encodes the amino acid sequence <SEQ ID 1114>. Analysis of this protein sequence reveals the following:

```
     Possible site: 28
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.01   Transmembrane  229 - 245 ( 222 - 249)
   INTEGRAL    Likelihood = -6.90   Transmembrane   45 -  61 (  40 -  68)
   INTEGRAL    Likelihood = -4.41   Transmembrane   17 -  33 (  11 -  35)
   INTEGRAL    Likelihood = -4.14   Transmembrane   87 - 103 (  86 - 106)
   INTEGRAL    Likelihood = -0.27   Transmembrane  170 - 186 ( 170 - 186)

----- Final Results -----
               bacterial membrane --- Certainty=0.3803(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:AAC80171 GB:U75480 putative prolipoprotein diacylglycerol
           transferase [Streptococcus mutans] (ver 3)
 Identities = 176/258 (68%), Positives = 217/258 (83%)

Query: 1   MINPIALKCGPLAIHWYALCILSGLVLAVYLASKEAPKKGISSDAIFDFILIAFPLAIVG 60
           MINPIA+K GPL I WY++CI++GL+LAVYL  +EAPKK I SD + DFILIAFPLAIVG
Sbjct: 1   MINPIAIKLGPLTIRWYSICIVTGLILAVYLTIREAPKKNIKSDDVLDFILIAFPLAIVG 60

Query: 61  ARIYYVIFEWSYYVKHLDEIIAIWNGGIAIYGGLITGALVLLAYCYNKVLNPIHFLDIAA 120
           AR+YYVIF+W YY+K+  EI  IW+GGIAIYGGL+TGALVL  + Y +++ PI FLD+AA
Sbjct: 61  ARLYYVIFDWDYYLKNPSEIPVIWHGGIAIYGGLLTGALVLFIFSYYRMIKPIDFLDVAA 120

Query: 121 PSVMVAQAIGRWGNFINQEAYGKAVSQLNYLPSFIQKQMFIEGSYRIPTFLYESLWNLLG 180
           P VM+AQ+IGRWGNF+NQEAYGK V+QLNYLP FI+KQM+I+G YR PTFLYESLWNLLG
Sbjct: 121 PGVMLAQSIGRWGNFVNQEAYGKTVTQLNYLPDFIRKQMYIDGHYRTPTFLYESLWNLLG 180

Query: 181 FVIIMMWRRKPKSLLDGEIFAFYLIWYGSGRLVIEGMRTDSLMFLGIRISQYVSALLIII 240
           F+IIM+ RR+P  L +GE+  FYLIWYGSGR VIEGMRTDSLMF  +R+SQ++S LL+++
Sbjct: 181 FIIIMILRRRPNLLKEGEVAFFYLIWYGSGRFVIEGMRTDSLMFASLRVSQWLSVLLVVV 240

Query: 241 GLIFVIKRRRQKGISYYQ 258
           G+I ++ RRR   I YYQ
Sbjct: 241 GVILMVIRRRNHAIPYYQ 258
```

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

```
 Identities = 176/257 (68%), Positives = 221/257 (85%)

Query: 2   MINPVAIRLGPFSIRWYAICIVSGMLLAVYLAMKEAPRKNIKSDDILDFILMAFPLSIVG 61
           MINP+A++ GP +I WYA+CI+SG++LAVYLA KEAP+K I SD I DFIL+AFPL+IVG
Sbjct:. 1   MINPIALKCGPLAIHWYALCILSGLVLAVYLASKEAPKKGISSDAIFDFILIAFPLAIVG 60
```

```
Query: 62  ARIYYVIFEWAYYSKHPVEIIAIWNGGIAIYGGLITGAILLVIFSYRRLINPIDFLDIAA 121
           ARIYYVIFEW+YY KH  EIIAIWNGGIAIYGGLITGA++L+ + Y +++NPI FLDIAA
Sbjct: 61  ARIYYVIFEWSYYVKHLDEIIAIWNGGIAIYGGLITGALVLLAYCYNKVLNPIHFLDIAA 120

Query: 122 PGVMIAQAIGRWGNFINQEAYGRAVKNLNYVPNFIKNQMYIDGAYRVPTFLYESLWNFLG 181
           P VM+AQAIGRWGNFINQEAYG+AV  LNY+P+FI+ QM+I+G+YR+PTFLYESLWN LG
Sbjct: 121 PSVMVAQAIGRWGNFINQEAYGKAVSQLNYLPSFIQKQMFIEGSYRIPTFLYESLWNLLG 180

Query: 182 FVIIMSIRHRPRTLKQGEVACFYLVWYGCGRFIIEGMRTDSLYLAGLRVSQWLSVILVII 241
           FVIIM  R +P++L  GE+  FYL+WYG GR +IEGMRTDSL   G+R+SQ++S +L+II
Sbjct: 181 FVIIMMWRRKPKSLLDGEIFAFYLIWYGSGRLVIEGMRTDSLMFLGIRISQYVSALLIII 240

Query: 242 GIVMIIYRRRREQHISYY 258
           G++ +I RRR++ ISYY
Sbjct: 241 GLIFVIKRRRQKGISYY 257
```

A related GBS gene <SEQ ID 8557> and protein <SEQ ID 8558> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
McG: Discrim Score:       2.45
GvH: Signal Score (-7.5): -2.9
     Possible site: 39
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 3 value:  -8.39 threshold:  0.0
   INTEGRAL    Likelihood = -8.39   Transmembrane  209 - 225 ( 203 - 229)
   INTEGRAL    Likelihood = -7.64   Transmembrane   67 -  83 (  65 -  85)
   INTEGRAL    Likelihood = -1.86   Transmembrane   24 -  40 (  24 -  42)
   PERIPHERAL  Likelihood =  0.79      92
 modified ALOM score:   2.18

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4354(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:

```
ORF01400(238 - 1008 of 1308)
SP|P72482|LGT_STRMU(1 - 257 of 259) PROLIPOPROTEIN DIACYLGLYCERYL TRANSFERASE (EC 2.4.99.-
). GP|4583534|gb|AAC80171.3||U75480 putative prolipoprotein diacylglycerol transferase
(Streptococcus mutans) PIR|T11569|T11569 prolipoprotein diacylglyceryl transferase (EC
2.4.99.-) - Streptococcus mutans
%Match = 46.9
%Identity = 71.6  %Similarity = 89.5
Matches = 184  Mismatches = 27  Conservative Sub.s = 46

198       228       258       288       318       348       378       408
WGLMLPRLLRIV*HI*LVRTRSMMINPVAIRLGPFSIRWYAICIVSGMLLAVYLAMKEAPRKNIKSDDILDFILMAFPLS
                          ||||:||:|||::|||||:||||:|::|||||  ::|||:|||||||:|||||:||||:
                         MINPIAIKLGPLTIRWYSICIVTGLILAVYLTIREAPKKNIKSDDVLDFILIAFPLA
                                   10        20        30        40        50

438       468       498       528       558       588       618       648
IVGARIYYVIFEWAYYSKHPVEIIAIWNGGIAIYGGLITGAILLVIFSYRRLINPIDFLDIAAPGVMIAQAIGRWGNFIN
|||||:|||||:| || |:| ||  ||:||||||||||:|||::| |||| |:| ||||||:|||||||:||:||||||||:|
IVGARLYYVIFDWDYYLKNPSEIPVIWHGGIAIYGGLLTGALVLFIFSYYRMIKPIDFLDVAAPGVMLAQSIGRWGNFVN
          70        80        90        100       110       120       130

678       708       738       768       798       828       858       888
QEAYGRAVKNLNYVPNFIKNQMYIDGAYRVPTFLYESLWNFLGFVIIMSIRHRPRTLKQGEVACFYLVWYGCGRFIIEGM
|||||: |  |||:|:||: ||||||  || ||||||||||:|||:||| :| ||  ||:|||| |||:||| |||:||||
QEAYGKTVTQLNYLPDFIRKQMYIDGHYRTPTFLYESLWNLLGFIIIMILRRRPNLLKEGEVAFFYLIWYGSGRFVIEGM
          150       160       170       180       190       200       210
```

```
918       948       978       1008      1038      1068      1098      1128
RTDSLYLAGLRVSQWLSVILVIIGIVMIIYRRREQHISYY*TEEVL**KLLY*LLPLRLLF*F*EYFSF*KKYQKRLRKP
|||||  :| |||||||||:||::|:::::  ||| : | ||
RTDSLMFASLRVSQWLSVLLVVVGVILMVIRRRNHAIPYYQC
          230       240       250
```

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

Example 344

A DNA sequence (GBSx0375) was identified in *S.agalactiae* <SEQ ID 1115> which encodes the amino acid sequence <SEQ ID 1116>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2817(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:BAA77782 GB:AB027460 Hpr kinase [Streptococcus bovis]
 Identities = 264/309 (85%), Positives = 292/309 (94%)

Query: 1   MAVTVQMLVDRLKLNVIYGDEHLLSKRITTADISRPGLEMTGYFDYYAPERLQLVGMKEW 60
           M+VTV+MLVD++KL+VIYGD+ LLSK ITT+DISRPGLEMTGYFDYY+PERLQL+GMKEW
Sbjct: 1   MSVTVKMLVDKVKLDVIYGDDDLLSKEITTSDISRPGLEMTGYFDYYSPERLQLLGMKEW 60

Query: 61  SYLMAMTGHNRYQVLREMFQKETPAIVVARDLEIPEEMYEAAKDTGIAILQSKAPTSRLS 120
           SYL  MT HNR  VLREM + ETPAI+VAR+L IPEEM  AAK+ GIAILQS  PTSRLS
Sbjct: 61  SYLTKMTSHNRRHVLREMIKPETPAIIVARNLAIPEEMISAAKEKGIAILQSHVPTSRLS 120

Query: 121 GEVSWYLDSCLAERTSVHGVLMDIYGMGVLIQGDSGIGKSETGLELVKRGHRLVADDRVD 180
           GE+SWYLDSCLAERTSVHGVLMDIYGMGVLIQGDSGIGKSETGLELVKRGHRLVADDRVD
Sbjct: 121 GEMSWYLDSCLAERTSVHGVLMDIYGMGVLIQGDSGIGKSETGLELVKRGHRLVADDRVD 180

Query: 181 VYAKDEETLWGEPAEILRHLLEIRGVGIIDIMSLYGASAVKDSSQVQLAIYLENFETGKV 240
           V+AKDEETLWGEPAEILRHLLEIRGVGIID+MSLYGASAVKDSSQVQLAIYLEN+E+GKV
Sbjct: 181 VFAKDEETLWGEPAEILRHLLEIRGVGIIDVMSLYGASAVKDSSQVQLAIYLENYESGKV 240

Query: 241 FDRLGNGNEEIELSGVKVPRIRIPVKTGRNVSVVIEAAAMNHRAKQMGFDATQTFEDRLT 300
           FDRLGNGNEE+ELSGVK+PR+RIPV+TGRN+SVVIEAAAMN+RAKQMGFDAT+TFE+RLT
Sbjct: 241 FDRLGNGNEELELSGVKIPRLRIPVQTGRNMSVVIEAAAMNYRAKQMGFDATKTFEERLT 300

Query: 301 HLISQNEVN 309
            LI++NE N
Sbjct: 301 QLITKNEGN 309
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1117> which encodes the amino acid sequence <SEQ ID 1118>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2391(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 = 255/309 (82%), Positives = 288/309 (92%)
```

```
Query: 1   MAVTVQMLVDRLKLNVIYGDEHLLSKRITTADISRPGLEMTGYFDYYAPERLQLVGMKEW 60
           M VTV+MLV ++KL+V+Y  ++LLSK ITT+DISRPGLEMTGYFDYYAPERLQL GMKEW
Sbjct: 32  MTVTVKMLVQKVKLDVVYATDNLLSKEITTSDISRPGLEMTGYFDYYAPERLQLFGMKEW 91

Query: 61  SYLMAMTGHNRYQVLREMFQKETPAIVVARDLEIPEEMYEAAKDTGIAILQSKAPTSRLS 120
           SYL  MT HNRY VL+EMF+K+TPA+VV+R+L IP+EM +AAK+ GI++L S+  TSRL+
Sbjct: 92  SYLTQMTSHNRYSVLKEMFKKDTPAVVVSRNLAIPKEMVQAAKEEGISLLSSRVSTSRLA 151

Query: 121 GEVSWYLDSCLAERTSVHGVLMDIYGMGVLIQGDSGIGKSETGLELVKRGHRLVADDRVD 180
           GE+S++LD+ LAERTSVHGVLMDIYGMGVLIQGDSGIGKSETGLELVKRGHRLVADDRVD
Sbjct: 152 GEMSYFLDASLAERTSVHGVLMDIYGMGVLIQGDSGIGKSETGLELVKRGHRLVADDRVD 211

Query: 181 VYAKDEETLWGEPAEILRHLLEIRGVGIIDIMSLYGASAVKDSSQVQLAIYLENFETGKV 240
           VYAKDEETLWGEPAEILRHLLEIRGVGIID+MSLYGASAVKDSSQVQLAIYLENFE GKV
Sbjct: 212 VYAKDEETLWGEPAEILRHLLEIRGVGIIDVMSLYGASAVKDSSQVQLAIYLENFEAGKV 271

Query: 241 FDRLGNGNEEIELSGVKVPRIRIPVKTGRNVSVVIEAAAMNHRAKQMGFDATQTFEDRLT 300
           FDRLGNGNEEI  SGV++PRIRIPVKTGRNVSVVIEAAAMNHRAK+MGFDAT+TFEDRLT
Sbjct: 272 FDRLGNGNEEITFSGVRIPRIRIPVKTGRNVSVVIEAAAMNHRAKEMGFDATKTFEDRLT 331

Query: 301 HLISQNEVN 309
            LI++NEV+
Sbjct: 332 QLITKNEVS 340
```

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

Example 345

A DNA sequence (GBSx0376) was identified in *S.agalactiae* <SEQ ID 1119> which encodes the amino acid sequence <SEQ ID 1120>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1836(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 9719> which encodes amino acid sequence <SEQ ID 9720> was also identified.

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

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

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

Example 346

A DNA sequence (GBSx0377) was identified in *S.agalactiae* <SEQ ID 1121> which encodes the amino acid sequence <SEQ ID 1122>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -4.88   Transmembrane   35 -  51 (  31 -  59)

----- Final Results -----
            bacterial membrane --- Certainty=0.2954(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:AAC67275 GB:AF017113 YvlC [Bacillus subtilis]
 Identities = 21/63 (33%), Positives = 36/63 (56%), Gaps = 2/63 (3%)

Query: 3  SSFYKQRKGKLVCGVVAGLADKYNWDLALSRVLIALILYFTKF--GLLLYILLAVFLPYK 60
          +  Y+  K K + GV+ GLA+ +NWD +L RV+  ++   T     LL+YI+    +P +
Sbjct: 2  NKLYRSEKNKKIAGVIGGLAEYFNWDASLLRVITVILAIMTSVLPVLLIYIIWIFIVPSE 61

Query: 61 EDI 63
           D+
Sbjct: 62 RDM 64
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1123> which encodes the amino acid sequence <SEQ ID 1124>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.26   Transmembrane   39 -  55 (  31 -  61)

----- Final Results -----
               bacterial membrane --- Certainty=0.3102(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 = 60/90 (66%), Positives = 77/90 (84%), Gaps = 3/90 (3%)

Query: 1  MKSSFYKQRKGKLVCGVVAGLADKYNWDLALSRVLIALILYFTKFGLLLYILLAVFLPYK 60
          +++ FYKQRK +LV GV+AGLADKY WDLAL+RVL AL++Y T FG+LLYILLA+FLPYK
Sbjct: 1  VETKFYKQRKNRLVAGVIAGLADKYGWDLALARVLAALLIYGTGFGVLLYILLAIFLPYK 60

Query: 61 EDIIETR-RQGPRRRKDAEPV--DDDGWFW 87
          ED++E R  +GPRRRKDA+ +  ++DGWFW
Sbjct: 61 EDLLEERYGRGPRRRKDADVLNEEEDGWFW 90
```

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

Example 347

A DNA sequence (GBSx0378) was identified in *S.agalactiae* <SEQ ID 1125> which encodes the amino acid sequence <SEQ ID 1126>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3577(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 9717> which encodes amino acid sequence <SEQ ID 9718> was also identified.

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

```
>GP:BAB04250 GB:AP001508 unknown conserved protein [Bacillus halodurans]
 Identities = 379/729 (51%), Positives = 515/729 (69%), Gaps = 25/729 (3%)

Query: 29  ENLNITQIAIDLGIKASQIEKVLELTDEGNTIPFIARYRKEMTGNLDEVQIKSIIDLDKS 88
```

```
            E   I  +A +L +K + I++V++L  EGNT+PFIARYRKE+TG +DEV+I+ + +
Sbjct: 8    EEHTIKTLAKELSLKPNYIKQVIQLLHEGNTVPFIARYRKELTGGMDEVKIREVSEKWTY 67

Query: 89   MTALSDRKTTVLAKIEEQGKLTQELKKAIEEATKLADVEELYLPYKEKRRTKATIAREAG 148
               L +RK  V+  +EEQGKLT E KK +E+A KL +VE+LY PYK+KRRT+AT+A+E G
Sbjct: 68   ANQLHERKEEVIRLVEEQGKLTDEWKKTVEQAQKLQEVEDLYRPYKQKRRTRATVAKEKG 127

Query: 149  LFPLARLI--LQNKDNLEEEAQNYLTDGFETTT--KALSGAVDILIEAFSEDNKLRSWTY 204
            L PLA  +  L    +  +EA+ YL+   E T     L GA DI+ E  ++D  LR
Sbjct: 128  LEPLAEWLFSLPRDGDPLQEAEVYLSVEHELTKVEDVLQGAQDIIAEWIADDADLRKRIR 187

Query: 205  NEIWNYSSITAVVKDESLDEKQVFKIYYDFSEKISKLHGYQVLALNRGEKMGVLKVNFEH 264
            +  +   S+ A VK E LDEK V+++YYD+ E +  L  ++ LALNRGEK  VL+V
Sbjct: 188  SLGFKEGSVIAKVKKEELDEKGVYEMYYDYEEPVRTLVPHRTLALNRGEKEDVLRVTIRF 247

Query: 265  NLEKMFRF----FAVRFKETS-QYIDDLIVQTVKKKIVPAMERRIRTELSEGAEDGAISL 319
             ++++       F  RF   +  Y+   I    K+ I P++ER IR EL+E AE+ AI +
Sbjct: 248  PVDRIIEMSEKTFIRRFGSPAVPYVKAAIEDGYKRLIEPSIEREIRHELTEKAEEQAIHI 307

Query: 320  FSENLRNLLLVSPLKGKMVLGFDPAFRTGAKLAVVDQTGKLMTTQVIYPVPPANQAKIEQ 379
            F+ENLR+LLL  P+KGK+VLG DPA+RTG KLA+VD+TGK++  QVIYP PP N+  +
Sbjct: 308  FAENLRSLLLQPPIKGKVVLGLDPAYRTGCKLAIVDETGKVLDIQVIYPTPPKNE--VAA 365

Query: 380  SKIELAKLIKEFNIEIIAIGNGTASRESEAFVAEVLQDFPD-VSYVIVNESGASVYSASE 438
            +K  + KLI ++ +E+IAIGNGTASRESE F+A++++D P  + Y+IVNE+GASVYSASE
Sbjct: 366  AKKIVKKLIADYGVEMIAIGNGTASRESEQFIADLIKDLPQTIYYLIVNEAGASVYSASE 425

Query: 439  LARHEFPDLTVEKRSAISIARRLQDPLAELVKIDPKSIGVGQYQHDVSQKKLAENLDFVV 498
            + R EFPDL VE+RSA+SIARRLQDPLAELVKIDPKS+GVGQYQHDVSQK+L E+L FVV
Sbjct: 426  IGREEFPDLQVEERSAVSIARRLQDPLAELVKIDPKSVGVGQYQHDVSQKRLNESLTFVV 485

Query: 499  ETVVNQVGVNVNTASPALLAHVSGLNKTISENIVKYREENGQIKSRAEIKKVPRLGAKAF 558
            ETVVNQVGVNVNTASP+LL +V+GL+KT+++NIVK REE G+  +RA++K +PRLGAK +
Sbjct: 486  ETVVNQVGVNVNTASPSLLQYVAGLSKTVAKNIVKKREEAGRFTARAQLKDIPRLGAKTY 545

Query: 559  EQAAGFLRIPNAKNFLDNTGVHPESYEAVKKLLDQLTIKELD---DLAKEKLQNLDLIAT 615
            EQ  GFLRI +  N LD T +HPESY+   KLL ++     D   +  K+KLQ LD+ A
Sbjct: 546  EQCIGFLRIMDGDNLLDATAIHPESYKVTDKLLSEVGATAADVGIEDLKKKLQALDVSAM 605

Query: 616  AESIGVGQETLKDIIEDLLKPGRDLRDDFEAPVLRHDVLDVSDLKVGQELQGTVRNVVDF 675
            A ++ VG  TLKD+I+ L++P RD RD+   P+L+ DVL + DL  G ELQGTVRNVVDF
Sbjct: 606  AATLDVGVPTLKDMIDALIRPTRDPRDEVAKPLLKQDVLQLEDLLPGMELQGTVRNVVDF 665

Query: 676  GAFVDIGVHEDGLIHQSRLIKRKRDKKTRKMPPLQHPSKYLSVGDIVTVWVVEVDAERSR 735
            G FVDIGV +DGL+H S+L  R            ++HP + ++VG+IVTVWV +VD ++ R
Sbjct: 666  GVFVDIGVKQDGLVHISKLANRY----------IKHPLEVVTVGEIVTVWVEDVDIKKGR 715

Query: 736  IGLSLIKPD 744
            I L++++P+
Sbjct: 716  LALTMLRPE 724
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1127> which encodes the amino acid sequence <SEQ ID 1128>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2207(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 = 532/716 (74%), Positives = 619/716 (86%), Gaps = 10/716 (1%)

Query: 28  MENLNITQIAIDLGIKASQIEKVLELTDEGNTIPFIARYRKEMTGNLDEVQIKSIIDLDK 87
```

```
            MEN N   IA  L +   QIE+VL LT +GNTIPFIARYRKE+TGNLDEV IKSIID+DK
Sbjct: 1    MENNNNHNIAEALSVSLHQIEQVLALTAQGNTIPFIARYRKEVTGNLDEVVIKSIIDMDK 60

Query: 88   SMTALSDRKTTVLAKIEEQGKLTQELKKAIEEATKLADVEELYLPYKEKRRTKATIAREA 147
            S+T L++RK T+LAKIEEQGKLT +L+ +IE   KLAD+EELYLPYKEKRRTKATIAREA
Sbjct: 61   SLTTLNERKATILAKIEEQGKLTDQLRTSIEATEKLADLEELYLPYKEKRRTKATIAREA 120

Query: 148  GLFPLARLILQNKDNLEEEAQNYLTDGFETTTKALSGAVDILIEAFSEDNKLRSWTYNEI 207
            GLFPLARLILQN  NLE  A+ ++T+GF +  +AL+GAVDIL+EA SED KLRSWTYNEI
Sbjct: 121  GLFPLARLILQNAQNLETAAEPFVTEGFASPQEALAGAVDILVEAMSEDAKLRSWTYNEI 180

Query: 208  WNYSSITAVVKDESLDEKQVFKIYYDFSEKISKLHGYQVLALNRGEKMGVLKVNFEHNLE 267
             W YS + + +KDE LDEK+VF+IYYDFS+++S + GY+ LALNRGEK+G+LKV+FEHNLE
Sbjct: 181  WQYSRLVSTLKDEQLDEKKVFQIYYDFSDQVSNMQGYRTLALNRGEKLGILKVSFEHNLE 240

Query: 268  KMFRFFAVRFKETSQYIDDLIVQTVKKKIVPAMERRIRTELSEGAEDGAISLFSENLRNL 327
            KM RFF+VRFKET+ YI+++I QT+KKKIVPAMERR+R+ELS+ AEDGAI LFSENLR+L
Sbjct: 241  KMQRFFSVRFKETNPYIEEVINQTIKKKIVPAMERRVRSELSDAAEDGAIHLFSENLRHL 300

Query: 328  LLVSPLKGKMVLGFDPAFRTGAKLAVVDQTGKLMTTQVIYPVPPANQAKIEQSKIELAKL 387
            LLVSPLKGKMVLGFDPAFRTGAKLA+VDQTGKL+TTQVIYPV PA+Q KI+ +K  L +L
Sbjct: 301  LLVSPLKGKMVLGFDPAFRTGAKLAIVDQTGKLLTTQVIYPVAPASQTKIQAAKETLTQL 360

Query: 388  IKEFNIEIIAIGNGTASRESEAFVAEVLQDFPDVSYVIVNESGASVYSASELARHEFPDL 447
            I+ + I+IIAIGNGTASRESEAFVA+VL+DFP+ SYVIVNESGASVYSASELARHEFPDL
Sbjct: 361  IETYQIDIIAIGNGTASRESEAFVADVLKDFPNTSYVIVNESGASVYSASELARHEFPDL 420

Query: 448  TVEKRSAISIARRLQDPLAELVKIDPKSIGVGQYQHDVSQKKLAENLDFVVETVVNQVGV 507
            TVEKRSAISIARRLQDPLAELVKIDPKSIGVGQYQHDVSQKKL+ENL FVV+TVVNQVGV
Sbjct: 421  TVEKRSAISIARRLQDPLAELVKIDPKSIGVGQYQHDVSQKKLSENLGFVVDTVVNQVGV 480

Query: 508  NVNTASPALLAHVSGLNKTISENIVKYREENGQIKSRAEIKKVPRLGAKAFEQAAGFLRI 567
            NVNTASP+LLAHVSGLNKTISENIVKYREENG + SRA+IKKVPRLGAKAFEQAAGFLRI
Sbjct: 481  NVNTASPSLLAHVSGLNKTISENIVKYREENGALTSRADIKKVPRLGAKAFEQAAGFLRI 540

Query: 568  PNAKNFLDNTGVHPESYEAVKKLLDQLTIKELDDLAKEKLQNLDLIATAESIGVGQETLK 627
            P AKN LDNTGVHPESY AVK+L   L I++LDD AK  L  + +   AE++ +GQETLK
Sbjct: 541  PGAKNILDNTGVHPESYPAVKELFKVLGIQDLDDAAKATLAAVQVPQMAETLAIGQETLK 600

Query: 628  DIIEDLLKPGRDLRDDFEAPVLRHDVLDVSDLKVGQELQGTVRNVVDFGAFVDIGVHEDG 687
            DII DLLKPGRDLRDDFEAP+LR D+LD+ DL++GQ+L+GTVRNVVDFGAFVDIGVHEDG
Sbjct: 601  DIIADLLKPGRDLRDDFEAPILRQDILDLKDLEIGQKLEGTVRNVVDFGAFVDIGVHEDG 660

Query: 688  LIHQSRLIKRKRDKKTRKMPPLQHPSKYLSVGDIVTVWVVEVDAERSRIGLSLIKP 743
            LIH S + K             + HPS+ +SVGD+VTVWV ++D +R ++ LSL+ P
Sbjct: 661  LIHISEMSKTF----------VNHPSQVVSVGDLVTVWVSKIDLDRHKVNLSLLPP 706
```

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

Example 348

A DNA sequence (GBSx0379) was identified in *S.agalactiae* <SEQ ID 1129> which encodes the amino acid sequence <SEQ ID 1130>. This protein is predicted to be N5,N10-methylenetetrahydromethanopterin reductase homolog. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4864(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:AAB94650 GB:U96107 N5,N10-methylenetetrahydromethanopterin
          reductase homolog [Staphylococcus carnosus]
 Identities = 164/300 (54%), Positives = 217/300 (71%), Gaps = 1/300 (0%)

Query: 45  VYGIGEHHREDFAVSAPEIVLAAGAVRTNNIRLSSAVTILSSNDPIRVYQQFSTIDALSN 104
           +YG+GEHHR D+AVS P  VLAA A  T  I+LSSAVT+LSS+DP+ VY++F+T+DA+SN
Sbjct: 1   MYGLGEHHRSDYAVSDPVTVLAAAASLTQRIKLSSAVTVLSSDDPVCVYERFATLDAVSN 60

Query: 105 GRAEIMAGRGSFIESFPLFGYDLADYDDLFNEKMDMLLAINSATNLDWKGHLTQTVNERP 164
           GRAEIM GRGSFIESFPLFGYDL DYD LF EK+++L  IN    + W+G +   +
Sbjct: 61  GRAEIMVGRGSFIESFPLFGYDLDDYDRLFVEKLELLKEINQHEVVTWEGTMRPAIKGLG 120

Query: 165 IYPRALQRQLPIWVATGGNVDSTIRIAEQGLPIVYATIGGNPKAFRQLVHIYKEVGSRNG 224
           +YPRA+Q ++PIW+ATGG  +S+IR AE GLPI YA IGGNPK F++ + IY+ V    G
Sbjct: 121 VYPRAVQDEIPIWLATGGTPESSIRAAEFGLPITYAIIGGNPKRFKRNIAIYRAVAESRG 180

Query: 225 HKPEQLKVAAHSWGWIEEDNQTAIDRYFFPTKQTVDNIAKGRPHWSEMTKEQYLRSVGPE 284
           +    + VA HSWG+I + ++ A   ++ PTK   + IAK R +W   T+  + R +  E
Sbjct: 181 YDLADMPVAVHSWGYIADTDEQAQREFYEPTKVHHEIIAKER-NWPPYTEAHFQREISDE 239

Query: 285 GAIFVGSPEVVAHKIIGLVEALELDRFMLHLPVGSMPHKDVLNAIKLYGKEVAPIVRKYF 344
           GA+FVGSPE VA K+I ++E L L+RFMLH+PVGSMPH+ ++ AIKLYGK V PI+  YF
Sbjct: 240 GAMFVGSPETVARKMIKVIEELGLNRFMLHIPVGSMPHERIMKAIKLYGKRVKPIIEDYF 299
```

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 349

A DNA sequence (GBSx0380) was identified in *S.agalactiae* <SEQ ID 1131> which encodes the amino acid sequence <SEQ ID 1132>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1310(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 9715> which encodes amino acid sequence <SEQ ID 9716> 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 1133> which encodes the amino acid sequence <SEQ ID 1134>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0915(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 = 20/40 (50%), Positives = 27/40 (67%), Gaps = 3/40 (7%)

Query: 4  MAITHKRQDDLESMFASFAKVP---KPKKVDSDSKPEQKD 40
          MAITHK+ D+LE M A FA +P    KP +V++D K   K+
```

```
Sbjct: 1  MAITHKKNDELEKMLAGFASIPSFDKPLEVNTDGKLATKE 40
```

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

Example 350

A DNA sequence (GBSx0381) was identified in *S.agalactiae* <SEQ ID 1135> which encodes the amino acid sequence <SEQ ID 1136>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1453(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 351

A DNA sequence (GBSx0382) was identified in *S.agalactiae* <SEQ ID 1137> which encodes the amino acid sequence <SEQ ID 1138>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.15   Transmembrane  216 - 232 ( 210 - 240)
   INTEGRAL    Likelihood = -9.18   Transmembrane   15 -  31 (  10 -  39)
   INTEGRAL    Likelihood = -9.02   Transmembrane  283 - 299 ( 276 - 299)
   INTEGRAL    Likelihood = -8.76   Transmembrane  128 - 144 ( 119 - 150)
   INTEGRAL    Likelihood = -4.62   Transmembrane  243 - 259 ( 237 - 265)
   INTEGRAL    Likelihood = -2.44   Transmembrane   65 -  81 (  65 -  81)
   INTEGRAL    Likelihood = -2.44   Transmembrane   94 - 110 (  93 - 111)

----- 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 GENPEPT database:

```
>GP:CAB12119 GB:Z99105 ycgR [Bacillus subtilis]
 Identities = 141/283 (49%), Positives = 198/283 (69%), Gaps = 3/283 (1%)

Query: 10  SVLQWFAIFISIIIEALPFVLLGTILSGIIEVFITPDIVNKFLPKNKFLRVLFGTFVGFV 69
           S LQ  +IFISI+IEA+PF+L+G ILSGII++F++ +++ + +PKN+FL VLFG   G +
Sbjct: 6   SFLQLNSIFISILIEAIPFILIGVILSGIIQMFVSEEMIARIMPKNRFLAVLFGALAGVL 65

Query: 70  FPSCECGIIPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSIRFLILRFVG 129
           FP+CECGIIPI  R L K VP +  V F+ TAPIINPIVLF+TY AFGN    +  R
Sbjct: 66  FPACECGIIPITRRLLLKGVPLHAGVAFMLTAPIINPIVLFSTYIAFGNRWSVVFYRGGL 125

Query: 130 ATIVAIALGVMLAPLVDDNILKEDAKPTHFHDYSDKKWYQKIFLALAHAIDEFFDTGRYL 189
           A  V++ +GV+L++   DN L +  +P H H +      QK+   L HAIDEFF  G+YL
Sbjct: 126 ALAVSLIIGVILSYQFKDNQLLKPDEPGHHHHHHGTL-LQKLGGTLRHAIDEFFSVGKYL 184

Query: 190 VFGTLIASAMQIYLPTRVLTTIGHSPITAILVMMLLAFILSLCSEADAFIGASLLSTFGI 249
```

```
          + G  IA+AMQ Y+ T  L  IG + +++ LVMM LAF+LSLCSE DAFI +S  STF +
Sbjct: 185 IIGAFIAAAMQTYVKTSTLLAIGQNDVSSSLVMMGLAFVLSLCSEVDAFIASSFSSTFSL 244

Query: 250 APVMAFLLIGPMIDIKNLMMMVNSFKTRFIVQFISVSSLIIII 292
             ++AFL+ G M+DIKNL+MM+ +FK RF+  F+ ++ +++I+
Sbjct: 245 GSLIAFLVFGAMVDIKNLLMMLAAFKKRFV--FLLITYIVVIV 285
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1139> which encodes the amino acid sequence <SEQ ID 1140>. Analysis of this protein sequence reveals the following:

```
     Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.92   Transmembrane  216 - 232 ( 211 - 237)
   INTEGRAL    Likelihood = -9.45   Transmembrane  283 - 299 ( 276 - 299)
   INTEGRAL    Likelihood = -8.76   Transmembrane  128 - 144 ( 119 - 150)
   INTEGRAL    Likelihood = -7.80   Transmembrane   15 -  31 (  10 -  39)
   INTEGRAL    Likelihood = -5.47   Transmembrane  243 - 259 ( 237 - 265)
   INTEGRAL    Likelihood = -2.44   Transmembrane   65 -  81 (  65 -  81)
   INTEGRAL    Likelihood = -2.44   Transmembrane   94 - 110 (  93 - 111)

----- 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 databases:

```
>GP:CAB12119 GB:Z99105 ycgR [Bacillus subtilis]
 Identities = 143/288 (49%), Positives = 196/288 (67%), Gaps = 1/288 (0%)

Query: 10  SVLQWFAIFMSIIIEALPFVLLGTILSGCIEVFVTPELVQKYLPKQKCLRILFGTFVGFV 69
           S LQ  +IF+SI+IEA+PF+L+G ILSG I++FV+ E++ + +PK + L +LFG   G +
Sbjct: 6   SFLQLNSIFISILIEAIPFILIGVILSGIIQMFVSEEMIARIMPKNRFLAVLFGALAGVL 65

Query: 70  FPSCECGIIPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSLRFLILRLVG 129
           FP+CECGIIPI  R L K VP +  V F+ TAPIINPIVLF+TY AFGN    +  R
Sbjct: 66  FPACECGIIPITRRLLLKGVPLHAGVAFMLTAPIINPIVLFSTYIAFGNRWSVVFYRGGL 125

Query: 130 AALVAITLGVMLAFIVDDNILKDNAQPVHFHDYSHESLPKRIYLALVHAIDEFFDTGRYL 189
           A  V++ +GV+L++   DN L    +P H H + H +L +++   L HAIDEFF  G+YL
Sbjct: 126 ALAVSLIIGVILSYQFKDNQLLKPDEPGH-HHHHHGTLLQKLGGTLRHAIDEFFSVGKYL 184

Query: 190 VFGTLIASAMQIYVPTRVLTTIGHNPLTAILIMMLMAFILSLCSEADAFIGASLLSTFGV 249
           + G  IA+AMQ YV T  L  IG N +++ L+MM +AF+LSLCSE DAFI +S  STF +
Sbjct: 185 IIGAFIAAAMQTYVKTSTLLAIGQNDVSSSLVMMGLAFVLSLCSEVDAFIASSFSSTFSL 244

Query: 250 APVLAFLLIGPMVDIKNLMMMVKAFKGRFIVQFIGVSVLMIAVYCLLV 297
             ++AFL+ G MVDIKNL+MM+ AFK RF+  I   V+++    LLV
Sbjct: 245 GSLIAFLVFGAMVDIKNLLMMLAAFKKRFVFLLITYIVVIVLAGSLLV 292
```

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

```
 Identities = 248/300 (82%), Positives = 278/300 (92%)

Query: 1   MDIFNQLPDSVLQWFAIFISIIIEALPFVLLGTILSGIIEVFITPDIVNKFLPKNKFLRV 60
           M +F+ LP SVLQWFAIF+SIIIEALPFVLLGTILSG IEVF+TP++V K+LPK K LR+
Sbjct: 1   MSLFSNLPPSVLQWFAIFMSIIIEALPFVLLGTILSGCIEVFVTPELVQKYLPKQKCLRI 60

Query: 61  LFGTFVGFVFPSCECGIIPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSI 120
           LFGTFVGFVFPSCECGIIPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNS+
Sbjct: 61  LFGTFVGFVFPSCECGIIPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSL 120

Query: 121 RFLILRFVGATIVAIALGVMLAFLVDDNILKEDAKPTHFHDYSDKKWYQKIFLALAHAID 180
           RFLILR VGA +VAI LGVMLAF+VDDNILK++A+P HFHDYS +   ++I+LAL HAID
Sbjct: 121 RFLILRLVGAALVAITLGVMLAFIVDDNILKDNAQPVHFHDYSHESLPKRIYLALVHAID 180

Query: 181 EFFDTGRYLVFGTLIASAMQIYLPTRVLTTIGHSPITAILVMMLLAFILSLCSEADAFIG 240
```

```
            EFFDTGRYLVFGTLIASAMQIY+PTRVLTTIGH+P+TAIL+MML+AFILSLCSEADAFIG
Sbjct: 181 EFFDTGRYLVFGTLIASAMQIYVPTRVLTTIGHNPLTAILIMMLMAFILSLCSEADAFIG 240

Query: 241 ASLLSTFGIAPVMAFLLIGPMIDIKNLMMMVNSFKTRFIVQFISVSSLIIIIYCLFVGVI 300
           ASLLSTFG+APV+AFLLIGPM+DIKNLMMMV +FK RFIVQFI VS L+I +YCL VGV+
Sbjct: 241 ASLLSTFGVAPVLAFLLIGPMVDIKNLMMMVKAFKGRFIVQFIGVSVLMIAVYCLLVGVL 300
```

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

Example 352

A DNA sequence (GBSx0383) was identified in *S.agalactiae* <SEQ ID 1141> which encodes the amino acid sequence <SEQ ID 1142>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4703(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 353

A DNA sequence (GBSx0384) was identified in *S.agalactiae* <SEQ ID 1143> which encodes the amino acid sequence <SEQ ID 1144>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.44   Transmembrane   45 -  61 (  39 -  65)
   INTEGRAL    Likelihood = -8.12   Transmembrane   83 -  99 (  77 - 101)
   INTEGRAL    Likelihood = -0.00   Transmembrane    2 -  18 (   1 -  19)

----- Final Results -----
               bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8559> which encodes amino acid sequence <SEQ ID 8560> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 2
SRCFLG: 0
McG: Length of UR:   8
     Peak Value of UR:   2.23
     Net Charge of CR: 1
McG: Discrim Score:      0.46
GvH: Signal Score (-7.5): -3.54
     Possible site: 42
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 2 value:  -8.44 threshold:  0.0
   INTEGRAL    Likelihood = -8.44   Transmembrane   37 -  53 (  31 -  57)
   INTEGRAL    Likelihood = -8.12   Transmembrane   75 -  91 (  69 -  93)
   PERIPHERAL  Likelihood =  2.76     200
```

```
 modified ALOM score:   2.19
icml HYPID: 7  CFP: 0.438

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB12118 GB:Z99105 ycgQ [Bacillus subtilis]
 Identities = 100/290 (34%), Positives = 159/290 (54%), Gaps = 25/290 (8%)

Query: 9   MIRFLILAGYFELSMYLKLSGKLNQYINTHYTYLAYISMVLSFILAIVQLIIWVKNMKMH 68
           M R L+L G+      +L  SG L +YIN  Y YL++I++ L  IL  VQ  +++K+ +
Sbjct: 1   MFRLLVLMGFTFFFYHLHASGNLTKYINMKYAYLSFIAIFLLAILTAVQAYLFIKSPEKS 60

Query: 69  SHLHGKIA----------KSTSP--------MILVFPVLVGLLVPTVSLDSTTVSAKGYN 110
            H H              +   P         ++ +FP++ G+  P  +LDS+ V  KG++
Sbjct: 61  GHHHDHDCGCGHDHEHDHEQNKPFYQRYLIYVVFLFPLVSGIFFPIATLDSSIVKTKGFS 120

Query: 111 FPLAAGSTGTVSQDGTRVQYLKPDTSTYFTSSAYEKEMQKELKKYKGSGTLTITTENYME 170
           F  A  S    SQ      QYL+PD S Y+   +Y+K+M++   KY      +++T +++++
Sbjct: 121 FK-AMESGDHYSQ----TQYLRPDASLYYAQDSYDKQMKQLFNKYSSKKEISLTDDDFLK 175

Query: 171 VMELIYLYPEQFMDRQIQYTGFVY-NEPKHEGYQFIFRFGIIHCIADSGVYGLLTT-GNQ 228
            ME IY YP +F+ R I++ GF Y     ++    F+ RFGIIHCIADSGVYG+L
Sbjct: 176 GMETIYNYPGEFLGRTIEFHGFAYKGNAINKNQLFVLRFGIIHCIADSGVYGMLVEFPKD 235

Query: 229 KSYPDNTWVTVRGTIKSEYNQLLQQNLPVLHIEESRQVSKANNPYVYRVF 278
               D+ W+ ++GT+ SEY Q  +  LPV+ + +    + K ++PYVYR F
Sbjct: 236 MDIKDDEWIHIKGTLASEYYQPFKSTLPVVKVTDWNTIKKPDDPYVYRGF 285
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1145> which encodes the amino acid sequence <SEQ ID 1146>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.33   Transmembrane   83 -  99 (  74 - 101)
   INTEGRAL    Likelihood = -6.21   Transmembrane   42 -  58 (  39 -  62)

----- 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>
```

A related sequence was also identified in GAS <SEQ ID 9115> which encodes the amino acid sequence <SEQ ID 9116>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 54
>>> Seems to have an uncleavable N-term signal seq
     INTEGRAL    Likelihood = -8.33   Transmembrane   75 -  91 (  66 -  93)
     INTEGRAL    Likelihood = -6.21   Transmembrane   34 -  50 (  31 -  54)
     PERIPHERAL  Likelihood =  2.76

----- Final Results -----
              bacterial membrane --- Certainty= 0.433(Affirmative) < succ>
               bacterial outside --- Certainty= 0.000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

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

```
Identities = 208/279 (74%), Positives = 244/279 (86%), Gaps = 1/279 (0%)
```

```
Query: 1   MFICGGNIMIRFLILAGYFELSMYLKLSGKLNQYINTHYTYLAYISMVLSFILAIVQLII 60
           +F CGG +MIRFLILAGYFEL+MYL+LSGKL+QYIN  Y+YLAYISM+LSFILA+VQL
Sbjct: 1   LFTCGGALMIRFLILAGYFELTMYLQLSGKLDQYINVRYSYLAYISMILSFILALVQLYT 60

Query: 61  WVKNMKMHSHLHGKIAKSTSPMILVFPVLVGLLVPTVSLDSTTVSAKGYNFPLAAGSTGT 120
           W+KN+K+HSHL GKIA+ TSP ILVFPVL+GLLVPTV+LDSTTVSAKGY FPLAAG++ T
Sbjct: 61  WMKNIKVHSHLTGKIARLTSPFILVFPVLIGLLVPTVTLDSTTVSAKGYTFPLAAGASKT 120

Query: 121 -VSQDGTRVQYLKPDTSTYFTSSAYEKEMQKELKKYKGSGTLTITTENYMEVMELIYLYP 179
            VS DGT +QYLKPDTS YFT SAY+KEM++EL KYKG   +TITTENYMEVMELIYLYP
Sbjct: 121 GVSDDGTTIQYLKPDTSLYFTKSAYQKEMRQELHKYKGKKPVTITTENYMEVMELIYLYP 180

Query: 180 EQFMDRQIQYTGFVYNEPKHEGYQFIFRFGIIHCIADSGVYGLLTTGNQKSYPDNTWVTV 239
           ++F+DR IQYTGFVYNEP H+ YQF+FRFGIIHCIADSGVYGLLTTGNQ SYP+NTW+TV
Sbjct: 181 DEFLDRDIQYTGFVYNEPGHDNYQFLFRFGIIHCIADSGVYGLLTTGNQTSYPNNTWLTV 240

Query: 240 RGTIKSEYNQLLQQNLPVLHIEESRQVSKANNPYVYRVF 278
           +G +  EY++ L+Q+LPVL + E  Q  + NNPYVYRVF
Sbjct: 241 KGRLHMEYDKNLEQHLPVLQLAEVHQTKEPNNPYVYRVF 279
```

SEQ ID 8560 (GBS235d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 14 & 15; MW 48.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 17 & 18; MW 23.4kDa), in Figure 150 (lane 15; MW 23kDa) and in Figure 182 (lane 5; MW 23kDa).

GBS235d-His was purified as shown in Figure 235, lane 6-7.

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

Example 354

A DNA sequence (GBSx0385) was identified in *S.agalactiae* <SEQ ID 1147> which encodes the amino acid sequence <SEQ ID 1148>. This protein is predicted to be signal recognition particle (ftsY). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3301(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:BAB06205 GB:AP001515 signal recognition particle (docking
           protein) [Bacillus halodurans]
 Identities = 175/304 (57%), Positives = 227/304 (74%)

Query: 233 EKYNRSLKKTRTGFSARLNAFLSNFRRVDEEFFEELEEMLILSDVGVNVATQLTEDLRYE 292
           EK+   L+KTR  F+ ++N  +  +R VDE+FFEELEE+LI +DVGV     L E+L+ E
Sbjct: 20  EKFKAGLEKTRDSFAGKMNDLVYKYRSVDEDFFEELEEILIGADVGVTTVMDLVEELKDE 79

Query: 293 AKLENAKKSEDLKRVIVEKLVEIYEKDGIYNEAINFQEGLTVMLFVGVNGVGKTTSIGKL 352
            + +N K S+D++ +I EKL E+ EK+G   E      GL+V+L VGVNGVGKTTSIGKL
Sbjct: 80  VRRQNIKDSKDIQPIISEKLAELLEKEGGETEVNLQPAGLSVILVVGVNGVGKTTSIGKL 139

Query: 353 AHQYKSQGKKVMLVAADTFRAGAVAQLVEWGRRVDVPVVTGEEKADPASVVFDGMEKAVA 412
           AH YK QGKKV+L A DTFRAGA+ QL  WG R  V V+   E +DPA+V+FD ++ A +
Sbjct: 140 AHMYKQQGKKVILAAGDTFRAGAIEQLEVWGERAGVDVIKQSEGSDPAAVMFDAIQAAKS 199

Query: 413 QGVDVLLIDTAGRLQNKENLMAELEKIGRIIKRVVPDAPHETLLALDASTGQNALSQAKE 472
```

```
           +  D+L+ DTAGRLQNK NLM ELEK+ R+I R +P APHE L+ALDA+TGQNA+SQAK
Sbjct: 200 READILICDTAGRLQNKVNLMKELEKVKRVISREIPGAPHEVLIALDATTGQNAMSQAKT 259

Query: 473 FSKITPLTGLILTKIDGTAKGGVVLAIRQELDIPVKFIGFGEKIDDIGEFNSEDFMRGLL 532
           F + T +TG+ILTK+DGTAKGG+VLAIR ELDIPVKF+G GEKIDD+  F+SE F+ GL
Sbjct: 260 FKETTDVTGIILTKLDGTAKGGIVLAIRHELDIPVKFVGLGEKIDDLQPFDSEQFVYGLF 319

Query: 533 EGIL 536
           + ++
Sbjct: 320 KDMV 323
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1149> which encodes the amino acid sequence <SEQ ID 1150>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4384(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 = 339/549 (61%), Positives = 404/549 (72%), Gaps = 46/549 (8%)

Query: 1   MGLFDRLFGHKKKDKEPEIEASESVVLEDEDSVIDKEEGSNFSKESTLNRTSEVPVAEDD 60
           MGLFDRLFG K+  K  E +  E+++ E       KEE S  + E        ++   + +
Sbjct: 1   MGLFDRLFGKKETPKVAEEKLEENLLTE----TTQKEELSEKANEQ-----DKIEAVQQE 51

Query: 61  SFLELERDTALSESHQPVTSEIHPLESEDTDEIPVKEDDSFLELEDRAKTKVADTSEVEN 120
              ++  + A S  + P  + ++ L  E+T                D +     DT+E
Sbjct: 52  ---DVSSEGAGSVENGPEAASVNALVEEETG--------------DNSNHPSEDTNEF-- 92

Query: 121 VVPDSTTLSDNVSAKSEASFSDKEQLSDSQASDQFSETPLQEEMS--SGKTEVQTESEDT 178
              D T L   VS  S+++ S+ + L D    +QF     Q + S  S   E    S++
Sbjct: 93  -AADKTDLK--VSELSQSTASEPKDLVDQPVVEQFPTKQAQADASNDSANEEAVDTSKEQ 149

Query: 179 SAADAFLADYYAKRKAIEKEISSNSLST---------DESEFSEAQEVLSQSQA--DTIK 227
           S++   + DYY ++ A+EK +   + +T          E++ S + E  SQ++A  DTI
Sbjct: 150 SSSQQVMEDYYRRKAALEKSLQEKAAATVPVMPEEVPQENQASTSAEA-SQNKATHDTIP 208

Query: 228 AESQEEKYNRSLKKTRTGFSARLNAFLSNFRRVDEEFFEELEEMLILSDVGVNVATQLTE 287
            E+ +EKY RSLKKTRTGFSARLN+F +NFRRVDEEFFE+LEEMLILSDVGV+VAT LTE
Sbjct: 209 -ETDQEKYKRSLKKTRTGFSARLNSFFANFRRVDEEFFEDLEEMLILSDVGVHVATTLTE 267

Query: 288 DLRYEAKLENAKKSEDLKRVIVEKLVEIYEKDGIYNEAINFQEGLTVMLFVGVNGVGKTT 347
           +LRYEAKLENAKK + LKRVIVEKLV+IYEKDG YNEAIN+Q+GLTVMLFVGVNGVGKTT
Sbjct: 268 ELRYEAKLENAKKPDALKRVIVEKLVDIYEKDGRYNEAINYQDGLTVMLFVGVNGVGKTT 327

Query: 348 SIGKLAHQYKSQGKKVMLVAADTFRAGAVAQLVEWGRRVDVPVVTGEEKADPASVVFDGM 407
           SIGKLA++YK +GKKVMLVAADTFRAGAVAQLVEWGRRVDVPV+TG EKADPASVVFDGM
Sbjct: 328 SIGKLAYRYKQEGKKVMLVAADTFRAGAVAQLVEWGRRVDVPVITGPEKADPASVVFDGM 387

Query: 408 EKAVAQGVDVLLIDTAGRLQNKENLMAELEKIGRIIKRVVPDAPHETLLALDASTGQNAL 467
           EKAVA+GVD+LLIDTAGRLQNKENLMAELEK+GRIIKRV+PDAPHETLLALDASTGQNAL
Sbjct: 388 EKAVAKGVDILLIDTAGRLQNKENLMAELEKMGRIIKRVLPDAPHETLLALDASTGQNAL 447

Query: 468 SQAKEFSKITPLTGLILTKIDGTAKGGVVLAIRQELDIPVKFIGFGEKIDDIGEFNSEDF 527
           SQAKEFSKITPLTGLILTKIDGTAKGGVVLAIRQELDIPVKFIGFGEK+DDIGEF+SEDF
Sbjct: 448 SQAKEFSKITPLTGLILTKIDGTAKGGVVLAIRQELDIPVKFIGFGEKVDDIGEFHSEDF 507

Query: 528 MRGLLEGIL 536
           M+GLLEGIL
Sbjct: 508 MKGLLEGIL 516
```

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

Example 355

A DNA sequence (GBSx0386) was identified in *S.agalactiae* <SEQ ID 1151> which encodes the amino acid sequence <SEQ ID 1152>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3592(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:AAA62048 GB:L10328 f270 [Escherichia coli]
 Identities = 101/273 (36%), Positives = 160/273 (57%), Gaps = 10/273 (3%)

Query: 4   IKILALDLDGTLFTTDKKVSEENKVALKAAREKGIKVVITTGRPLKAIGNLLEDLELVSD 63
           IK++A+D+DGTL   D  +S   K A+ AAR +G+ VV+TTGRP   + N L++L +
Sbjct: 3   IKLIAIDMDGTLLLPDHTISPAVKNAIAAARARGVNVVLTTGRPYAGVHNYLKELHMEQP 62

Query: 64  EDYSITFNGGLVQQNT-GKILAKTAMTRQEVEDIHEELYQVGLPTDILSEGTVYS----I 118
            DY IT+NG LVQ+   G  +A+TA++  +   + +   +VG     L   T+Y+    I
Sbjct: 63  GDYCITYNGALVQKAADGSTVAQTALSYDDYRXLEKLSREVGSHFHALDRTTLYTANRDI 122

Query: 119 ANKGHHSQYHLANPLLEFIEVDDLEQVPKDVVYNKIVSVIDATYLDQQIAKLPDRLKVDY 178
           +    H  +    PL+ F E    E++  +  + K++ + +   LDQ IA++P  +K  Y
Sbjct: 123 SYYTVHESFVATIPLV-FCEA---EKMDPNTQFLKVMMIDEPAILDQAIARIPQXVKEKY 178

Query: 179 EMFKSRDIILELMPKGVHKAVGLELLTKHLGLDSSQVMAMGDEANDLSMLEWAGLGVAMA 238
            + KS    LE++ K V+K  G++ L   LG+   ++MA+GD+ ND++M+E+AG+GVAM
Sbjct: 179 TVLKSAPYFLEILDKRVNKGTGVKSLADVLGIKPEEIMAIGDQENDIAMIEYAGVGVAMD 238

Query: 239 NGIPEAKAIAKATTICNNDESGVAEAIGKYILS 271
           N IP  K +A   T  +N E GVA AI KY+L+
Sbjct: 239 NAIPSVKEVANFVT-KSNLEDGVAFAIEKYVLN 270
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1153> which encodes the amino acid sequence <SEQ ID 1154>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3502(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 = 180/273 (65%), Positives = 218/273 (78%), Gaps = 1/273 (0%)

Query: 3   DIKILALDLDGTLFTTDKKVSEENKVALKAAREKGIKVVITTGRPLKAIGNLLEDLELVS 62
           +I+ILALDLDGTL+ T+K V++ NK AL AAREKG+KVVITTGRPLKAIGNLLE+L+L+
Sbjct: 2   NIRILALDLDGTLYNTEKIVTDANKKALAAAREKGVKVVITTGRPLKAIGNLLEELDLLD 61

Query: 63  DEDYSITFNGGLVQQNTGKILAKTAMTRQEVEDIHEELYQVGLPTDILSEGTVYSIANK- 121
            +DYSITFNGGLVQ+NTG++L K++++  +V  I + L  VGLPTDI+S G VYSI +K
Sbjct: 62  HDDYSITFNGGLVQRNTGEVLDKSSLSFDQVCQIQQALEAVGLPTDIISGGDVYSIPSKD 121

Query: 122 GHHSQYHLANPLLEFIEVDDLEQVPKDVVYNKIVSVIDATYLDQQIAKLPDRLKVDYEMF 181
```

```
            G HSQYHLANPLL FIEV  + ++PKD+ YNKIV+V D  +LDQQI KL   L  D+E F
Sbjct: 122 GRHSQYHLANPLLTFIEVTSVAELPKDITYNKIVTVTDPDFLDQQIIKLSPSLFEDFEAF 181

Query: 182 KSRDIILELMPKGVHKAVGLELLTKHLGLDSSQVMAMGDEANDLSMLEWAGLGVAMANGI 241
           KSRDII E+MPKG+ KA GL LL +HLGLD+  VMAMGDEAND +MLEWAGLGVAMANG+
Sbjct: 182 KSRDIIFEIMPKGIDKAFGLNLLCQHLGLDARHVMAMGDEANDFAMLEWAGLGVAMANGV 241

Query: 242 PEAKAIAKATTICNNDESGVAEAIGKYILSEEN 274
             AKA A A T   NDESGVAEA+  +IL EE+
Sbjct: 242 SGAKADADAVTTLTNDESGVAEAVKTFILEEES 274
```

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

Example 356

A DNA sequence (GBSx0387) was identified in *S.agalactiae* <SEQ ID 1155> which encodes the amino acid sequence <SEQ ID 1156>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4648(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:BAA35556 GB:D90723 Hypothetical 30.2 kd protein in idh-deoR
          intergenic region. [Escherichia coli]
 Identities = 91/264 (34%), Positives = 146/264 (54%), Gaps = 4/264 (1%)

Query: 2   IKLVATDMDGTFLDENGTYDKKRLANVLKKFKEQGIVFTAASGRSLLSLEQLFADFRDQM 61
           IKL+A DMDGTFL +  TY+++R     ++ K QGI F  ASG     L   F +  +++
Sbjct: 4   IKLIAVDMDGTFLSDQKTYNRERFMAQYQQMKAQGIRFVVASGNQYYQLISFFPEIANEI 63

Query: 62  AFIAENGSAAVLFNRLAYEQHLSREQYLDIIDHLSKSPYMENNEYVLSGKDGAYILSDAN 121
           AF+AENG   V   +  LS++ +  +++HL   P +   E +  GK+ AY L   +
Sbjct: 64  AFVAENGGWVVSEGKDVFNGELSKDAFATVVEHLLTRPEV---EIIACGKNSAYTLKKYD 120

Query: 122 PDYIEFITHYYDNLQKVSHFEDVDDIIFKVTANFTEETVRQAEEWVNQAI-PYATAVTTG 180
                 YY  L+ V +F++++DI FK   N ++E + Q ++ +++AI     +V TG
Sbjct: 121 DAMKTVAEMYYHRLEYVDNFDNLEDIFFKFGLNLSDELIPQVQKALHEAIGDIMVSVHTG 180

Query: 181 FKSIDIILSSVNKRNGLEHLCEQYGIRAEEVLSFGDNINDLEMLEWSGKAIATENARPEV 240
             SID+I+  V+K NGL  L + +GI   EV+ FGD  ND+EML  +G + A ENA   V
Sbjct: 181 NGSIDLIIPGVHKANGLRQLQKLWGIDDSEVVVFGDGGNDIEMLRQAGFSFAMENAGSAV 240

Query: 241 KEIADCIIGHHNNQAVMAYLESMV 264
             A    G +N + V+  ++ ++
Sbjct: 241 VAAAKYRAGSNNREGVLDVIDKVL 264
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1157> which encodes the amino acid sequence <SEQ ID 1158>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3401(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 = 138/265 (52%), Positives = 193/265 (72%), Gaps = 1/265 (0%)

Query: 1   MIKLVATDMDGTFLDENGTYDKKRLANVLKKFKEQGIVFTAASGRSLLSLEQLFADFRDQ 60
           MIKL+ATDMDGTFL E+GTY++++LA +L K  E+GI+F  +SGRSLL+++QLF  F DQ
Sbjct: 1   MIKLIATDMDGTFLAEDGTYNQEQLAALLPKLAEKGILFAVSSGRSLLAIDQLFEPFLDQ 60

Query: 61  MAFIAENGSAAVLFNRLAYEQHLSREQYLDIIDHLSKSPYMENNEYVLSGKDGAYILSDA 120
           +A IAENGS       + +   +++EQY ++   +  +P+      V SG+  AYIL  A
Sbjct: 61  IAVIAENGSVVQYRGEILFADMMTKEQYTEVAKKILANPHYVETGMVFSGQKAAYILKGA 120

Query: 121 NPDYIEFITHYYDNLQKVSHFEDVD-DIIFKVTANFTEETVRQAEEWVNQAIPYATAVTT 179
           + +YI+   HYY N++ ++ FED++ D IFKV+ NFT  TV +  +W+NQA+PYATAVTT
Sbjct: 121 SEEYIQKTKHYYANVKVINGFEDMENDAIFKVSTNFTGHTVLEGSDWLNQALPYATAVTT 180

Query: 180 GFKSIDIILSSVNKRNGLEHLCEQYGIRAEEVLSFGDNINDLEMLEWSGKAIATENARPE 239
           GF SIDIIL  VNK  G+EHLC+  GI+  E ++FGDN ND +MLE++G+AIATENARPE
Sbjct: 181 GFDSIDIILKEVNKGFGMEHLCQALGIKKAETIAFGDNFNDYQMLEFAGRAIATENARPE 240

Query: 240 VKEIADCIIGHHNNQAVMAYLESMV 264
           +K I+D +IGH N+ AV+ YL+ +V
Sbjct: 241 IKVISDQVIGHCNDGAVLTYLKGLV 265
```

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

Example 357

A DNA sequence (GBSx0388) was identified in *S.agalactiae* <SEQ ID 1159> which encodes the amino acid sequence <SEQ ID 1160>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2428(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 358

A DNA sequence (GBSx0389) was identified in *S.agalactiae* <SEQ ID 1161> which encodes the amino acid sequence <SEQ ID 1162>. This protein is predicted to be p115 protein (smc). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.99   Transmembrane 1092 -1108 (1088 -1110)

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

A related GBS nucleic acid sequence <SEQ ID 9713> which encodes amino acid sequence <SEQ ID 9714> was also identified.

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

```
>GP:CAB13467 GB:Z99112 chromosome segregation SMC protein homolg
           [Bacillus subtilis]
 Identities = 458/1193 (38%), Positives = 728/1193 (60%), Gaps = 27/1193 (2%)

Query: 1    MFLKEIEMQGFKSFADKTKVEFDQGVTAVVGPNGSGKSNITESLRWALGESSAKSLRGGK 60
            MFLK +++ GFKSFA++  V+F +GVTAVVGPNGSGKSNIT+++RW LGE SA+SLRGGK
Sbjct: 1    MFLKRLDVIGFKSFAERISVDFVKGVTAVVGPNGSGKSNITDAIRWVLGEQSARSLRGGK 60

Query: 61   MPDVIFAGTENRKPLNYAQVSVTLDNSDHFIENIADEVRVERRIFRNGDSEYLIDGRKVR 120
            M D+IFAG+++RK LN A+V++TLDN DHF+      EV V RR++R+G+SE+LI+ +  R
Sbjct: 61   MEDIIFAGSDSRKRLNLAEVTLTLDNDDHFLPIDFHEVSVTRRVYRSGESEFLINNQPCR 120

Query: 121  LRDIHDLFMDTGLGRDSFSIISQGRVEAIFNSKPEERRAIFEEAAGVLKYKTRKKETQSK 180
            L+DI DLFMD+GLG+++FSIISQG+VE I +SK E+RR+IFEEAAGVLKYKTRKK+ ++K
Sbjct: 121  LKDIIDLFMDSGLGKEAFSIISQGKVEEILSSKAEDRRSIFEEAAGVLKYKTRKKKAENK 180

Query: 181  LEQTQGNLDRLEDIIYELDMQVQPLEKQASIAKRFLVLDEERQGLHLSILIEDILQHQSD 240
            L +TQ NL+R+EDI++EL+ QV+PL+ QASIAK +L   +E + + +++   DI +
Sbjct: 181  LFETQDNLNRVEDILHELEGQVEPLKIQASIAKDYLEKKKELEHVEIALTAYDIEKLHGK 240

Query: 241  LTTVEEKLLTVRKELATYYQQRQSLEDENQSLKQKRHHLSEEIEAKQLALLDVTKLKSDL 300
             +T++EK+   ++E         + E + +  + K   L E +   Q  LL  ++    L
Sbjct: 241  WSTLKEKVQMAKEEELAESSAISAKEAKIEDTRDKIQALDESVNELQQVLLVTSEELEKL 300

Query: 301  ERQIDLIRLESNQKAEKKEEAGQRLAELEIKAKDCSDQITQKNIELTTLSEKIAQIRSEI 360
            E + ++++       + +E+  + + + + K     ++++++     TL  ++ Q+R+++
Sbjct: 301  EGRKEVLKERKKNAVQNQEQLEEAIVQFQQKETVLKEELSKQEAVFETLQAEVKQLRAQV 360

Query: 361  VSTESSLERFSTNPDQIIEKLREDFVTLMQEEADTSNALTALLADIENQKQASQAKSQEI 420
               + +L   + N ++ IE+L+ D+  L+  +A   N L  LL D  +Q   +  +  +
Sbjct: 361  KEKQQALSLHNENVEEKIEQLKSDYFELLNSQASIRNEL-QLLDDQMSQSAVTLQRLADN 419

Query: 421  QEVSKNLEVLKSNAKVALE-RFEAAKKNVRQLLSHYQDLGQTLQNLEGEYKNQQSILFDH 479
             E         S  K A E  F   ++ +   + Y+D+   .  +  + +Y+  +S L+
Sbjct: 421  NEKHLQERHDISARKAACETEFARIEQEIHSQVGAYRDMQTKYEQKKRQYEKNESALYQA 479

Query: 480  LDEIKSKQARISSLESILKNHSNFYAGVKSVLQAKDQLGGIIGAVSEHLSFDKHYQTALE 539
              ++  +++   LE++  + S FY GVK VL+AK++LGGI GAV E +S ++ Y+TA+E
Sbjct: 479  YQYVQQARSKKDMLETMQGDFSGFYQGVKEVLKAKERLGGIRGAVLELISTEQKYETAIE 539

Query: 540  IALGGSSQHIIVEDESAAKRSIAFLKKNRQGRATFLPLTTIKPRELAQHYLSKLQSSQGF 599
            IALG S+QH++ +DE +A+++I +LK+N  GRATFLPL+ I+ R+L             F
Sbjct: 481  IALGASAQHVVTDDEQSARKAIQYLKQNSFGRATFLPLSVIRDRQLQSRDAETAARHSSF 599

Query: 600  LGIASELVTYDQRLSNIFKNNLGLTAIFDTVDNANVAARQLNYQVRLVTLDGTELRPGGS 659
            LG+ASELVT+D     ++ +N LG   I + +  AN  A+ L ++ R+VTL+G  + PGGS
Sbjct: 601  LGVASELVTFDPAYRSVIQNLLGTVLITEDLKGANELAKLLGHRYRIVTLEGDVVNPGGS 659

Query: 660  YSGGANRQNNTVFI--KPELDNLKKELKQAQSKQLIQEKEVATLLEQLKEKQETLAQLKN 717
             +GGA ++ N   +    EL+++ K L + + K  + E+EV TL   +++ ++ LA L+
Sbjct: 661  MTGGAVKKKNNSLLGRSRELEDVTKRLAEMEEKTALLEQEVKTLKHSIQDMEKKLADLRE 719

Query: 718  DGEQARLEEQRADIEYQQLSEKLADLNKLYNGLQLSSGALEQTTSENE--KNRLEKELEQ 775
             GE  RL++Q    +  +L      ++N              AL ++  E +  K +LE+EL
Sbjct: 721  TGEGLRLKQQDVKGQLYELQVAEKNINTHLELYDQEKSALSESDEERKVRKRKLEEELSA 779

Query: 776  FAIKKEELTTSIAQIKEDKDSIQEKVNNLTTLLSEAQLEERDLLNEQKFERANCTRL--- 832
             + K ++L   I ++ + K +       +L+  L+E ++          K E  N  RL
Sbjct: 781  VSEKMKQLEEDIDRLTKQKQTQSSTKESLSNELTELKIAAAKKEQACKGEEDNLARLKKE 839

Query: 833  ----EITLSEIKRDISNLQTLLSHQDSQLDKEELPRIEKQLLQVNNRRENDEEKLVSLRF 888
                E+ L E K D+S L + +S   S          E++L +    + ND+ K + L
Sbjct: 841  LTETELALKEAKEDLSFLTSEMSSSTSG---------EEKLEEAAKHKLNDKTKTIELIA 891

Query: 889  ELEDCEAALDDLAASLAKEGQKNESLIRQQAQL----ESQCEQLSQQLMIFSRQLSEDYQ 944
              D    L     +E ++ + L +Q+  L    E +  ++  +L     + L E+Y
```

```
Sbjct: 891  LRRDQRIKLQHGLDTYERELKEMKRLYKQKTTLLKDEEVKLGRMEVELDNLLQYLREEYS 950

Query: 945  MTLDEAKVKANVLEDILMAREQLKSLQAKIKALGPVNIDAIAQFEEVHERLTFLNTQRDD 1004
             ++ + AK K  +  D   AR+++K ++  I+ LG VN+ +I +FE V+ER  FL+ Q++D
Sbjct: 951  LSFEGAKEKYQLETDPEEARKRVKLIKLAIEELGTVNLGSIDEFERVNERYKFLSEQKED 1010

Query: 1005 LVHAKNLLLETITDMDDEVKTRFKSTFEAIRHSFKETFVQMFGGGSADLILTE-GDLLSA 1063
              L  AKN L + I +MD+E+  RF  TF  IR  F + F  +FGGG A+L LT+  DLL +
Sbjct: 1011 LTEAKNTLFQVIEEMDEEMTKRFNDTFVQIRSHFDQVFRSLFGGGRAELRLTDPNDLLHS 1070

Query: 1064 GVDISVQPPGKKIQSLNLMSGGEKALSALALLFAIIRVKTIPFVILDEVEAALDEANVKR 1123
              GV+I  QPPGKK+Q+LNL+SGGE+AL+A+ALLF+I++V+ +PF +LDEVEAALDEANV R
Sbjct: 1071 GVEIIAQPPGKKLQNLNLLSGGERALTAIALLFSILKVRPVPFCVLDEVEAALDEANVFR 1130

Query: 1124 FGDYLNRFDKSSQFIVVTHRKGTMSAADSIYGVTMQESGVSKIVSVKLKEAQE 1176
              F  YL ++   +QFIV+THRKGTM  AD +YGVTMQESGVSK++SVKL+E +E
Sbjct: 1131 FAQYLKKYSSDTQFIVITHRKGTMEEADVLYGVTMQESGVSKVISVKLEETKE 1183
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1163> which encodes the amino acid sequence <SEQ ID 1164>. Analysis of this protein sequence reveals the following:

```
     Possible site: 15

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.99   Transmembrane 1092 -1108 (1088 -1110)

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

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

```
>GP:CAB13467 GB:Z99112 chromosome segregation SMC protein homolg
            [Bacillus subtilis]
 Identities = 441/1192 (36%), Positives = 729/1192 (60%), Gaps = 25/1192 (2%)

Query: 1    MFLKEIELEGFKSFADKTKIEFDKGVTAVVGPNGSGKSNITESLRWALGESSAKNLRGGK 60
             MFLK +++ GFKSFA++  ++F KGVTAVVGPNGSGKSNIT+++RW LGE SA++LRGGK
Sbjct: 1    MFLKRLDVIGFKSFAERISVDFVKGVTAVVGPNGSGKSNITDAIRWVLGEQSARSLRGGK 60

Query: 61   MPDVIFAGTQNRNPLNYAKVAVVLDNSDHFIKTAKKEIRVERHIYRNGDSDYLIDGRKVR 120
             M D+IFAG+ +R  LN A+V + LDN DHF+      E+ V R +YR+G+S++LI+ +  R
Sbjct: 61   MEDIIFAGSDSRKRLNLAEVTLTLDNDDHFLPIDFHEVSVTRRVYRSGESEFLINNQPCR 120

Query: 121  LRDIHDLFMDTGLGRDSFSIISQGRVEEIFNSKPEERRAIFEEAAGVLKYKTRKKETQIK 180
             L+DI DLFMD+GLG+++FSIISQG+VEEI +SK E+RR+IFEEAAGVLKYKTRKK+ + K
Sbjct: 121  LKDIIDLFMDSGLGKEAFSIISQGKVEEILSSKAEDRRSIFEEAAGVLKYKTRKKKAENK 180

Query: 181  LNQTQDNLDRLEDIIYELDTQLAPLEKQAKVAKQFLELDANRKQLQLDILVKDIDIAQER 240
             L +TQDNL+R+EDI++EL+ Q+ PL+ QA +AK +LE     + +++ +   DI+    +
Sbjct: 181  LFETQDNLNRVEDILHELEGQVEPLKIQASIAKDYLEKKKELEHVEIALTAYDIEKLHGK 240

Query: 241  QTKDTEALAALQQDLASYYAKRQSMEEDYQKFKQKKQVLSQESDQTQTTLLELTKLIADL 300
              +   E +   +++  +  +  + E   +  + K Q L +  ++ Q  LL  ++ +  L
Sbjct: 241  WSTLKEKVQMAKEEELAESSAISAKEAKIEDTRDKIQALDESVNELQQVLLVTSEELEKL 300

Query: 301  EKQIELVKLESGQEAEKKAEAKKHLEQLQEQLDGFQAEEKQCTEQLLH-------IDQQL 353
             E + E++K        E+K  A ++ EQL+E +  FQ +E     E+L          +  ++
Sbjct: 301  EGRKEVLK-------ERKKNAVQNQEQLEEAIVQFQQKETVLKEELSKQEAVFETLQAEV 353

Query: 354  CDVKQQLNELSNALERFSSDPDQLMETLREEFVLLMQKEAALSNQLTALKAHLDKEKQAR 413
               ++ Q+ E   AL   + + ++ +E L+ ++  L+  +A++ N+L  L   + +
Sbjct: 354  KQLRAQVKEKQQALSLHNENVEEKIEQLKSDYFELLNSQASIRNELQLLDDQMSQSAVTL 413

Query: 414  QHKAQEYQLLVTKLDQLNDESQKAQAHYKAQKEQVEMLLQNYQEGDKRVQELERDYQLNQ 473
             Q  A   +  + +   ++      +  +   ++++   +  Y++   + ++ +R Y+ N+
```

```
Sbjct: 414  QRLADNNEKHLQERHDISARKAACETEFARIEQEIHSQVGAYRDMQTKYEQKKRQYEKNE 473

Query: 474  ERLFDLLDQKKGKEARKASLESIQKSHSQFYAGVRAVLQSQKKLGGIIGAVSEHLSFDSD 533
             L+      +   ++K  LE++Q   S FY GV+ VL+++++LGGI GAV E +S +
Sbjct: 474  SALYQAYQYVQQARSKKDMLETMQGDFSGFYQGVKEVLKAKERLGGIRGAVLELISTEQK 533

Query: 534  YQTALEVALGANSQHIIVTDEAAAKRAIAYLKKNRQGRATFLPLTTIKARSLSEHYHRQL 593
             Y+TA+E+ALGA++QH++  DE +A++AI YLK+N  GRATFLPL+ I+ R L
Sbjct: 534  YETAIEIALGASAQHVVTDDEQSARKAIQYLKQNSFGRATFLPLSVIRDRQLQSRDAETA 593

Query: 594  ATCEGYLGTAESLIRYDDSLSAIIQNLLSSTAIFETIDQANIAARLLGYKVRIVTLDGTE 653
             A   +LG A  L+ +D +  ++IQNLL +  I E +  AN  A+LLG++ RIVTL+G
Sbjct: 594  ARHSSFLGVASELVTFDPAYRSVIQNLLGTVLITEDLKGANELAKLLGHRYRIVTLEGDV 653

Query: 654  LRPGGSFSGGANRQSNTTFI--KPELEQISEELTRLVEQLKITEKEVAALQSDLIAKKEE 711
             + PGGS +GGA ++ N + +     ELE +++ L  + E+  + E+EV  L+  +   +++
Sbjct: 654  VNPGGSMTGGAVKKKNNSLLGRSRELEDVTKRLAEMEEKTALLEQEVKTLKHSIQDMEKK 713

Query: 712  LTQLKLAGDQARLAEQ--RAQMAYQQLQEKQEDSKALLAALDQSQTTHSDESLLAEQARI 769
             L  L+  G+  RL +Q  + Q+   Q+ EK  ++   L   ++S  + SDE     + ++
Sbjct: 714  LADLRETGEGLRLKQQDVKGQLYELQVAEKNINTHLELYDQEKSALSESDEERKVRKRKL 773

Query: 770  EEALTAIAKKKNALTCDIDDIKENKDLIRQKTQNIHQALSQARLQERDLLNEKKFEQANQ 829
             EE L+A+++K   L  DID + + K        +++   L++ ++          K E+ N
Sbjct: 774  EEELSAVSEKMKQLEEDIDRLTKQKQTQSSTKESLSNELTELKIAAAKKEQACKGEEDNL 833

Query: 830  SRLRTQLKQCQQNILKLESILNNNVSQDSIQRLPQWQKQLQDATEHKSGAQKRLVQLRFE 889
             +RL+ +L + +  + + + L+   S+ S        +++L++A +HK   + + ++L
Sbjct: 834  ARLKKELTETELALKEAKEDLSFLTSEMSSS--TSGEEKLEEAAKHKLNDKTKTIELIAL 891

Query: 890  IEDYEARLEETAEKITKESEKNDTFIRRQTKL----ETHLEQVANRLRAYAKSLSEDFQM 945
              D   +L+   +  +E ++      +++T L    E  L ++   L     + L E++ +
Sbjct: 892  RRDQRIKLQHGLDTYERELKEMKRLYKQKTTLLKDEEVKLGRMEVELDNLLQYLREEYSL 951

Query: 946  TLADAKEVTNSIDHLESAKEKLHHLQKTIRALGPINSDAINQYEEVHERLTFLTSQKTDL 1005
             +  AKE         E A++++  ++  I  LG +N  +I+++E V+ER  FL+ QK DL
Sbjct: 952  SFEGAKEKYQLETDPEEARKRVKLIKLAIEELGTVNLGSIDEFERVNERYKFLSEQKEDL 1011

Query: 1006 TKAKNLLLETINSMDSEVKARFKVTFEAIQKSFKETFTQMFGGGSADLVLTE-TDLLSAG 1064
             T+AKN L + I  MD E+  RF  TF  I+  F + F  +FGGG A+L LT+  DLL +G
Sbjct: 1012 TEAKNTLFQVIEEMDEEMTKRFNDTFVQIRSHFDQVFRSLFGGGRAELRLTDPNDLLHSG 1071

Query: 1065 IEISVQPPGKKIQSLNLMSGGEKALSALALLFAIIRVKTIPFVILDEVEAALDEANVKRF 1124
             +EI  QPPGKK+Q+LNL+SGGE+AL+A+ALLF+I++V+ +PF +LDEVEAALDEANV RF
Sbjct: 1072 VEIIAQPPGKKLQNLNLLSGGERALTAIALLFSILKVRPVPFCVLDEVEAALDEANVFRF 1131

Query: 1125 GDFLNRFDKDSQFIVVTHRKGTMAAADSIYGITMQESGVSKIVSVKLKEAQE 1176
              +L ++  D+QFIV+THRKGTM  AD +YG+TMQESGVSK++SVKL+E +E
Sbjct: 1132 AQYLKKYSSDTQFIVITHRKGTMEEADVLYGVTMQESGVSKVISVKLEETKE 1183
```

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

```
Identities = 732/1179 (62%), Positives = 911/1179 (77%)

Query: 1    MFLKEIEMQGFKSFADKTKVEFDQGVTAVVGPNGSGKSNITESLRWALGESSAKSLRGGK 60
             MFLKEIE++GFKSFADKTK+EFD+GVTAVVGPNGSGKSNITESLRWALGESSAK+LRGGK
Sbjct: 1    MFLKEIELEGFKSFADKTKIEFDKGVTAVVGPNGSGKSNITESLRWALGESSAKNLRGGK 60

Query: 61   MPDVIFAGTENRKPLNYAQVSVTLDNSDHFIENIADEVRVERRIFRNGDSEYLIDGRKVR 120
             MPDVIFAGT+NR PLNYA+V+V LDNSDHFI+    E+RVER I+RNGDS+YLIDGRKVR
Sbjct: 61   MPDVIFAGTQNRNPLNYAKVAVVLDNSDHFIKTAKKEIRVERHIYRNGDSDYLIDGRKVR 120

Query: 121  LRDIHDLFMDTGLGRDSFSIISQGRVEAIFNSKPEERRAIFEEAAGVLKYKTRKKETQSK 180
             LRDIHDLFMDTGLGRDSFSIISQGRVE IFNSKPEERRAIFEEAAGVLKYKTRKKETQ K
Sbjct: 121  LRDIHDLFMDTGLGRDSFSIISQGRVEEIFNSKPEERRAIFEEAAGVLKYKTRKKETQIK 180

Query: 181  LEQTQGNLDRLEDIIYELDMQVQPLEKQASIAKRFLVLDEERQGLHLSILIEDILQHQSD 240
             L QTQ NLDRLEDIIYELD Q+ PLEKQA +AK+FL LD  R+ L L IL++DI   Q
```

```
Sbjct: 181  LNQTQDNLDRLEDIIYELDTQLAPLEKQAKVAKQFLELDANRKQLQLDILVKDIDIAQER 240

Query: 241  LTTVEEKLLTVRKELATYYQQRQSLEDENQSLKQKRHHLSEEIEAKQLALLDVTKLKSDL 300
             T   E L  ++++LA+YY +RQS+E++ Q  KQK+  LS+E +  Q  LL++TKL +DL
Sbjct: 241  QTKDTEALAALQQDLASYYAKRQSMEEDYQKFKQKKQVLSQESDQTQTTLLELTKLIADL 300

Query: 301  ERQIDLIRLESNQKAEKKEEAGQRLAELEIKAKDCSDQITQKNIELTTLSEKIAQIRSEI 360
             E+QI+L++LES Q+AEKK EA + L +L+ +       +  Q   +L  + +++  ++ ++
Sbjct: 301  EKQIELVKLESGQEAEKKAEAKKHLEQLQEQLDGFQAEEKQCTEQLLHIDQQLCDVKQQL 360

Query: 361  VSTESSLERFSTNPDQIIEKLREDFVTLMQEEADTSNALTALLADIENQKQASQAKSQEI 420
                 ++LERFS++PDQ++E LRE+FV LMQ+EA  SN LTAL A ++ +KQA Q K+QE
Sbjct: 361  NELSNALERFSSDPDQLMETLREEFVLLMQKEAALSNQLTALKAHLDKEKQARQHKAQEY 420

Query: 421  QEVSKNLEVLKSNAKVALERFEAAKKNVRQLLSHYQDLGQTLQNLEGEYKNQQSILFDHL 480
             Q +    L+ L    ++ A    ++A K+ V  LL +YQ+  +Q LE +Y+  Q  LFD L
Sbjct: 421  QLLVTKLDQLNDESQKAQAHYKAQKEQVEMLLQNYQEGDKRVQELERDYQLNQERLFDLL 480

Query: 481  DEIKSKQARISSLESILKNHSNFYAGVKSVLQAKDQLGGIIGAVSEHLSFDKHYQTALEI 540
             D+ K K+AR +SLESI K+HS FYAGV++VLQ++ +LGGIIGAVSEHLSFD  YQTALE+
Sbjct: 481  DQKKGKEARKASLESIQKSHSQFYAGVRAVLQSQKKLGGIIGAVSEHLSFDSDYQTALEV 540

Query: 541  ALGGSSQHIIVEDESAAKRSIAFLKKNRQGRATFLPLTTIKPRELAQHYLSKLQSSQGFL 600
             ALG +SQHIIV DE+AAKR+IA+LKKNRQGRATFLPLTTIK R L++HY  +L + +G+L
Sbjct: 541  ALGANSQHIIVTDEAAAKRAIAYLKKNRQGRATFLPLTTIKARSLSEHYHRQLATCEGYL 600

Query: 601  GIASELVTYDQRLSNIFKNNLGLTAIFDTVDNANVAARQLNYQVRLVTLDGTELRPGGSY 660
             G A  L+ YD  LS I +N L  TAIF+T+D AN+AAR L Y+VR+VTLDGTELRPGGS+
Sbjct: 601  GTAESLIRYDDSLSAIIQNLLSSTAIFETIDQANIAARLLGYKVRIVTLDGTELRPGGSF 660

Query: 661  SGGANRQNNTVFIKPELDNLKKELKQAQSKQLIQEKEVATLLEQLKEKQETLAQLKNDGE 720
             SGGANRQ+NT FIKPEL+ + +EL +   + I EKEVA L   L  K+E L QLK  G+
Sbjct: 661  SGGANRQSNTTFIKPELEQISEELTRLVEQLKITEKEVAALQSDLIAKKEELTQLKLAGD 720

Query: 721  QARLEEQRADIEYQQLSEKLADLNKLYNGLQLSSGALEQTTSENEKNRLEKELEQFAIKK 780
             QARL EQRA + YQQL EK  D   L   L S       +  E+ R+E+ L   A KK
Sbjct: 721  QARLAEQRAQMAYQQLQEKQEDSKALLAALDQSQTTHSDESLLAEQARIEEALTAIAKKK 780

Query: 781  EELTTSIAQIKEDKDSIQEKVNNLTTLLSEAQLEERDLLNEQKFERANCTRLEITLSEIK 840
              LT  I  IKE+KD I++K  N+   LS+A+L+ERDLLNE+KFE+AN +RL   L + +
Sbjct: 781  NALTCDIDDIKENKDLIRQKTQNIHQALSQARLQERDLLNEKKFEQANQSRLRTQLKQCQ 840

Query: 841  RDISNLQTLLSHQDSQLDKEELPRIEKQLLQVNNRRENDEEKLVSLRFELEDCEAALDDL 900
             ++I  L+++L++  SQ  + LP+ +KQL      +  +++LV LRFE+ED EA L++
Sbjct: 841  QNILKLESILNNNVSQDSIQRLPQWQKQLQDATEHKSGAQKRLVQLRFEIEDYEARLEET 900

Query: 901  AASLAKEGQKNESLIRQQAQLESQCEQLSQQLMIFSRQLSEDYQMTLDEAKVKANVLEDI 960
             A  + KE +KN++ IR+Q +LE+  EQ++ +L  +++ LSED+QMTL +AK   N ++ +
Sbjct: 901  AESLAKEGQKNESLIRQQAQLESQCEQLSQQLMIFSRQLSEDYQMTLDEAKVKANVLEDI 960

Query: 961  LMAREQLKSLQAKIKALGPVNIDAIAQFEEVHERLTFLNTQRDDLVHAKNLLLETITDMD 1020
              A+E+L  LQ  I+ALGP+N DAI Q+EEVHERLTFL +Q+ DL  AKNLLLETI  MD
Sbjct: 961  ESAKEKLHHLQKTIRALGPINSDAINQYEEVHERLTFLTSQKTDLTKAKNLLLETINSMD 1020

Query: 1021 DEVKTRFKSTFEAIRHSFKETFVQMFGGGSADLILTEGDLLSAGVDISVQPPGKKIQSLN 1080
              EVK RFK TFEAI+ SFKETF QMFGGGSADL+LTE DLLSAG++ISVQPPGKKIQSLN
Sbjct: 1021 SEVKARFKVTFEAIQKSFKETFTQMFGGGSADLVLTETDLLSAGIEISVQPPGKKIQSLN 1080

Query: 1081 LMSGGEKALSALALLFAIIRVKTIPFVILDEVEAALDEANVKRFGDYLNRFDKSSQFIVV 1140
             LMSGGEKALSALALLFAIIRVKTIPFVILDEVEAALDEANVKRFGD+LNRFDK SQFIVV
Sbjct: 1081 LMSGGEKALSALALLFAIIRVKTIPFVILDEVEAALDEANVKRFGDFLNRFDKDSQFIVV 1140

Query: 1141 THRKGTMSAADSIYGVTMQESGVSKIVSVKLKEAQEMTN 1179
             THRKGTM+AADSIYG+TMQESGVSKIVSVKLKEAQEMTN
Sbjct: 1141 THRKGTMAAADSIYGITMQESGVSKIVSVKLKEAQEMTN 1179
```

SEQ ID 1162 (GBS199) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 2; MW 75kDa).

GBS199-GST was purified as shown in Figure 208, lane 3.

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

Example 359

A DNA sequence (GBSx0390) was identified in *S.agalactiae* <SEQ ID 1165> which encodes the amino acid sequence <SEQ ID 1166>. This protein is predicted to be ribonuclease III (rnc). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3372(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 9711> which encodes amino acid sequence <SEQ ID 9712> was also identified.

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

```
>GP:CAB13466 GB:Z99112 ribonuclease III [Bacillus subtilis]
 Identities = 115/230 (50%), Positives = 154/230 (66%), Gaps = 1/230 (0%)

Query: 13  KKMKELRSKLEKDYGIVFANQELLDTAFTHTSYANEHRLLNISHNERLEFLGDAVLQLLI 72
           KK+++ +   E+   + F N++LL  AFTH+SY NEHR       NERLEFLGDAVL+L I
Sbjct: 15  KKVEQFKEFQER-ISVHFQNEKLLYQAFTHSSYVNEHRKKPYEDNERLEFLGDAVLELTI 73

Query: 73  SQYLFTKYPQKAEGDLSKLRSMIVREESLAGFSRLCGFDHYIKLGKGEEKSGGRNRDTIL 132
           S++LF KYP  +EGDL+KLR+ IV E SL   +    F   + LGKGEE +GGR R  +L
Sbjct: 74  SRFLFAKYPAMSEGDLTKLRAAIVCEPSLVSLAHELSFGDLVLLGKGEEMTGGRKRPALL 133

Query: 133 GDLFEAFLGALLLDKGVEVVHAFVNKVMIPHVEKGTYERVKDYKTSLQELLQSHGDVKID 192
            D+FEAF+GAL LD+G+E V +F+   + P +  G +  V D+K+ LQE +Q  G   ++
Sbjct: 134 ADVFEAFIGALYLDQGLEPVESFLKVYVFPKINDGAFSHVMDFKSQLQEYVQRDGKGSLE 193

Query: 193 YQVTNESGPAHAKEFEVTVSVNQENLSQGIGRSKKAAEQDAAKNALATLQ 242
           Y+++NE GPAH +EFE  VS+  E L  G GRSKK AEQ AA+ ALA LQ
Sbjct: 194 YKISNEKGPAHNREFEAIVSLKGEPLGVGNGRSKKEAEQHAAQEALAKLQ 243
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1167> which encodes the amino acid sequence <SEQ ID 1168>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1414(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 = 170/227 (74%), Positives = 192/227 (83%)

Query: 15  MKELRSKLEKDYGIVFANQELLDTAFTHTSYANEHRLLNISHNERLEFLGDAVLQLLISQ 74
```

```
           MK+L   L   + I F +  LL+TAFTHTSYANEHRLLN+SHNERLEFLGDAVLQL+IS+
Sbjct: 1   MKQLEELLSTSFDIQFNDLTLLETAFTHTSYANEHRLLNVSHNERLEFLGDAVLQLIISE 60

Query: 75  YLFTKYPQKAEGDLSKLRSMIVREESLAGFSRLCGFDHYIKLGKGEEKSGGRNRDTILGD 134
           YLF KYP+K EGD+SKLRSMIVREESLAGFSR C FD YIKLGKGEEKSGGR RDTILGD
Sbjct: 61  YLFAKYPKKTEGDMSKLRSMIVREESLAGFSRFCSFDAYIKLGKGEEKSGGRRRDTILGD 120

Query: 135 LFEAFLGALLLDKGVEVVHAFVNKVMIPHVEKGTYERVKDYKTSLQELLQSHGDVKIDYQ 194
           LFEAFLGALLLDKG++ V  F+ +VMIP VEKG +ERVKDYKT LQE LQ+ GDV.IDYQ
Sbjct: 121 LFEAFLGALLLDKGIDAVRRFLKQVMIPQVEKGNFERVKDYKTCLQEFLQTKGDVAIDYQ 180

Query: 195 VTNESGPAHAKEFEVTVSVNQENLSQGIGRSKKAAEQDAAKNALATL 241
           V +E GPAHAK+FEV++ VN   LS+G+G+SKK AEQDAAKNALA L
Sbjct: 181 VISEKGPAHAKQFEVSIVVNGAVLSKGLGKSKKLAEQDAAKNALAQL 227
```

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

Example 360

A DNA sequence (GBSx0391) was identified in *S.agalactiae* <SEQ ID 1169> which encodes the amino acid sequence <SEQ ID 1170>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -4.19   Transmembrane  100 - 116 (  99 - 117)
   INTEGRAL    Likelihood = -2.44   Transmembrane   81 -  97 (  81 -  97)

----- Final Results -----
               bacterial membrane --- Certainty=0.2678(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAC12789 GB:AJ279090 hypothetical protein [Staphylococcus
           carnosus]
 Identities = 50/114 (43%), Positives = 72/114 (62%)

Query: 3   KIFYISLGFISLGIGIAGIVLPVVPTTPLVLLSAFCFSRSSEKFDIWLRQTKVYKYYAAD 62
           K   ++LG I  GIG  GIV+P++PTTP +LL+A CFSRSS+KF+ WL  TK++  Y
Sbjct: 2   KYVLMTLGLIFAGIGFVGIVVPLLPTTPFLLLAAICFSRSSKKFNRWLVNTKIHDEYVES 61

Query: 63  FVESRSIAPARKKSMIWQIYILMGISIYFAPLMWLKLGLLIGTIVGTYVLFYVV 116
           F   +     +K  ++  +YILMGISI+    +++++ LLI   V T VLF  V
Sbjct: 62  FKRDKGFTLKKKFKLLTSLYILMGISIFIIDNLYIRITLLIMLFVQTVVLFTFV 115
```

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 361

A DNA sequence (GBSx0392) was identified in *S.agalactiae* <SEQ ID 1171> which encodes the amino acid sequence <SEQ ID 1172>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1908(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 1173> which encodes the amino acid sequence <SEQ ID 1174>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1610(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 = 225/269 (83%), Positives = 248/269 (91%)

Query: 1   MSEIGFKYSILASGSTGNCFYIETPQKRLLIDAGLTGKKVTSLLAEINRKPEDLDAILVT 60
           M+E GFKYSILASGSTGNCFY+ETP+KRLLIDAGLTGKK+TSLLAEI+RKPEDLDAIL+T
Sbjct: 1   MNESGFKYSILASGSTGNCFYLETPKKRLLIDAGLTGKKITSLLAEIDRKPEDLDAILIT 60

Query: 61  HEHSDHIKGVGVLARKYHLDIYANEQTWKVMDERNMLGKVDVSQKHVFGRGKTLTFGDLD 120
           HEHSDHIKGVGV+ARKYHLDIYANE+TW++MDE NMLGK+D SQKH+F R K LTFGD+D
Sbjct: 61  HEHSDHIKGVGVMARKYHLDIYANEKTWQLMDECNMLGKLDASQKHIFQRDKVLTFGDVD 120

Query: 121 IESFGVSHDAVDPQFYRMMKDDKSFVMLTDTGYVSDRMAGLIENADGYLIESNHDIEILR 180
           IESFGVSHDA+DPQFYR+MKD+KSFVMLTDTGYVSDRM G+IENADGYLIESNHDIEILR
Sbjct: 121 IESFGVSHDAIDPQFYRIMKDNKSFVMLTDTGYVSDRMTGIIENADGYLIESNHDIEILR 180

Query: 181 SGSYPWTLKQRILSDKGHLSNEDGSETMIRTIGNRTKHIYLGHLSKENNIKELAHMTMEN 240
           SGSYPW+LKQRILSD GHLSNEDG+  MIR++G  TK IYLGHLSKENNIKELAHMTM N
Sbjct: 181 SGSYPWSLKQRILSDLGHLSNEDGAGAMIRSLGYNTKKIYLGHLSKENNIKELAHMTMVN 240

Query: 241 NLMRADFGVGTDFSVHDTSPDSATPLTRI 269
            L  AD  VGTDF+VHDTSPD+A PLT I
Sbjct: 241 QLAMADLAVGTDFTVHDTSPDTACPLTDI 269
```

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

Example 362

A DNA sequence (GBSx0393) was identified in *S.agalactiae* <SEQ ID 1175> which encodes the amino acid sequence <SEQ ID 1176>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.94   Transmembrane   15 -  31 (   5 -  34)

----- Final Results -----
             bacterial membrane --- Certainty=0.5776(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1177> which encodes the amino acid sequence <SEQ ID 1178>. Analysis of this protein sequence reveals the following:

```
     Possible site: 26

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

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

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

```
Identities = 335/443 (75%), Positives = 392/443 (87%)

Query: 7   NIRSFELALLFLLVFVAVYFVYLAVRDFKMSKNIRLLNWKVRDLIAGNYSDSILIQGDAD 66
           N+ +FELA+L LLVFVA YF++LAVRD++ ++ IR+++ K+RDLI G Y+D I  + D +
Sbjct: 8   NLSTFELAILILLVFVAFYFIHLAVRDYRNARIIRMMSHKIRDLINGRYTDIIDEKADIE 67

Query: 67  LVELGESLNDLSDVFRMAHDNLEQEKNRLASILTYMTDGVLATDRSGKIVMINETAQQQF 126
           L+EL + LNDLSDVFR+ H+NL QEKNRLASIL YM+DGVLATDRSGKI+MINETA++Q
Sbjct: 68  LMELSDQLNDLSDVFRLTHENLAQEKNRLASILAYMSDGVLATDRSGKIIMINETARKQL 127

Query: 127 NLAYDEALSMNIVDMLGSGSPYSFQDLVSKTPEVVLNRRDENGEFVTLRIRFALNRRESG 186
           NL+ +EAL  NI D+L   + Y+++DLVSKTP V +N R++ GEFV+LR+RFALNRRESG
Sbjct: 128 NLSKEEALKKNITDLLEGDTSYTYRDLVSKTPVVTVNSRNDMGEFVSLRLRFALNRRESG 187

Query: 187 FISGLVAVSHDATEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGALNEEVAPSF 246
           FISGLV V HD TEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGAL E++APSF
Sbjct: 188 FISGLVVVLHDTTEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGALKEDIAPSF 247

Query: 247 IKVSLDETNRMMRMISDLLSLSRIDNEVTHLDVEMTNFTAFMTSILNRFDQIRNQKTVTG 306
           IKVSLDETNRMMRMISDLL+LSRIDN+VT L VEMTNFTAF+TSILNRFD ++NQ T TG
Sbjct: 248 IKVSLDETNRMMRMISDLLNLSRIDNQVTQLAVEMTNFTAFITSILNRFDLVKNQHTGTG 307

Query: 307 KVYEIVRDYPLKSIWVEIDTDKMTQVIDNILNNAVKYSPDGGKITVNLRTTKTQMILSIS 366
           KVYEIVRDYP+ S+W+EID DKMTQVI+NILNNA+KYSPDGGKITV ++TT TQ+I+SIS
Sbjct: 308 KVYEIVRDYPITSVWIEIDNDKMTQVIENILNNAIKYSPDGGKITVRMKTTDTQLIISIS 367

Query: 367 DQGLGIPKKDLPLIFDRFYRVDKARSRKQGGTGLGLSIAKEIVKQHKGFIWAKSEYGKGS 426
           DQGLGIPK DLPLIFDRFYRVDKARSR QGGTGLGL+IAKEI+KQH GFIWAKS+YGKGS
Sbjct: 368 DQGLGIPKTDLPLIFDRFYRVDKARSRAQGGTGLGLAIAKEIIKQHHGFIWAKSDYGKGS 427

Query: 427 TFTIVLPYDKDAVTYEEWEDVED 449
           TFTIVLPY+KDA  YEEWE+  D
Sbjct: 428 TFTIVLPYEKDAAIYEEWEEDVD 450
```

A related GBS gene <SEQ ID 8561> and protein <SEQ ID 8562> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      8.59
GvH: Signal Score (-7.5): -3.38
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value: -11.94 threshold:  0.0
   INTEGRAL    Likelihood =-11.94   Transmembrane   15 -  31 (   5 -  34)
   PERIPHERAL  Likelihood =  8.27     178
 modified ALOM score:   2.89

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5776(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
67.5/83.5% over 439aa
                              Streptococcus pneumoniae
 GP|5830524| histidine kinase Insert characterized

ORF01458(331 - 1647 of 1947)
GP|5830524|emb|CAB54569.1||AJ006392(10  -  449  of  449)  histidine  kinase  (Streptococcus
pneumoniae)
%Match = 45.6
```

```
%Identity = 67.5  %Similarity = 83.4
Matches = 297  Mismatches = 70  Conservative Sub.s = 70

126       156       186       216       246       276       306       336
ITSPFSDTYRTSHDTRTFIGNSLGI*LFWRCPYS*CDGETFT*KD*RYSWSSRIYFDSTWCRXIT*SLMNNSAANIRSFE
                                                                          |
                                                                        MLDLLKQTIFT
                                                                              10

366       396       426       450       480       510       540       570
LALLFLLVFVAVYFVYLAVRDFKMSKNIRL--LNWKVRDLIAGNYSDSILIQGDADLVELGESLNDLSDVFRMAHDNLEQ
  ::|:|:::   :|   :       ||:|  :| ||:|||||:||  : :|| :::  :  :|||||:| |: ::||||
RDFIFILILLGFILVVTLLLLENRRDNIQLKQVNQKVKDLIAGDYSKVLDMQGGSEITNITNNLNDLSEVIRLTQENLEQ
                  30        40        50        60        70        80        90

600       630       660       690       720       750       780       810
EKNRLASILTYMTDGVLATDRSGKIVMINETAQQQFNLAYDEALSMNIVDMLGSGSPYSFQDLVSKTPEVVLNRRDENGE
|  || ||| |||||||||:| |:|:|||:||::|: |  :: |: :|:::|     | ::||::::||::|: :| |||
ESKRLNSILFYMTDGVLATNRRGQIIMINDTAKKQLGLVKEDVLNRSILELLKIEENYELRDLITQSPELLLDSQDINGE
                  110       120       130       140       150       160       170

840       870       900       930       960       990       1020      1050
FVTLRIRFALNRRESGFISGLVAVSHDATEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGALNEEVAPSFIKVS
:: ||:|||| |||||||||||||| || |||||||||||||||||||||||||||||||||||||| | ||| |||||
YLNLRVRFALIRRESGFISGLVAVLHDTTEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGALCETVAPDFIKVS
                  190       200       210       220       230       240       250

1080      1110      1140      1170      1200      1230      1260      1290
LDETNRMMRMISDLLSLSRIDNEVTHLDVEMTNFTAFMTSILNRFDQIRNQKTVTGKVYEIVRDYPLKSIWVEIDTDKMT
||||||||||::||| |||||| :|||||: |||||:| |||||||::: |     | ||:|||||: |||:||||||||
LDETNRMMRMVTDLLHLSRIDNATSHLDVELINFTAFITFILNRFDKMKGQ--EKEKKYELVRDYPINSIWMEIDTDKMT
                  270       280       290       300       310       320

1320      1350      1380      1410      1440      1470      1500      1530
QVIDNILNNAVKYSPDGGKITVNLRTTKTQMILSISDQGLGIPKKDLPLIFDRFYRVDKARSRKQGGTGLGLSIAKEIVK
||:|||||||:|||||||||| ::||: ||||||||:||||||:||| |||||||||:|||| |||||||||||||:|
QVVDNILNNAIKYSPDGGKITVRMKTTEDQMILSISDHGLGIPKQDLPRIFDRFYRVDRARSRAQGGTGLGLSIAKEIIK
          340       350       360       370       380       390       400

1560      1590      1620      1647      1677      1707      1737      1767
QHKGFIWAKSEYGKGSTFTIVLPYDKDAVTYEEWED-VED*NMSEIGFKYSILASGSTGNCFYIETPQKRLLIDAGLTGK
|||||||||||||||||||||||||||||| | ||| |||
QHKGFIWAKSEYGKGSTFTIVLPYDKDAVKEEVWEDEVED
          420       430       440
```

SEQ ID 1176 (GBS41) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 7; MW 50kDa), in Figure 168 (lane 2-4; MW 65kDa – thioredoxin fusion) and in Figure 238 (lane 4; MW 65kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 7; MW 75kDa).

Purified Thio-GBS41-His is shown in Figure 244, lane 10.

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

Example 363

A DNA sequence (GBSx0394) was identified in *S.agalactiae* <SEQ ID 1179> which encodes the amino acid sequence <SEQ ID 1180>. This protein is predicted to be VicR protein (regX3). Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2754(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 1181> which encodes the amino acid sequence <SEQ ID 1182>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2754(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 = 205/236 (86%), Positives = 221/236 (92%)

Query: 1   MKKILIVDDEKPISDIIKFNLTKEGYETATAFDGREALVQYAEFQPDLIILDLMLPELDG 60
           MKKILIVDDEKPISDIIKFNLTKEGY+  TAFDGREA+  + E +PDLIILDLMLPELDG
Sbjct: 1   MKKILIVDDEKPISDIIKFNLTKEGYDIVTAFDGREAVTIFEEEKPDLIILDLMLPELDG 60

Query: 61  LEVAKEVRKTSHIPIIMLSAKDSEFDKVIGLEIGADDYVTKPFSNRELLARVKAHLRRTE 120
           LEVAKE+RKTSH+PIIMLSAKDSEFDKVIGLEIGADDYVTKPFSNRELLARVKAHLRRTE
Sbjct: 61  LEVAKEIRKTSHVPIIMLSAKDSEFDKVIGLEIGADDYVTKPFSNRELLARVKAHLRRTE 120

Query: 121 NIETAVAEESAQNASSDITIGELQILPDAFIAKKRGEEIELTHREFELLHHLATHIGQVM 180
            IETAVAEE+A + + ++TIG LQILPDAF+AKK G+E+ELTHREFELLHHLA H+GQVM
Sbjct: 121 TIETAVAEENASSGTQELTIGNLQILPDAFVAKKHGQEVELTHREFELLHHLANHMGQVM 180

Query: 181 TREHLLETVWGYDYFGDVRTVDVTVRRLREKIEDTPGRPEYILTRRGVGYYMKSYE 236
           TREHLLE VWGYDYFGDVRTVDVTVRRLREKIEDTP RPEYILTRRGVGYYMKSY+
Sbjct: 181 TREHLLEIVWGYDYFGDVRTVDVTVRRLREKIEDTPSRPEYILTRRGVGYYMKSYD 236
```

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

Example 364

A DNA sequence (GBSx0395) was identified in *S.agalactiae* <SEQ ID 1183> which encodes the amino acid sequence <SEQ ID 1184>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3791(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:CAB14701 GB:Z99118 glutamine ABC transporter (ATP-binding
          protein) [Bacillus subtilis]
 Identities = 149/244 (61%), Positives = 200/244 (81%), Gaps = 2/244 (0%)

Query: 3   LISYKNVNKYYGDYHALRQINLEIEPGQVVVLLGPSGSGKSTLIRTMNALESIDDGSLVV 62
           +I+++NVNK+YGD+H L+QINL+IE G+VVV++GPSGSGKSTL+R +N LESI++G L V
Sbjct: 1   MITFQNVNKHYGDFHVLKQINLQIEKGEVVVIIGPSGSGKSTLLRCINRLESINEGVLTV 60
```

```
Query: 63  NGHELANISSKELVNLRKEVGMVFQHFNLYPHKTVLENITLAPIKVLKQSKKEAMEIAEK 122
           NG  + N    ++  +R+ +GMVFQHF+LYPHKTVL+NI LAP+KVL+QS ++A E A
Sbjct: 61  NGTAI-NDRKTDINQVRQNIGMVFQHFHLYPHKTVLQNIMLAPVKVLRQSPEQAKETARY 119

Query: 123 YLKFVNMWERKDSYPSMLSGGQKQRIAIARGLAMHPKLLLFDEPTSALDPETIGDVLSVM 182
           YL+ V + ++ D+YPS LSGGQ+QR+AIARGLAM P+++LFDEPTSALDPE IG+VL VM
Sbjct: 120 YLEKVGIPDKADAYPSQLSGGQQQRVAIARGLAMKPEVMLFDEPTSALDPEMIGEVLDVM 179

Query: 183 QKLANDGMNMVVVTHEMGFAREVADRIIFMADGEILVDTTDVQDFFDNPREPRAKQFLSN 242
           + LA +GM MVVVTHEMGFA+EVADRI+F+ +G+IL +     +F+ NP+E RA+ FLS
Sbjct: 180 KTLAKEGMTMVVVTHEMGFAKEVADRIVFIDEGKILEEAVPA-EFYANPKEERARLFLSR 238

Query: 243 IINH 246
           I+NH
Sbjct: 239 ILNH 242
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1185> which encodes the amino acid sequence <SEQ ID 1186>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3763(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 = 131/243 (53%), Positives = 179/243 (72%), Gaps = 2/243 (0%)

Query: 2   SLISYKNVNKYYGDYHALRQINLEIEPGQVVVLLGPSGSGKSTLIRTMNALESIDDGSLV 61
           ++IS K+++KYYG    L+ I+L+I PG+VVV++GPSGSGKSTL+RTMN LE    G +
Sbjct: 5   AIISIKDLHKYYGHNEVLKGIDLDIMPGEVVVIIGPSGSGKSTLLRTMNLLEVPTKGQIR 64

Query: 62  VNGHELANISSKELVNLRKEVGMVFQHFNLYPHKTVLENITLAPIKVLKQSKKEAMEIAE 121
             G ++ +    ++ ++R+++GMVFQ FNL+P+ T+LENITL+PIK    +K EA + A
Sbjct: 65  FEGIDITD-KKNDIFSMREKMGMVFQQFNLFPNMTILENITLSPIKTKGMAKAEADKTAL 123

Query: 122 KYLKFVNMWERKDSYPSMLSGGQKQRIAIARGLAMHPKLLLFDEPTSALDPETIGDVLSV 181
             L  V + E+  +YP+ LSGGQ+QRIAIARGLAM P +LLFDEPTSALDPE +G+VL+V
Sbjct: 124 SLLDKVGLSEKAKAYPASLSGGQQQRIAIARGLAMDPDVLLFDEPTSALDPEMVGEVLAV 183

Query: 182 MQKLANDGMNMVVVTHEMGFAREVADRIIFMADGEILVDTTDVQDFFDNPREPRAKQFLS 241
           MQ LA  GM MV+VTHEMGFA+EVADR++FM DG ++V+       FD  +E R K FLS
Sbjct: 184 MQDLAKSGMTMVIVTHEMGFAKEVADRVMFM-DGGVIVEEGSPNQLFDLTKEERTKDFLS 242

Query: 242 NII 244
           ++
Sbjct: 243 RVL 245
```

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

Example 365

A DNA sequence (GBSx0396) was identified in *S.agalactiae* <SEQ ID 1187> which encodes the amino acid sequence <SEQ ID 1188>. This protein is predicted to be glutamine-binding. 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:CAB73178 GB:AL139076 probable ABC-type amino-acid transporter
          periplasmic solute-binding protein [Campylobacter
          jejuni]
 Identities = 99/240 (41%), Positives = 141/240 (58%), Gaps = 3/240 (1%)

Query: 1   MLRRKRLTFYLLSCIFIFLLFYPNSTSANQLSEIKKSGVLKVGVKQDVPNFGYYNAETNQ 60
           M+ RK L    +  +   + F   + +  +L  IK  G L VGVK DVP++   +  T +
Sbjct: 1   MVFRKSLLLKLAVFALGACVAFSNANAAEGKLESIKSKGQLIVGVKNDVPHYALLDQATGE 60

Query: 61  YEGMEIDIAKKIAKSL---GVKPVFVPTTAQTREPLMDNGQIDILIATYTITPERKANYN 117
            +G E+D+AK +AKS+      K   V   A+TR PL+DNG +D +IAT+TITPERK  YN
Sbjct: 61  IKGFEVDVAKLLAKSILGDDKKIKLVAVNAKTRGPLLDNGSVDAVIATFTITPERKRIYN 120

Query: 118 ISKAYYHDEIGFLVRKNSHIKTIKELDGKHIGVAQGATTKVNLEKYAKEHKLKFSYAQLG 177
            S+ YY D IG LV K    K++ ++ G +IGVAQ ATTK  + + AK+  +   +++
Sbjct: 121 FSEPYYQDAIGLLVLKEKKYKSLADMKGANIGVAQAATTKKAIGEAAKKIGIDVKFSEFP 180

Query: 178 SFPELAISLYANRIDAFSVDKSILSGYLSPHTTILKEGFNTQEYGIATSKQDKVLIPYVN 237
           +P +  +L A R+DAFSVDKSIL GY+   + IL + F  Q YGI T K D     YV+
Sbjct: 181 DYPSIKAALDAKRVDAFSVDKSILLGYVDDKSEILPDSFEPQSYGIVTKKDDPAFAKYVD 240
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1189> which encodes the amino acid sequence <SEQ ID 1190>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.16   Transmembrane   17 -  33 (  15 -  35)

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

A related sequence was also identified in GAS <SEQ ID 9097> which encodes the amino acid sequence <SEQ ID 9098>. Analysis of this protein sequence reveals the following:

```
>>> May be a lipoprotein

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

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

```
Identities = 66/251 (26%), Positives = 111/251 (43%), Gaps = 27/251 (10%)

Query: 23  PNSTSANQLSEIKKSGVLKVGVKQDVPNFGYYNAETNQYEGMEIDIAKKIAKSLGVKPVF 82
           P+ +  +    IK+ GVLKV          +YN + N+  G E+D+ K+I K L +K  F
Sbjct: 34  PHQSQKSSWDTIKEKGVLKVATPGTYQPTSFYN-DNNELVGYEVDMVKEIGKRLNIKVKF 92

Query: 83  VPTTAQTREPLMDNGQIDILIATYTITPERKANYNISKAYYHDEIGFLVR----KNSHIK 138
           V T        +D+G++DI +  + ITP+R+  YNIS  Y +   G +VR     N   K
Sbjct: 93  VETGFDQAFTSVDSGRVDISLNNFDITPKRQKKYNISTPYKYGVGGMIVRADGSSNIAKK 152

Query: 139 TIKELDGKHIGVAQGATTKVNLEKYAKEHKLKFSYAQLGSFPELAISLYANRI------- 191
            + +  GK    A G       +K          A+L ++  +  +Y N +
Sbjct: 153 DLSDWKGKKAAGASGTEYMKVAQKQG---------AELVTYDNVTGDVYLNDVANGRTDF 203

Query: 192 --DAFSVDKSILSGYLSPHTTILKE----GFNTQEYGIATSKQDKVLIPYVNKLLVSWEK 245
```

```
            + +   K  +   LS +  +  +     +N  E GI  +K+D  L   ++ ++    K
Sbjct: 204 IPNDYPAQKLFVDYMLSQNPNLNVKMSDVQYNPTEQGIVMNKKDDSLKKKIDAVIKDMIK 263

Query: 246 DGSLKHIYQKF 256
           DGSLK I + +
Sbjct: 264 DGSLKKISETY 274
```

SEQ ID 1188 (GBS136) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 5; MW 29.9kDa).

The GBS136-His fusion product was purified (Figure 200, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 284), 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 366

A DNA sequence (GBSx0397) was identified in *S.agalactiae* <SEQ ID 1191> which encodes the amino acid sequence <SEQ ID 1192>. This protein is predicted to be integral membrane. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.34   Transmembrane   32 -  48 (  27 -  55)
   INTEGRAL    Likelihood = -5.04   Transmembrane  200 - 216 ( 196 - 219)
   INTEGRAL    Likelihood = -3.13   Transmembrane   93 - 109 (  93 - 113)
   INTEGRAL    Likelihood = -2.02   Transmembrane   74 -  90 (  74 -  92)

----- Final Results -----
               bacterial membrane --- Certainty=0.4736(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:CAB73177 GB:AL139076 putative ABC-type amino-acid transporter
          permease protein [Campylobacter jejuni]
 Identities = 112/226 (49%), Positives = 160/226 (70%), Gaps = 3/226 (1%)

Query: 5   NISPFAISRWGAFFNHFDLFFKGFLYTLGISFGALLLALILGILSGGLSTSKSKVGKLIS 64
           +ISPFA+ ++    ++ D F  GF+YTL +S  ALL+A I G + G ++TS+ K+ +  +
Sbjct: 25  SISPFAVWKFLDALDNKDAFINGFIYTLEVSILALLIATIFGTIGGVMATSRFKIIRAYT 84

Query: 65  RIYVEVFQNTPLLVQMVFVYYGLAIISNGHVMISAFFTAVLCVGLYHGAYISEVIRSGIE 124
           RIYVE+FQN PL++Q+ F++Y L ++    + +  F   VL VG YHGAY+SEV+RSGI
Sbjct: 85  RIYVELFQNVPLVIQIFFLFYALPVLG---IRLDIFTIGVLGVGAYHGAYVSEVVRSGIL 141

Query: 125 AVPKGQTEAALAQGFTANQTMQLIILPQAVRTILPPMTNQVVNLIKNTSTVAIISGADIM 184
           AVP+GQ EA+ +QGFT  Q M+ II+PQ +R ILPPMTNQ+VNLIKNTS + I+ GA++M
Sbjct: 142 AVPRGQFEASASQGFTYIQQMRYIIVPQTIRIILPPMTNQMVNLIKNTSVLLIVGGAELM 201

Query: 185 FVAKAWAYDTTNYIPAFAGAAIFYFVICFPLASWARKQEELNKKTY 230
             A ++A D  NY PA+  AA+ YF+IC+PLA +A+  E   KK +
Sbjct: 202 HSADSYAADYGNYAPAYIFAAVLYFIICYPLAYFAKAYENKLKKAH 247
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1193> which encodes the amino acid sequence <SEQ ID 1194>. Analysis of this protein sequence reveals the following:

```
     Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
```

```
   INTEGRAL    Likelihood = -6.26   Transmembrane  307 - 323 ( 303 - 327)
   INTEGRAL    Likelihood = -5.89   Transmembrane  485 - 501 ( 479 - 502)
   INTEGRAL    Likelihood = -1.12   Transmembrane  375 - 391 ( 375 - 391)

----- Final Results -----
              bacterial membrane --- Certainty=0.3506(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:BAA17584 GB:D90907 glutamine-binding periplasmic protein
          [Synechocystis sp.]
 Identities = 146/532 (27%), Positives = 244/532 (45%), Gaps = 59/532 (11%)

Query: 6   YMKKLILSCLVALALLFGGMSRAQANQYLRVGMEAAYAPFNWTQDDASNGAVPIEGTSQY 65
           Y   L L  L+A+A+          + Q + V  E  + PF  T            E T Q
Sbjct: 16  YYLLLALGVLLAIAIPLLPAFSQVSRQTIIVATEPTFPPFEMTD----------EATGQL 65

Query: 66  ANGYDVQVAKKVAKAMNKELLVVKTSWTGLIPALTSGKIDMIAAGMSPTKERRNEISFSN 125
             G+DV + + + +A    + +    + G+IPAL S  +    + ++ T ER   +SFS+
Sbjct: 66  T-GFDVDLIQAIGEAAQVTVDIQGYPFDGIIPALQSNTVGAAISAITITPERAQSVSFSS 124

Query: 126 SSYTSQPVLVVTANGKYADATSLKDFSGAKVTAQQGVWHVNLLTQLKGAKLQTPMGDFSQ 185
             + S  VL +          +LKD  G ++    G     + T + GAK+ T     +
Sbjct: 125 PYFKS--VLAIAVQDGNDTIKNLKDLEGKRLAVAIGTTGAMVATNVPGAKV-TNFDSITS 181

Query: 186 MRQALTSGVIDAYISERPEAMTAEAADSRLKMITLKKGFAVAESDAAIAVGMKKNDDRMA 245
             Q L +G  DA I++RP  + A   D+ L+ + +      +E    IA+ +    +
Sbjct: 182 ALQELVNGNADAVINDRPVLLYA-IKDAGLRNVKISADVG-SEDYYGIAMPLAPPGE--- 236

Query: 246 TVNQVLEGFSQTDRMALMDDMVTKQPVEKKAEDAKASFLGQMWAIFKGN---------- 294
            +NQ  E  +Q     ++++       EK   +    FL  +    G
Sbjct: 237 -INQTREVLNQ-GLFQIIENGTYNAIYEKWFGEKNPPFLPLVAPSLVGKVGTAQSLTERS 294

Query: 295 ----------WKQFLRGTGMTLLISMVGTITGLFIGLLIGIFRTAPKAKHKVAALGQK 342
                    ++   +G+ +T+L++      GL  G  + I   +            K
Sbjct: 295 QANPNDNFLITLFRNLFKGSILTVLLTAFSVFFGLIGGTGVAIALISDI----------K 344

Query: 343 LFGWLLTIYIEIFRGTPMIVQSMVIYYGTAQAF-----GISIDRTLAAIFIVSINTGAYM 397
             +  IY+E FRGTPM+VQ  +IY+G    F      GI+IDR  AAI  +S+N  AY+
Sbjct: 345 PLQLIFRIYVEFFRGTPMLVQLFIIYFGLPALFKEIGLGITIDRFPAAIIALSLNVAAYL 404

Query: 398 SEIVRGGIFAVDKGQFKAATALGFTHGQTMRKIVLPQVVRNILPATGNEFVINIKDTSVL 457
           +EI+RGGI ++D+GQ++A  +LG +  QTM++++ PQ  R ILP  GNEF+  IKDTS+
Sbjct: 405 AEIIRGGIQSIDQGQWEACESLGMSPWQTMKEVIFPQAFRRILPPLGNEFITLIKDTSLT 464

Query: 458 NVISVVELYFSGNTVATQTYQYFQTFTIIAIIYFVLTFTVTRILRYIERRFD 509
            VI   EL+  G  +   TY+ F+ +  +A++Y +LT   + + +++E   D
Sbjct: 465 AVIGFQELFREGQLIVATTYRAFEVYIAVALVYLLLTTISSFVFKWLENYMD 516
```

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

```
 Identities = 82/210 (39%), Positives = 113/210 (53%), Gaps = 12/210 (5%)

Query: 14  WGAFFNHFDLFFKGFLYTLGISFGALLLALILGILSGGLSTS---KSKVGKL-------I 63
           W  F  ++  F +G   TL IS    +  L +G+L G   T+   K KV  L       +
Sbjct: 288 WAIFKGNWKQFLRGTGMTLLISMVGTITGLFIGLLIGIFRTAPKAKHKVAALGQKLFGWL 347

Query: 64  SRIYVEVFQNTPLLVQMVFVYYGLAIISNGHVMISAFFTAVLCVGLYHGAYISEVIRSGI 123
             IY+E+F+ TP++VQ + +YYG A       + I     A+  V +  GAY+SE++R GI
Sbjct: 348 LTIYIEIFRGTPMIVQSMVIYYGTAQAFG--ISIDRTLAAIFIVSINTGAYMSEIVRGGI 405

Query: 124 EAVPKGQTEAALAQGFTANQTMQLIILPQAVRTILPPMTNQVVNLIKNTSTVAIISGADI 183
            AV KGQ +AA A GFT  QTM+ I+LPQ VR ILP   N+ V  IK+TS + +IS  ++
Sbjct: 406 FAVDKGQFKAATALGFTHGQTMRKIVLPQVVRNILPATGNEFVINIKDTSVLNVISVVEL 465

Query: 184 MFVAKAWAYDTTNYIPAFAGAAIFYFVICF 213
```

```
            F    A  T  Y   F   AI YFV+ F
Sbjct: 466 YFSGNTVATQTYQYFQTFTIIAIIYFVLTF 495
```

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

Example 367

A DNA sequence (GBSx0398) was identified in *S.agalactiae* <SEQ ID 1195> which encodes the amino acid sequence <SEQ ID 1196>. This protein is predicted to be amino acid ABC transporter, permease protein. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.95   Transmembrane   25 -  41 (  16 -  42)
   INTEGRAL    Likelihood = -3.61   Transmembrane   66 -  82 (  65 -  86)
   INTEGRAL    Likelihood = -2.44   Transmembrane  184 - 200 ( 182 - 201)
   INTEGRAL    Likelihood = -0.59   Transmembrane  119 - 135 ( 119 - 135)

----- Final Results -----
              bacterial membrane --- Certainty=0.3781(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB14704 GB:Z99118 glutamine ABC transporter (integral membrane
          protein) [Bacillus subtilis]
 Identities = 84/206 (40%), Positives = 129/206 (61%), Gaps = 6/206 (2%)

Query: 10  ILFLLQGFGLTLYISFISILLSMFFGTLLAIMRNSKNPIWKLIASIYIEFVRNVPNLLWI 69
           + FL  GF +TLY++FISI+LS FFG +   +R +K P+   + ++ +E +RN+P LL I
Sbjct: 12  LAFLWDGFLVTLYVAFISIILSFFFGLIAGTLRYAKVPVLSQLIAVLVETIRNLPLLLII 71

Query: 70  FIIFLVF-----QMKSVSAGITSFTIFTSAALAEIIRGGLNGVDKGQTEAGLSQGFTYLQ 124
           F  F        +++  +A IT+ TIF SA L+EIIR GL  +DKGQ EA  S G +Y Q
Sbjct: 72  FFTFFALPEIGIKLEITAAAITALTIFESAMLSEIIRSGLKSIDKGQIEAARSSGLSYTQ 131

Query: 125 VFIIIIFPQAFRKMLPAIISQFVTVIKDTSLLYSVIAIQEIFGKSQILMGRYFEAGQVFT 184
               I+ PQA R+M+P I+SQF++++KDTSL   VIA+ E+   +QI+ G+  +    F
Sbjct: 132 TLFFIVMPQALRRMVPPIVSQFISLLKDTSLAV-VIALPELIHNAQIINGQSADGSYFFP 190

Query: 185 LYAIITAVYFITNFIISSFSRKLSKR 210
           ++ +   +YF  N+ +S  +R+L  R
Sbjct: 191 IFLLAALMYFAVNYSLSLAARRLEVR 216
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1197> which encodes the amino acid sequence <SEQ ID 1198>. Analysis of this protein sequence reveals the following:

```
     Possible site: 20
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.51   Transmembrane  529 - 545 ( 517 - 551)
   INTEGRAL    Likelihood =-10.30   Transmembrane  697 - 713 ( 693 - 719)
   INTEGRAL    Likelihood = -4.41   Transmembrane  560 - 576 ( 555 - 585)
   INTEGRAL    Likelihood = -0.32   Transmembrane  662 - 678 ( 662 - 678)

----- 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 databases:

```
>GP:BAA17584 GB:D90907 glutamine-binding periplasmic protein
```

```
            [Synechocystis sp.]
 Identities = 153/475 (32%), Positives = 251/475 (52%), Gaps = 27/475 (5%)

Query: 273 IVSDSSFAPFEFQN-GKGKYVGIDIELIKAIAKQQGFKIEIANPGFDAALNAVQSSQADG 331
           + ++ +F PFE +   G+  G D++LI+AI +      ++I    FD  + A+QS+
Sbjct: 46  VATEPTFPPFEMTDEATGQLTGFDVDLIQAIGEAAQVTVDIQGYPFDGIIPALQSNTVGA 105

Query: 332 VIAGATITDARKAIFDFSDPYYTSNIILAVKAGKN-IKNYEDLDRKTVGAKNGTSSYSWL 390
            I+  TIT  R     FS PY+ S + +AV+ G + IKN +DL+ K +    GT+  + +
Sbjct: 106 AISAITITPERAQSVSFSSPYFKSVLAIAVQDGNDTIKNLKDLEGKRLAVAIGTTG-AMV 164

Query: 391 KENAPKYGYNVKAFDDGSSMYDSLNSGSVDAIMDDEAVLKYAISQG--RRFETPLEGIST 448
             N P  G  V  FD  +S    L +G+ DA+++D  VL YAI     R  +   + S
Sbjct: 165 ATNVP--GAKVTNFDSITSALQELVNGNADAVINDRPVLLYAIKDAGLRNVKISADVGSE 222

Query: 449 GEVGFAVKKGTNPELI---EMFNNGLAALKKSGQYDDIIDKYLDSKKA-----ATPSEKG 500
             G A+      E+    E+ N GL  + ++G Y+ I +K+   K          PS  G
Sbjct: 223 DYYGIAMPLAPPGEINQTREVLNQGLFQIIENGTYNAIYEKWFGEKNPPFLPLVAPSLVG 282

Query: 501 -----------ADESTISGLLSNNYKQLLAGLGTTLSLTLISFAIAIIIGIIFGMMAVSP 549
                    +  +    L   ++ L  G   T+ LT  S    +I G    +  +S
Sbjct: 283 KVGTAQSLTERSQANPNDNFLITLFRNLFKGSILTVLLTAFSVFFGLIGGTGVAIALISD 342

Query: 550 TKSLRLISTVFVDVVRGIPLMIVAAFIFWGVPNLIESMTGHQSPINDFLAATIALSLNGG 609
            K L+LI  ++V+  RG P+++     I++G+P L + + G    I+ F AA IALSLN
Sbjct: 343 IKPLQLIFRIYVEFFRGTPMLVQLFIIYFGLPALFKEI-GLGITIDRFPAAIIALSLNVA 401

Query: 610 AYIAEIVRGGIEAVPAGQMEASRSLGLSYGTTMRKVILPQAVKLMLPNFINQFVISLKDT 669
           AY+AEI+RGGI+++  GQ EA  SLG+S   TM++VI PQA + +LP   N+F+  +KDT
Sbjct: 402 AYLAEIIRGGIQSIDQGQWEACESLGMSPWQTMKEVIFPQAFRRILPPLGNEFITLIKDT 461

Query: 670 TIVSAIGLVELFQTGKIIIARNYQSFRMYAILAIIYLIMIILLTRLAKRLEKRLN 724
           ++ + IG  ELF+ G++I+A  Y++F +Y  +A++YL++  + + + K LE  ++
Sbjct: 462 SLTAVIGFQELFREGQLIVATTYRAFEVYIAVALVYLLLTTISSFVFKWLENYMD 516
 Identities = 68/247 (27%), Positives = 106/247 (42%), Gaps = 11/247 (4%)

Query: 7   VLLLAIMSIFLTCNIASAETIAIVSDTAYAPFEFKD--SDQIYKGIDVDIINEVAKRQSW 64
           VLL   + +    + S +TI + ++  + PFE  D  + Q+  G DVD+I  + +
Sbjct: 24  VLLAIAIPLLPAFSQVSRQTIIVATEPTFPPFEMTDEATGQL-TGFDVDLIQAIGEAAQV 82

Query: 65  DFSMSFPGFDAAVNAVQSGQASALMAGTTITNARKKVFHFSEPYYDTKIVIATRKAN-AI 123
             +   FD  + A+QS    A ++  TIT  R +   FS PY+ + + IA +  N  I
Sbjct: 83  TVDIQGYPFDGIIPALQSNTVGAAISAITITPERAQSVSFSSPYFKSVLAIAVQDGNDTI 142

Query: 124 KKYSDLKGKTVGVKNGTAAQAFLNNYKKKYDYTVKTFDTGDLMYNSLSAGSIAAVMDDEA 183
           K  DL+GK + V  GT       N       V  FD+       L  G+  AV++D
Sbjct: 143 KNLKDLEGKRLAVAIGTTGAMVATNVP---GAKVTNFDSITSALQELVNGNADAVINDRP 199

Query: 184 VIQYAIS----QNQDIAINMKGEPIGSFGFAVKKGSGYDYLVNDFNTALKAMKADGTYQA 239
           V+ YAI     +N  I+ ++  E         +       N  L  +  GTY A
Sbjct: 200 VLLYAIKDAGLRNVKISADVGSEDYYGIAMPLAPPGEINQTREVLNQGLFQIIENGTYNA 259

Query: 240 IMTKWLG 246
           I  KW G
Sbjct: 260 IYEKWFG 266
```

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

```
 Identities = 68/210 (32%), Positives = 113/210 (53%), Gaps = 16/210 (7%)

Query: 13  LLQGFGLTLYISFISILLSMFFGTLLAIMRNSKNPIWKLIASIYIEFVRNVPNLLWIFII 72
           LL G G TL ++ IS  +++  G +  +M  S     +LI++++++ VR +P ++    I
Sbjct: 517 LLAGLGTTLSLTLISFAIAIIIGIIFGMMAVSPTKSLRLISTVFVDVVRGIPLMIVAAFI 576

Query: 73  F-----LVFQMKSVSAGITSFTIFT-------SAALAEIIRGGLNGVDKGQTEAGLSQGF 120
           F     L+  M    + I  F   T         A +AEI+RGG+  V  GQ EA  S G
Sbjct: 577 FWGVPNLIESMTGHQSPINDFLAATIALSLNGGAYIAEIVRGGIEAVPAGQMEASRSLGL 636
```

```
Query: 121 TYLQVFIIIIFPQAFRKMLPAIISQFVTVIKDTSLLYSVIAIQEIFGKSQILMGRYFEAG 180
           +Y      +I PQA + MLP  I+QFV  +KDT+++ S I + E+F   +I++ R +
Sbjct: 637 SYGTTMRKVILPQAVKLMLPNFINQFVISLKDTTIV-SAIGLVELFQTGKIIIARNY--- 692

Query: 181 QVFTLYAIITAVYFITNFIISSFSRKLSKR 210
           Q F +YAI+  +Y I   +++  +++L KR
Sbjct: 693 QSFRMYAILAIIYLIMIILLTRLAKRLEKR 722
```

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

Example 368

A DNA sequence (GBSx0399) was identified in *S.agalactiae* <SEQ ID 1199> which encodes the amino acid sequence <SEQ ID 1200>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood =-12.21   Transmembrane     7 -  23 (   1 -  30)

----- Final Results -----
                bacterial membrane --- Certainty=0.5883(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:BAB04094 GB:AP001508 unknown conserved protein in B. subtilis
           [Bacillus halodurans]
 Identities = 43/157 (27%), Positives = 83/157 (52%), Gaps = 9/157 (5%)

Query: 26  YQSQFQKTTNQALAIAYKDAKVAKK--DVIHQKIDKEFENFRGSYEIEFNTKSAEYSYHV 83
           +Q++      N+ L +A        ++  + +  + +K+ +N R  YEIE       EY + +
Sbjct: 38  HQAESVSADNEGLTLAEASDIALERAGNGVVTEAEKDRDNGRVVYEIEVKNDDDEYDFKI 97

Query: 84  DVKTGQILERDMDNNGFSKSTSQSSSSSSQKSHKISQEEAKKIAFKDANIEESEVSNLKI 143
           D +TG+IL+    +     SK      SSS    ++  IS +EAK+IA K+ +     ++ ++++
Sbjct: 98  DQQTGEILKEKQEQRKGSKPREGHSSSKGSEA-VISMDEAKEIALKEVS---GKIDDIEL 153

Query: 144 KEEIENGKSVYDIDF-VDLKNKNEVDYQIDAETGKII 179
             E ENG  VY+++   D  + ++V   +DA TG ++
Sbjct: 154 --ERENGSLVYEVEIESDHYDDDDVTVYVDAMTGNVL 188
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1201> which encodes the amino acid sequence <SEQ ID 1202>. Analysis of this protein sequence reveals the following:

```
     Possible site: 57

>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood = -5.15   Transmembrane    42 -  58 (  41 -  60)

----- Final Results -----
                bacterial membrane --- Certainty=0.3060(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
Identities = 37/96 (38%), Positives = 63/96 (65%), Gaps = 5/96 (5%)

Query: 94  DMDNNGFSKSTSQSSSSSSQKSHKISQEEAKKIAFKDANIEESEVSNLKIKEEIENGKSV 153
           DMD+       +Q  +S + K  K+S+++AK IA KDA++ E++    L + ++ E+GK+V
Sbjct: 59  DMDDKD-DHMDNQPKTSQTSKKVKLSEDKAKSIALKDASVTEADAQMLSVTQDNEDGKAV 117
```

```
Query: 154 YDIDFVDLKNKN-EVDYQIDAETGKIIERSRDHMND 188
           Y+I+F   +NK+ E  Y IDA +G I+E+S + +ND
Sbjct: 118 YEIEF---QNKDQEYSYTIDANSGDIVEKSSEPIND 150
 Identities = 23/62 (37%), Positives = 37/62 (59%)

Query: 35  NQALAIAYKDAKVAKKDVIHQKIDKEFENFRGSYEIEFNTKSAEYSYHVDVKTGQILERD 94
           ++A +IA KDA V + D     + ++ E+ +  YEIEF  K  EYSY +D  +G I+E+
Sbjct: 85  DKAKSIALKDASVTEADAQMLSVTQDNEDGKAVYEIEFQNKDQEYSYTIDANSGDIVEKS 144

Query: 95  MD 96
            +
Sbjct: 145 SE 146
```

A related GBS gene <SEQ ID 8563> and protein <SEQ ID 8564> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:     14.45
GvH: Signal Score (-7.5): -5.92
     Possible site: 39
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -8.92 threshold:  0.0
   INTEGRAL    Likelihood = -8.92   Transmembrane    7 -  23 (   2 -  28)
   PERIPHERAL  Likelihood = 10.93      37
 modified ALOM score:   2.28

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4567(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:

```
26.1/59.2% over 140aa
  Bacillus subtilis
  EGAD|107494| hypothetical protein Insert characterized
  GP|2632048|emb|CAA05607.1||AJ002571 YkoJ Insert characterized
  GP|2633682|emb|CAB13185.1||Z99110 similar to hypothetical proteins from B. subtilis
Insert characterized
  PIR|F69859|F69859 conserved hypothetical protein ykoJ - Insert characterized

ORF00925(379 - 852 of 1164)
EGAD|107494|BS1329(29 - 169 of 170) hypothetical protein (Bacillus subtilis)
GP|2632048|emb|CAA05607.1||AJ002571 YkoJ (Bacillus subtilis)
GP|2633682|emb|CAB13185.1||Z99110 similar to hypothetical proteins from B. subtilis
(Bacillus subtilis) PIR|F69859|F69859 conserved hypothetical protein ykoJ - Bacillus
subtilis
%Match = 6.2
%Identity = 26.1  %Similarity = 59.2
Matches = 37  Mismatches = 52  Conservative Sub.s = 47

297       327       357       387       417       447       468       498
NIIE**KEGCCMIKKNKVFLEVLLVLVVILEGGVLFYQSQFQKTTNQALAIAYKDAKVAKKDVIH---QKIDKEFENFRG
           |   :|  |                :: :: :::  |:  ::|:   | |:    : ||::: | :
         MLKKKWMVGLLAGCLAAGGFSYNAFATENNENRQASSKTDALTEQEAEAIAKTVVDGTVEDIDRDLYNGKE
                  10        20        30        40        50        60        70

528       558       588       618       648       672       702       732
SYEIEFNTKSAEYSYHVDVKTGQILERDMDNNGFSKSTSQSSSSSSQKSHK--ISQEEAKKIAFKDANIEESEVSNLKIK
 ||:|   :  :|  :||: | | |   :                  :|: :  |::|||::||:|        |:  |:
VYEVEIEKEGEDYDVYVDIHTKQALNDPL----------------KEKAEQVAITKEEAEEIALKQTG---GTVTESKLD
                  90        100                       110       120       130
```

```
762       792       822       852       882       912       942       972
EEIENGKSVYDIDFVDLKNKNEVDYQIDAETGKIIERSRDHMND*FK*DIKKRRSKRPSF*LLSSLLPTF*KFT*KT*DD
|:  :|  :|::: :  |   | :::| |: |:||::  |
ED--DGAYIYEME-IQTKQGTETEFEISAKDGRIIKQEIDD
       140       150       160       170
```

SEQ ID 8564 (GBS37) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 4; MW 22kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 10; MW 47kDa).

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

Example 369

A DNA sequence (GBSx0400) was identified in *S.agalactiae* <SEQ ID 1203> which encodes the amino acid sequence <SEQ ID 1204>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1499(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 9709> which encodes amino acid sequence <SEQ ID 9710> 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 1205> which encodes the amino acid sequence <SEQ ID 1206>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2808(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 = 128/297 (43%), Positives = 180/297 (60%), Gaps = 9/297 (3%)

Query: 54  IDDIKVGSPIFKYFWT-SLSLQAPLKALEFVLEQAKMPTELSGELSETQYLVAQFSDELA 112
           I D ++GSP F   W      Q+  + L F+L+  +MP ELSG+L ETQ L+ +F   L
Sbjct: 46  IIDNRLGSPTFWVIWPIEKENQSAKQLLTFLLDLVEMPFELSGQLHETQTLLTRFHPSLL 105

Query: 113 PHDDFWIALSQVIYDSFPGNSLAEDTVLNRKLHQFRYLISSQQAQYVRRYFKDVGMTDRD 172
           P   FW  L+ ++  +FPG +L++   L ++LHQFRY+ISSQQAQ +R ++K + MTD
Sbjct: 106 PDHMFWKELASLVDQAFPGKTLSQAGELEKRLHQFRYVISSQQAQSIRNHYKMIEMTDAQ 165

Query: 173 ALVNYL-----SCL-REPDSIAYYESARLHNKRRRNGEIFGFPDDEPVINSKLLISFHTE 226
           AL  +L      CL R+       +SARLHNK R       FP  E   N K+L+ FHTE
Sbjct: 166 ALALFLRSKKGPCLWRQAPDYTLMDSARLHNKLRFEDNKVIFPSQEVSYNIKVLLWFHTE 225

Query: 227 FIIDDKGNFLNEIDAEVITRNGIINGASFNYAFKNNTRHKELDVDPVK-LDPKFRNDMTR 285
           F +D  G FLNE+DAEV+T  GI+NGASFNY    +  RH +LDVDP+   DP+FR D  +
Sbjct: 226 FTLDSTGFFLNEVDAEVVTEKGIVNGASFNYG-TDGPRHWDLDVDPISHHDPQFRRDTLK 284
```

```
Query: 286 GYRSPNLSRRKWFFFKEEDYDCSYFNKKGYYAFGRRSAKQSVDKQVKYLKKAVQKMR 342
           G+RSP     R+WF  +++D+  SYFN KG +A+  +S+   V K  K  K+ +  ++
Sbjct: 285 GFRSPKRVFRQWFRAQKDDFMFSYFNAKGLFAYHNKSSFARVKKSAKQFKRQIHPIK 341
```

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

Example 370

A DNA sequence (GBSx0401) was identified in *S.agalactiae* <SEQ ID 1207> which encodes the amino acid sequence <SEQ ID 1208>. This protein is predicted to be similar to two-component response regulator [YcbM] (ompr-likeprotei). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3129(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:CAA55264 GB:X78502 gtcR [Brevibacillus brevis]
 Identities = 99/228 (43%), Positives = 149/228 (64%), Gaps = 3/228 (1%)

Query: 2   RTVLVVQGDDETIELLRSYLEGALYKVVMASDGEEAFSLFQQHQIDLAIIDITLPKIDGY 61
           +T+L+    + E IELL+ +LE   Y+++ A DGE+A++  +QH +DLAIIDI +P +DG+
Sbjct: 3   KTILIADDEPEIIELLKLFLERESYRIIEAYDGEQAWNYIRQHPVDLAIIDIMMPALDGF 62

Query: 62  ELTRLIRQDSQIPIIMLAAKTTDMDRILGLNIGADDFITKPFNSLEVLARINSQLRRYYE 121
           +L + +  + ++P+I+L+AK  D D+ILGL +GADDFI+KPFN LE +ARI +QLRR +E
Sbjct: 63  QLIKRLTNEYKLPVIILSAKNRDSDKILGLGLGADDFISKPFNPLEAVARIQAQLRRAFE 122

Query: 122 FNSLAKP--KNQFIKIGELELDEEHVELTKNGKHIKLTATEFKILHILMS-SPGRIYTKT 178
           FN   +     Q   +G L L       + +  +   +T  E+++L+  M  S   I+TK
Sbjct: 123 FNEPEEKAISTQSTTVGRLTLLHTACVVYRGDETYSVTPLEYRLLNTFMQCSRTSIFTKQ 182

Query: 179 QLYEKINGRYLEGDETTIMVHISNIRDKIEDDSKYPKYIKTLRGVGYK 226
           QL+E+         D+ TIMV IS +RDKIED  + P YIKT+RG+GYK
Sbjct: 183 QLFEQAWSETYWEDDNTIMVQISRLRDKIEDQPRQPVYIKTVRGLGYK 230
```

There is also homology to SEQ ID 1182:

```
 Identities = 87/230 (37%), Positives = 144/230 (61%), Gaps = 5/230 (2%)

Query: 1   MRTVLVVQGDDETIELLRSYLEGALYKVVMASDGEEAFSLFQQHQIDLAIIDITLPKIDG 60
           M+ +L+V  +    ++++  L    Y +V A DG EA ++F++ + DL I+D+ LP++DG
Sbjct: 1   MKKILIVDDEKPISDIIKFNLTKEGYDIVTAFDGREAVTIFEEEKPDLIILDLMLPELDG 60

Query: 61  YELTRLIRQDSQIPIIMLAAKTTDMDRILGLNIGADDFITKPFNSLEVLARINSQLRRYY 120
            E+ + IR+ S +PIIML+AK ++ D+++GL IGADD++TKPF++ E+LAR+ + LRR
Sbjct: 61  LEVAKEIRKTSHVPIIMLSAKDSEFDKVIGLEIGADDYVTKPFSNRELLARVKAHLRRTE 120

Query: 121 EFNSLAKPKN-----QFIKIGELELDEEHVELTKNGKHIKLTATEFKILHILMSSPGRIY 175
              +    +N     Q + IG L++  +      K+G+ ++LT  EF++LH L +  G++
Sbjct: 121 TIETAVAEENASSGTQELTIGNLQILPDAFVAKKHGQEVELTHREFELLHHLANHMGQVM 180

Query: 176 TKTQLYEKINGRYLEGDETTIMVHISNIRDKIEDDSKYPKYIKTLRGVGY 225
           T+  L E + G    GD  T+ V +  +R+KIED     P+YI T RGVGY
Sbjct: 181 TREHLLEIVWGYDYFGDVRTVDVTVRRLREKIEDTPSRPEYILTRRGVGY 230
```

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

Example 371

A DNA sequence (GBSx0402) was identified in *S.agalactiae* <SEQ ID 1209> which encodes the amino acid sequence <SEQ ID 1210>. This protein is predicted to be threonyl-tRNA synthetase 1 (thrS). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2353(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:BAB06860 GB:AP001517 threonyl-tRNA synthetase 1 [Bacillus halodurans]
 Identities = 413/638 (64%), Positives = 506/638 (78%), Gaps = 7/638 (1%)

Query: 1   MIKITFPDGAIREFESGITTFEIAQSISNSLAKKALAGKFNGQLIDTTRAIEEDGSIEIV 60
           MI ITFPDGA++EF  G TT EIA SIS  L KKALAG  +G L+D    IE+DG+I IV
Sbjct: 4   MINITFPDGAVKEFPKGTTTAEIAGSISPGLKKKALAGMLDGTLLDLNTPIEQDGTITIV 63

Query: 61  TPDHEDALGVLRHSAAHLFAQAAKRLFPD--LCLGVGPAIQDGFYYDTDNKSGQISNDDL 118
           TP+ ++AL VLRHS AH+ AQA KRLF D  + LGVGP I+ GFYYD D      ++ +DL
Sbjct: 64  TPESDEALEVLRHSTAHVMAQALKRLFKDRNVKLGVGPVIEGGFYYDVDMDES-LTPEDL 122

Query: 119 PRIEEEMKKIVKENHPCIREEISKEEALELFKD--DPYKVELISEHAEDG-LTVYRQGEF 175
           P+IE+EMKKI+ EN P  R  +S+EEAL  +++  DPYK+ELI++  ED  +T+Y QGEF
Sbjct: 123 PKIEKEMKKIIGENLPIERVVVSREEALARYEEVGDPYKIELINDLPEDETITIYEQGEF 182

Query: 176 VDLCRGPHVPSTGRIQVFHLLNVAGAYWRGNSDNAMMQRVYGTAWFDKKDLKAYLKRREE 235
            DLCRG HVPSTG+++ F LLN+AGAYWRG+S N M+QR+YGTA+F K DL  +L+  EE
Sbjct: 183 FDLCRGVHVPSTGKLKEFKLLNLAGAYWRGDSSNKMLQRIYGTAFFKKADLDEHLRLLEE 242

Query: 236 AKERDHRKLGKELDLFMVNPEVGQGLPFWLPNGATIRRELERYIVDKEIASGYQHVYTPP 295
           AKERDHRKLGKEL +F ++ +VGQGLP WLP GATIRR +ERYIVDKE   GYQHVYTP
Sbjct: 243 AKERDHRKLGKELGIFALSQKVGQGLPLWLPKGATIRRIIERYIVDKEEKLGYQHVYTPV 302

Query: 296 MASVEFYKTSGHWDHYREDMFPTMDMGDGEEFVLRPMNCPHHIEVYKHHVHSYRELPIRI 355
           +AS E YKTSGHWDHY++DMFPTM+M + EE VLRPMNCPHH+ VYK  + SYR+LP+RI
Sbjct: 303 LASSELYKTSGHWDHYKDDMFPTMEM-ENEELVLRPMNCPHHMMVYKTEMRSYRQLPLRI 361

Query: 356 AELGMMHRYEKSGALTGLQRVREMTLNDAHIFVTPEQIKDEFLKALNLIAEIYEDFNLTD 415
           AELG+MHRYE SGA++GLQRVR MTLNDAHIF  P+QIKDEF++ + LI  +YEDF L +
Sbjct: 362 AELGLMHRYEMSGAVSGLQRVRGMTLNDAHIFCRPDQIKDEFVRVVRLIQAVYEDFGLKN 421

Query: 416 YRFRLSYRDPEDKHKYYDNDEMWENAQAMLKEAMDDFGLDYFEAEGEAAFYGPKLDIQVK 475
           Y FRLSYRDPEDK KY+D+D MW  AQ MLKEAMD+  L+YFEAEGEAAFYGPKLD+QV+
Sbjct: 422 YSFRLSYRDPEDKEKYFDDDNMWNKAQGMLKEAMDELELEYFEAEGEAAFYGPKLDVQVR 481

Query: 476 TALGNEETLSTIQLDFLLPERFDLKYIGADGEEHRPIMIHRGGISTMERFTAILIETYKG 535
           TALG +ETLST+QLDFLLPERFDL Y+G DG+ HRP+++HRG +STMERF A L+E YKG
Sbjct: 482 TALGKDETLSTVQLDFLLPERFDLTYVGEDGQPHRPVVVHRGVVSTMERFVAFLLEEYKG 541

Query: 536 AFPTWLAPQQVSVIPISNEAHIDYAWEVARVLKDRGIRAEVDDRNEKMQYKIRAAQTQKI 595
           AFPTWLAP QV VIP+S EAH++YA  V   L+  GIR E+D+R+EK+ YKIR AQ QKI
Sbjct: 542 AFPTWLAPVQVQVIPVSPEAHLEYAKNVQETLQQAGIRVEIDERDEKIGYKIREAQMQKI 601

Query: 596 PYQLIVGDKEMEEKAVNVRRYGSKATETKSIEEFVESI 633
           PY L++GDKE+E    VNVR+YG K + +  ++EFV  +
Sbjct: 602 PYMLVLGDKEVEANGVNVRKYGEKDSSSMGLDEFVRHV 639
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1211> which encodes the amino acid sequence <SEQ ID 1212>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2566(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 = 564/644 (87%), Positives = 608/644 (93%)

Query: 1   MIKITFPDGAIREFESGITTFEIAQSISNSLAKKALAGKFNGQLIDTTRAIEEDGSIEIV 60
           MIKITFPDGA+REFESG+TTF+IA+SIS SLAKKALAGKFN QLIDTTRAIEEDGSIEIV
Sbjct: 1   MIKITFPDGAVREFESGVTTFDIAESISKSLAKKALAGKFNDQLIDTTRAIEEDGSIEIV 60

Query: 61  TPDHEDALGVLRHSAAHLFAQAAKRLFPDLCLGVGPAIQDGFYYDTDNKSGQISNDDLPR 120
           TPDH+DA  VLRHSAAHLFAQAAKRLFP+L LGVGPAI +GFYYDTDN  GQISN+DLPR
Sbjct: 61  TPDHKDAYEVLRHSAAHLFAQAAKRLFPNLHLGVGPAIAEGFYYDTDNAEGQISNEDLPR 120

Query: 121 IEEEMKKIVKENHPCIREEISKEEALELFKDDPYKVELISEHAEDGLTVYRQGEFVDLCR 180
           IE EM+KIV EN+PCIREE++KEEALELFKDDPYKVELI+EHA  GLTVYRQGEFVDLCR
Sbjct: 121 IEAEMQKIVTENYPCIREEVTKEEALELFKDDPYKVELINEHAGAGLTVYRQGEFVDLCR 180

Query: 181 GPHVPSTGRIQVFHLLNVAGAYWRGNSDNAMMQRVYGTAWFDKKDLKAYLKRREEAKERD 240
           GPHVPSTGRIQVFHLLNVAGAYWRGNSDN MMQR+YGTAWFDKKDLKAYL R EEAKERD
Sbjct: 181 GPHVPSTGRIQVFHLLNVAGAYWRGNSDNNMMQRIYGTAWFDKKDLKAYLTRLEEAKERD 240

Query: 241 HRKLGKELDLFMVNPEVGQGLPFWLPNGATIRRELERYIVDKEIASGYQHVYTPPMASVE 300
           HRKLGKELDLFM++ EVGQGLPFWLP+GATIRR LERYI DKE+ASGYQHVYTPP+ASVE
Sbjct: 241 HRKLGKELDLFMISQEVGQGLPFWLPDGATIRRTLERYITDKELASGYQHVYTPPLASVE 300

Query: 301 FYKTSGHWDHYREDMFPTMDMGDGEEFVLRPMNCPHHIEVYKHHVHSYRELPIRIAELGM 360
            YKTSGHWDHY+EDMFP MDMGDGEEFVLRPMNCPHHI+VYK+HV SYRELPIRIAELGM
Sbjct: 301 LYKTSGHWDHYQEDMFPVMDMGDGEEFVLRPMNCPHHIQVYKNHVRSYRELPIRIAELGM 360

Query: 361 MHRYEKSGALTGLQRVREMTLNDAHIFVTPEQIKDEFLKALNLIAEIYEDFNLTDYRFRL 420
           MHRYEKSGAL+GLQRVREMTLND HIFVTPEQI++EF +AL LI ++Y DFNLTDYRFRL
Sbjct: 361 MHRYEKSGALSGLQRVREMTLNDGHIFVTPEQIQEEFQRALQLIIDVYADFNLTDYRFRL 420

Query: 421 SYRDPEDKHKYYDNDEMWENAQAMLKEAMDDFGLDYFEAEGEAAFYGPKLDIQVKTALGN 480
           SYRDP D HKYYDNDEMWENAQ+MLK A+D+ G+DYFEAEGEAAFYGPKLDIQVKTALGN
Sbjct: 421 SYRDPNDTHKYYDNDEMWENAQSMLKAALDEMGVDYFEAEGEAAFYGPKLDIQVKTALGN 480

Query: 481 EETLSTIQLDFLLPERFDLKYIGADGEEHRPIMIHRGGISTMERFTAILIETYKGAFPTW 540
           EETLSTIQLDFLLPERFDLKYIGADGEEHRP+MIHRG ISTMERFTAILIETYKGAFPTW
Sbjct: 481 EETLSTIQLDFLLPERFDLKYIGADGEEHRPVMIHRGVISTMERFTAILIETYKGAFPTW 540

Query: 541 LAPQQVSVIPISNEAHIDYAWEVARVLKDRGIRAEVDDRNEKMQYKIRAAQTQKIPYQLI 600
           LAP QV+VIPISNEAHIDYAWEVA+ L+DRG+RA+VDDRNEKMQYKIRA+QT KIPYQLI
Sbjct: 541 LAPHQVTVIPISNEAHIDYAWEVAKTLRDRGVRADVDDRNEKMQYKIRASQTSKIPYQLI 600

Query: 601 VGDKEMEEKAVNVRRYGSKATETKSIEEFVESILADIARKSRPD 644
           VGDKEME+K+VNVRRYGSK T T+S+EEFVE+ILADIARKSRPD
Sbjct: 601 VGDKEMEDKSVNVRRYGSKTTHTESVEEFVENILADIARKSRPD 644
```

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

Example 372

A DNA sequence (GBSx0403) was identified in *S.agalactiae* <SEQ ID 1213> which encodes the amino acid sequence <SEQ ID 1214>. Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1985(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:CAA72250 GB:Y11463 ORF5 [Streptococcus pneumoniae]
 Identities = 189/290 (65%), Positives = 234/290 (80%)

Query: 1   MRIGLFTDTYFPQVSGVSTSIRTLKEGLEKEGHEVYIFTTTDRNVKRFEDPTIIRLPSVP 60
           MRIGLFTDTYFPQVSGV+TSIRTLK  LEK+GH V+IFTTTD++V R+ED  IIR+PSVP
Sbjct: 1   MRIGLFTDTYFPQVSGVATSIRTLKTELEKQGHAVFIFTTTDKDVNRYEDWQIIRIPSVP 60

Query: 61  FISFTDRRVVYRGLISAYRIAKDYELDIIHTQTEFSLGLLGKLVAKALRIPVVHTYHTQY 120
           F +F DRR  YRG   A  IAK Y+LDIIHTQTEFSLGLLG  +A+ L+IPV+HTYHTQY
Sbjct: 61  FFAFKDRRFAYRGFSKALEIAKQYQLDIIHTQTEFSLGLLGIWIARELKIPVIHTYHTQY 120

Query: 121 EDYVGYIAKGKLIKPSMVKYIMRTYLSDLDGVICPSRIVLNLLDGYGVKIPKQVIPTGIP 180
           EDYV YIAKG LI+PSMVKY++R +L D+DGVICPS IV +LL  Y VK+ K+VIPTGI
Sbjct: 121 EDYVHYIAKGMLIRPSMVKYLVRGFLHDVDGVICPSEIVRDLLSDYKVKVEKRVIPTGIE 180

Query: 181 VENYRREDISEETIKNLRTELGLADNDTMLLSLSRVSFEKNIQAILMHLSAVVDENPHVK 240
           +  + R +I +E +K LR++LG+ D +  LLSLSR+S+EKNIQA+L+  + V+ E   VK
Sbjct: 181 LAKFERPEIKQENLKELRSKLGIQDGEKTLLSLSRISYEKNIQAVLVAFADVLKEEDKVK 240

Query: 241 LVIVGDGPYLSDLKELVHSLELENSVIFTGMVEHSQVAIYYKACDFFISA 290
           LV+ GDGPYL+DLKE   +LE+++SVIFTGM+  S+ A+YYKA DFFISA
Sbjct: 241 LVVAGDGPYLNDLKEQAQNLEIQDSVIFTGMIAPSETALYYKAADFFISA 290
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1215> which encodes the amino acid sequence <SEQ ID 1216>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1074(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 = 309/444 (69%), Positives = 370/444 (82%)

Query: 1   MRIGLFTDTYFPQVSGVSTSIRTLKEGLEKEGHEVYIFTTTDRNVKRFEDPTIIRLPSVP 60
           MRIGLFTDTYFPQVSGV+TSIRTLKE LEKEGHEVYIFTTTDR+VKRFEDPTIIRLPSVP
Sbjct: 1   MRIGLFTDTYFPQVSGVATSIRTLKEELEKEGHEVYIFTTTDRDVKRFEDPTIIRLPSVP 60

Query: 61  FISFTDRRVVYRGLISAYRIAKDYELDIIHTQTEFSLGLLGKLVAKALRIPVVHTYHTQY 120
           F+SFTDRRVVYRGLIS+Y+IAK Y LDIIHTQTEFSLGLLGK++ KALRIPVVHTYHTQY
Sbjct: 61  FVSFTDRRVVYRGLISSYKIAKHYNLDIIHTQTEFSLGLLGKMIGKALRIPVVHTYHTQY 120

Query: 121 EDYVGYIAKGKLIKPSMVKYIMRTYLSDLDGVICPSRIVLNLLDGYGVKIPKQVIPTGIP 180
           EDYV YIA GK+I+PSMVK ++R YL DLDGVICPSRIVLNLL+GY V IPK+VIPTGIP
Sbjct: 121 EDYVSYIANGKIIRPSMVKPLLRGYLKDLDGVICPSRIVLNLLEGYEVTIPKRVIPTGIP 180

Query: 181 VENYRREDISEETIKNLRTELGLADNDTMLLSLSRVSFEKNIQAILMHLSAVVDENPHVK 240
           +E Y R+DI+ E + NL+ ELG+A ++TMLLSLSR+S+EKNIQAI+  + A++ EN  +K
Sbjct: 181 LEKYIRDDITAEEVTNLKAELGIAGDETMLLSLSRISYEKNIQAIINQMPAILAENAKIK 240

Query: 241 LVIVGDGPYLSDLKELVHSLELENSVIFTGMVEHSQVAIYYKACDFFISASTSETQGLTY 300
           L+IVG+GPYL DLK L   LE++  V FTGMV H +VA+YYKACDFFISASTSETQGLTY
Sbjct: 241 LIIVGNGPYLQDLKHLAMQLEVDKHVTFTGMVPHDKVALYYKACDFFISASTSETQGLTY 300

Query: 301 IESLASGRPIIAQSNPYLDDVISDKMFGTLYKKESDLADAILDAIAETPKMTQEAYEQKL 360
```

```
            IESLASG PIIA  NPYLDDV++DKMFGTLY  E+DL DAI+DAI +TP M +    +K
Sbjct: 301 IESLASGTPIIAHGNPYLDDVVTDKMFGTLYYAETDLTDAIIDAILKTPVMDKRLLAKKR 360

Query: 361 YEISAENFSKSVYAFYLDFLISQKASVKEKVSLTIGNKDSHSTLRFVRKAVYLPKKVFTF 420
           YEISA++F KS+Y FYLD LI++ +   +K+SL + +    S+L+ V+ A++LPK+
Sbjct: 361 YEISAQHFGKSIYTFYLDTLIARNSKEAQKLSLYLNHSGKSSSLKLVQGAIHLPKRAAKV 420

Query: 421 TGRASKKVVKAPKRRISSIRDFLD 444
           T   S KVVKAP + + +I+DFLD
Sbjct: 421 TAITSVKVVKAPIKLVHAIKDFLD 444
```

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

Example 373

A DNA sequence (GBSx0404) was identified in *S.agalactiae* <SEQ ID 1217> which encodes the amino acid sequence <SEQ ID 1218>. 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 = 117/350 (33%), Positives = 178/350 (50%), Gaps = 29/350 (8%)

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 FYQKADAIITPTDYSKQLIKAYGIKKPIFVLSNGIDLSRYQRSEKKESAFRHYFHLSKDD 161
           FY +AD ++ P++Y+K +++AY +  PI  LSNG+DL   Q  E   +  R  F L  D
Sbjct: 121 FYSQADLVLCPSEYTKDVLRAYPVDAPIRQLSNGVDLESMQGYESFRADTRARFDL--DG 178

Query: 162 KVVMGAGLYFMRKGIDQFVEVAAKMPDIRFIWFGETNKWVIPRKVRQIVTKQHPSNVTFA 221
            VV   G  F RKG+   F E+ AK   D   F WFG  ++          +      P+NVTF
Sbjct: 179 TVVYAVGEVFERKGLTMFCEL-AKATDHEFAWFGPYDEGPQAGAATRKWVADPPANVTFT 237

Query: 222 GYIKGDVYEGAMSASDAFFFPSREETEGIVVLEALASHQHVVLRDIPVYHGWVTE-DSVE 280
           GY++      A  A D + FP++ E +GI VLEA+A  + VVLRDIPV+  + T+ +
Sbjct: 238 GYMEDK--RAAFGAGDIYLFPAKVENQGIAVLEAMACGKPVVLRDIPVFREFFTDGEDCL 295

Query: 281 LATDVDGFVEKLDKVLSGKSDKIKEGYH---VAESRSIERIAHELASVYQ 327
           + +  + F + +D++      + + G +    AES S++RI  ELAS+Y+
Sbjct: 296 MCSTFEAFRDAIDRLADDPELRTRLGENARETAESHSLDRIGEELASIYE 345
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1219> which encodes the amino acid sequence <SEQ ID 1220>. Analysis of this protein sequence reveals the following:

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

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

```
                bacterial cytoplasm --- Certainty=0.4088(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 = 236/332 (71%), Positives = 276/332 (83%)

Query: 1   MKVLLYLEAEEYLKKSGIGRAIKHQEKALQIAGIDYTTNPTDDFDLVHMNTYGIRSWLLM 60
           MKVLLYLEAE YL+KSGIGRAIKHQ KAL + G  +TTNP + +DLVH+NTYG++SWLLM
Sbjct: 1   MKVLLYLEAENYLRKSGIGRAIKHQAKALSLVGQHFTTNPRETYDLVHLNTYGLKSWLLM 60

Query: 61  SKAKKTGKKVIMHGHSTEEDFRNSFIGSNLVSPLFKWYLCRFYQKADAIITPTDYSKQLI 120
            KA+K GKKVIMHGHSTEEDFRNSFI SNL+SP FK YLC FY KADAIITPT YSK LI
Sbjct: 61  IKAQKAGKKVIMHGHSTEEDFRNSFIFSNLLSPWFKKYLCHFYNKADAIITPTLYSKSLI 120

Query: 121 KAYGIKKPIFVLSNGIDLSRYQRSEKKESAFRHYFHLSKDDKVVMGAGLYFMRKGIDQFV 180
           ++YG+K PIF +SNGIDL +Y    KKE+AFR YF + + +KVVMGAGL+F+RKGID FV
Sbjct: 121 ESYGVKSPIFAVSNGIDLEQYGADPKKEAAFRRYFDIKEGEKVVMGAGLFFLRKGIDDFV 180

Query: 181 EVAAKMPDIRFIWFGETNKWVIPRKVRQIVTKQHPSNVTFAGYIKGDVYEGAMSASDAFF 240
           +VA  MPD+RFIWFGETNKWVIP +VRQ+V   HP N+ F GYIKGDVYEGAM+ +DAFF
Sbjct: 181 KVAQAMPDVRFIWFGETNKWVIPAQVRQMVNGNHPKNLIFPGYIKGDVYEGAMTGADAFF 240

Query: 241 FPSREETEGIVVLEALASHQHVVLRDIPVYHGWVTEDSVELATDVDGFVEKLDKVLSGKS 300
           FPSREETEGIVVLEALAS QH+VLRDIPVY+GWV + S ELATD+ GF+E L KV SG S
Sbjct: 241 FPSREETEGIVVLEALASRQHLVLRDIPVYYGWVDQSSAELATDIPGFIEALKKVFSGAS 300

Query: 301 DKIKEGYHVAESRSIERIAHELASVYQKVMEL 332
           +K++ GY VA+SR +E + H L  VY+KVMEL
Sbjct: 301 NKVEAGYKVAQSRRLETVGHALVDVYKKVMEL 332
```

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

Example 374

A DNA sequence (GBSx0405) was identified in *S.agalactiae* <SEQ ID 1221> which encodes the amino acid sequence <SEQ ID 1222>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.5487(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 = 308/483 (63%), Positives = 378/483 (77%)

Query: 1   MTNELIMQAFEWYLPSDGNHWKKLEESISDLKKLGISKIWLPPAFKGTSSDDVGYGVYDL 60
           MTNE +MQ FEWYLP+DG HW+ L E  S LK +GISK+W+PPAFKGT S+DVGYGVYDL
Sbjct: 1   MTNETMMQYFEWYLPNDGKHWQHLAEDASHLKNIGISKVWMPPAFKGTGSNDVGYGVYDL 60

Query: 61  FDLGEFDQNGTIRTKYGRKEEYLKLIKSLKANGIKPFADIVLNHKANGDHKEKFQVIKVN 120
           +DLGEF+QNGT+RTKYG +E+YL  + +LK   I P +DIVLNHKANGD KE+FQV+KVN
Sbjct: 61  YDLGEFNQNGTVRTKYGSREDYLNAVNALKEQEIMPISDIVLNHKANGDAKERFQVVKVN 120

Query: 121 PENRQEALSEPYEIEGWTGFDFPGRQGEYNDFKWHWYHFTGLDYDAKNNETDIFMIVGDN 180
           P NRQE +SEPYEIEGWT F+FPGRQ  Y+DFKWHWYHFTG+DYDA +NE  I+MI+GDN
Sbjct: 121 PSNRQEKISEPYEIEGWTQFNFPGRQDNYSDFKWHWYHFTGVDYDALHNENGIYMILGDN 180

Query: 181 KGWADDDLIDDENGNFDYLMYNDIDFKHPEVIKNLQDWAKWFIETTGIEGFRLDAVKHID 240
```

```
            KGWA  + ID ENGN+DYLMY+DIDFKHPEV ++L+DW  WF+ET+G+ GFRLDA+KHID
Sbjct: 181 KGWASQENIDQENGNYDYLMYDDIDFKHPEVQEHLRDWVAWFLETSGVGGFRLDAIKHID 240

Query: 241 SYFIQTFINDIRTKIKPDLEVFGEYWKSDQTSMKDYLEATQFQFSLVDVTLHMNFFDASH 300
             F+  FI  IR  +K DL VFGEYWK     + DYL +   QF L+DV LHM+ F+A
Sbjct: 241 KTFMAQFIRYIREHLKADLYVFGEYWKDSHFDITDYLHSVDLQFDLIDVMLHMSLFEAGQ 300

Query: 301 QNRDFDMRTIFDDSLVIDNPEYAVTFVENHDTQSGQALESRVEDWFKPLAYGLILLRQQG 360
            +  DFD+ TI DDSL+  +P++AVTFV+NHD+Q GQALES V +WFKPLAYGLILLRQ+G
Sbjct: 301 KGSDFDLSTILDDSLMKSHPDFAVTFVDNHDSQRGQALESTVAEWFKPLAYGLILLRQEG 360

Query: 361 TPCLFYGDYYGIQGEFGQPSFKEVIDKMAELRQNYVFGKQVDYFTHSNCIGWTCLGDEEH 420
             PC+FYGDYYGI GEF Q SF+ V+DK+  +RQ +V+G +    T  NCIGWTCLGDEEH
Sbjct: 361 IPCVFYGDYYGISGEFAQESFQTVLDKLLYIRQYHVYGSKKIILTMPNCIGWTCLGDEEH 420

Query: 421 NSCLAVVLTNGDQGWKHMEVGEIYAGKTFVDYLGNCEQEVVIGDDGWGDFLVESASISAW 480
              +AV+++NG+   K M +GE   K FVDYL NC +EV++ D GWGDF V+ AS+SAW
Sbjct: 421 PDGVAVIISNGEANCKRMNMGEFNRNKVFVDYLNNCTEEVILDDQGWGDFFVQEASLSAW 480

Query: 481 VPK 483
            V K
Sbjct: 481 VNK 483
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1223> which encodes the amino acid sequence <SEQ ID 1224>. Analysis of this protein sequence reveals the following:

```
      Possible site: 30

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

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

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

```
>GP:AAB00845 GB:M57692 alpha-cyclodextrin glycosyltransferase
          [Thermoanaerobacterium thermosulfurigenes]
 Identities = 356/710 (50%), Positives = 468/710 (65%), Gaps = 16/710 (2%)

Query: 7   KTYKLLTKSAVLLGLISFPLT--VSAADNASVTNKADFSTDTIYQIVTDRFNDGNTSNNG 64
           KT+KL+    + L L+ F LT  + AA + +V+N  ++STD IYQIVTDRF DGNTSNN
Sbjct: 3   KTFKLILVLMLSLTLV-FGLTAPIQAASDTAVSNVVNYSTDVIYQIVTDRFVDGNTSNNP 61

Query: 65  KTDVFDKN--DLKKYHGGDWQGIIAKIKDGYLTDMGISAIWISSPVENIDSIDPSN---G 119
             D++D     LKKY GGDWQGII KI DGYLT MG++AIWIS PVENI ++ P +   G
Sbjct: 62  TGDLYDPTHTSLKKYFGGDWQGIINKINDGYLTGMGVTAIWISQPVENIYAVLPDSTFGG 121

Query: 120 SAAYHGYWAKDFFKTNQHFGTEADFQQLVKVAHQHHIKVVIDFAPNHTSTAEKEGTTFKE 179
           S +YHGYWA+DF +TN +FG+  DFQ L+  AH H+IKV+IDFAPNHTS A +   T+ E
Sbjct: 122 STSYHGYWARDFKRTNPYFGSFTDFQNLINTAHAHNIKVIIDFAPNHTSPASETDPTYAE 181

Query: 180 DGALYKNGKLVGKFSDDKDKIFNHESWTDFSTYENSIYHSMYGLADLNNINPKVDQYMKE 239
           +G LY NG L+G +++D +  F+H   TDFS+YE+ IY +++ LADLN  N  +D Y+K
Sbjct: 182 NGRLYDNGTLLGGYTNDTNGYFHHYGGTDFSSYEDGIYRNLFDLADLNQQNSTIDSYLKS 241

Query: 240 AIDKWLDLGVDGIRVDAVKHMSQGWQKNWLSHIYEKHNVFVFGEWFSGHTDDDYDMTTFA 299
           AI  WLD+G+DGIR+DAVKHM  GWQKN++  I     VF FGEWF G  + D + T FA
Sbjct: 242 AIKVWLDMGIDGIRLDAVKHMPFGWQKNFMDSILSYRPVFTFGEWFLGTNEIDVNNTYFA 301

Query: 300 NNSGMGLLDFRFANAIRQLYTGFSTFTMRDFYKVLENRDQVTNEVTDQVTFIDNHDMERF 359
           N SGM LLDFRF+  +RQ++   +T TM      ++++      N + D VTFIDNHDM+RF
Sbjct: 302 NESGMSLLDFRFSQKVRQVFRD-NTDTMYGLDSMIQSTASDYNFINDMVTFIDNHDMDRF 360

Query: 360 ATKVANNQTAVNQAYALLLTSRGVPNIYYGTEQYATGDKDPNNRGDMPSFNKESQAYKVI 419
                +  V QA A  LTSRGVP IYYGTEQY TG+ DP NR  M SFN  + AY VI
```

```
Sbjct: 361 YN--GGSTRPVEQALAPTLTSRGVPAIYYGTEQYMTGNGDPYNRAMMTSFNTSTTAYNVI 418

Query: 420 SKLAPLRKQNQALAYGTTEQRWISDHVLVFERKFGNHVALVAINRDQTNGYTITNAKTAL 479
           KLAPLRK N A+AYGTT+QRWI++ V ++ERKFGN+VALVAINR+ +  Y IT   TAL
Sbjct: 419 KKLAPLRKSNPAIAYGTTQQRWINNDVYIYERKFGNNVALVAINRNLSTSYNITGLYTAL 478

Query: 480 PQNSYKDKLEGLLGGQELIVGADGTISSFELGAGQVAVWTYEGEDKTPQLGDVDASVGIA 539
           P  +Y D L GLL G  + V +DG+++ F L AG+VAVW Y     +P +G V  ++  A
Sbjct: 479 PAGTYTDVLGGLLNGNSISVASDGSVTPFTLSAGEVAVWQYVSSSNSPLIGHVGPTMTKA 538

Query: 540 GNKITISGQGFGNSKGQVTFGEISAEILSWSDTLITLKVPTVPANYYNISVTTADKQTSN 599
           G  ITI G+GFG + GQV FG  +  I+SW DT + +KVP+V    YNIS+ T+   TSN
Sbjct: 539 GQTITIDGRGFGTTSGQVLFGSTAGTIVSWDDTEVKVKVPSVTPGKYNISLKTSSGATSN 598

Query: 600 SYQAFEVLTDKQIPVRLLINDFKTVPGEQLYLMGDVFEMGANDAKNAVGPLFNNTQTIAK 659
           +Y    +LT  QI VR ++N+  TV GE +YL G+V E+G  D   A+GP+FN  Q + +
Sbjct: 599 TYNNINILTGNQICVRFVVNNASTVYGENVYLTGNVAELGNWDTSKAIGPMFN--QVVYQ 656

Query: 660 YPNWFFDTHLPINKEIAVKLVKKDSIGNVLWT--SPETYSIKTGHEAQTI 707
           YP W++D  +P     I  K +KK+     + W    S  TY++ +       I
Sbjct: 657 YPTWYYDVSVPAGTTIQFKFIKKNG-NTITWEGGSNHTYTVPSSSTGTVI 705
```

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

```
 Identities = 112/509 (22%), Positives = 193/509 (37%), Gaps = 103/509 (20%)

Query: 18  GNHWKKLEESISD--LKKLGISKIWLPPAFKGTSSDDV--------GYGVYDLFDLGEFD 67
           G  W+ +   I D  L  +GIS IW+     +   S D          GY   D F   +
Sbjct: 79  GGDWQGIIAKIKDGYLTDMGISAIWISSPVENIDSIDPSNGSAAYHGYWAKDFFKTNQH- 137

Query: 68  QNGTIRTKYGRKEEYLKLIKSLKANGIKPFADIVLNHKANGDHKEKFQVIKVNPENRQEA 127
                  +G + ++ +L+K     + IK    D   NH +  + +
Sbjct: 138 --------FGTEADFQQLVKVAHQHHIKVVIDFAPNHTSTAEKE---------------- 173

Query: 128 LSEPYEIEGWTGFDFPGRQQGEYNDFKWHWYHFTGLDYDAKNNETDIFMIVGDNKGWADDD 187
                    G  F       Y + K       G   D K+          + +++ W D
Sbjct: 174 -----------GTTFKEDGALYKNGK-----LVGKFSDDKDK-------IFNHESWTDFS 210

Query: 188 LIDDE--NGNFDYLMYNDIDFKHPEVIKNLQDWAKWFIETTGIEGFRLDAVKHIDSYFIQ 245
             ++    +  +     N+I+ K  + +K   D  KW      G++G R+DAVKH+    + +
Sbjct: 211 TYENSIYHSMYGLADLNNINPKVDQYMKEAID--KWL--DLGVDGIRVDAVKHMSQGWQK 266

Query: 246 TFINDIRTKIKPDLEVFGEYWKSDQTSMKDYLEATQFQFSLVDVTLHMNFFDASHQ-NRD 304
            +++ I     K ++ VFGE W S  T   D  + T F  +     L   F +A  Q
Sbjct: 267 NWLSHIYE--KHNVFVFGE-WFSGHTD--DDYDMTTFANNSGMGLLDFRFANAIRQLYTG 321

Query: 305 FDMRTIFDDSLVIDNPEYA-------VTFVENHDTQSGQALESRVEDWFKPLAYGLILLR 357
           F   T+ D   V++N +          VTF++NHD +      +  +       AY L LL
Sbjct: 322 FSTFTMRDFYKVLENRDQVTNEVTDQVTFIDNHDMERFATKVANNQTAVNQ-AYALLLT 379

Query: 358 QQGTPCLFYGDYYGIQGE------FGQPSFK------EVIDKMAELR---QNYVFGKQVD 402
            +G P ++YG      G+          PSF       +VI K+A LR   Q   +G
Sbjct: 380 SRGVPNIYYGTEQYATGDKDPNNRGDMPSFNKESQAYKVISKLAPLRKQNQALAYGTTEQ 439

Query: 403 YFTHSNCIGWTCLGDEEHNSCLAVVLTNGDQ--GWKHMEVGEIYAGKTFVDYLGNC--EQ 458
           +   + +    + + +  + +A+V  N DQ  G+             ++ D L      Q
Sbjct: 440 RWISDHVLVFERKFGNHVALVAINRDQTNGYTITNAKTALPQNSYKDKLEGLLGGQ 495

Query: 459 EVVIGDDGW-GDFLVESASISAWVPKIEE 486
           E+++G DG     F + +  ++ W  + E+
Sbjct: 496 ELIVGADGTISSFELGAGQVAVWTYEGED 524
```

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

Example 375

A DNA sequence (GBSx0406) was identified in *S.agalactiae* <SEQ ID 1225> which encodes the amino acid sequence <SEQ ID 1226>. This protein is predicted to be catabolite control protein A. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2154(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 9707> which encodes amino acid sequence <SEQ ID 9708> was also identified.

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

```
>GP:BAA88121 GB:AB028599 catabolite control protein A [Streptococcus
          bovis] (ver 3)
 Identities = 304/332 (91%), Positives = 320/332 (95%)

Query: 1   MNTDDTITIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNAVARGLA 60
           MNTDDTITIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNAVARGLA
Sbjct: 1   MNTDDTITIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNAVARGLA 60

Query: 61  SKKTTTVGVVIPNIANSYFSILARGIDDIAAMYKYNIVLASSDEDDDKEVNVVNTLFAKQ 120
           SKKTTTVGVVIPNIANSYFSILA+GIDDIAAMYKYNIVLASSDEDDDKEVNVVNTLFAKQ
Sbjct: 61  SKKTTTVGVVIPNIANSYFSILAKGIDDIAAMYKYNIVLASSDEDDDKEVNVVNTLFAKQ 120

Query: 121 VDGIIFMGHHLTEKIRAEFSRSRTPIVLAGTVDLEHQLPSVNIDYKAAAVDVIDILAGNH 180
           VDGIIFMGHHLTEKIRAEFSRSRTP+VLAGTVDLEHQLPSVNIDYKAA  DV+DILA N+
Sbjct: 121 VDGIIFMGHHLTEKIRAEFSRSRTPVVLAGTVDLEHQLPSVNIDYKAAVADVVDILAKNN 180

Query: 181 KDIAFVSGPLIDDINGKVRLAGYKEGLKKNGLNFKEGLVFEANYRYAEGFALAQRVINAG 240
           KDIAFVSGPLIDDINGKVRLAGYKEGL+KN L+FKEGLVFEANY Y +G+ LAQRV+N+G
Sbjct: 181 KDIAFVSGPLIDDINGKVRLAGYKEGLEKNNLSFKEGLVFEANYNYKDGYELAQRVMNSG 240

Query: 241 ATAAYVAEDELAAGLLNGLFEAGKRVPEDFEIITSNDSPIAQYTRPNLTSISQPVYDLGA 300
           ATAAYVAEDELAAGLLNGLF AGK+VPEDFEI+TSNDSPI  YTRPNL+SISQPVYDLGA
Sbjct: 241 ATAAYVAEDELAAGLLNGLFAAGKKVPEDFEILTSNDSPITSYTRPNLSSISQPVYDLGA 300

Query: 301 VSMRMLTKIMHKEELEEKEIVLNHGIVKRGTT 332
           VSMRMLTKIM+KEELEEKEI+LNHG+  RGTT
Sbjct: 301 VSMRMLTKIMNKEELEEKEIILNHGLKLRGTT 332
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1227> which encodes the amino acid sequence <SEQ ID 1228>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2154(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 = 307/332 (92%), Positives = 320/332 (95%)

Query: 1   MNTDDTITIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNAVARGLA 60
           MNTDD +TIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNAVARGLA
Sbjct: 1   MNTDDPLTIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNAVARGLA 60
```

```
Query: 61   SKKTTTVGVVIPNIANSYFSILARGIDDIAAMYKYNIVLASSDEDDDKEVNVVNTLFAKQ 120
            SKKTTTVGVVIPNIANSYFSILA+GIDDIAAMYKYNIVLASSDEDDDKEVNVVNTLFAKQ
Sbjct: 61   SKKTTTVGVVIPNIANSYFSILAKGIDDIAAMYKYNIVLASSDEDDDKEVNVVNTLFAKQ 120

Query: 121  VDGIIFMGHHLTEKIRAEFSRSRTPIVLAGTVDLEHQLPSVNIDYKAAAVDVIDILAGNH 180
            VDGIIFMGHHLTEKIRAEFSRSRTP+VLAGTVDL+HQLPSVNIDY+AA  +V+DILA NH
Sbjct: 121  VDGIIFMGHHLTEKIRAEFSRSRTPVVLAGTVDLDHQLPSVNIDYRAAVSNVVDILAENH 180

Query: 181  KDIAFVSGPLIDDINGKVRLAGYKEGLKKNGLNFKEGLVFEANYRYAEGFALAQRVINAG 240
            K IAFVSGPLIDDINGKVRLAGYKEGLK N L+FKEGLVFEANY Y EGF LAQRVIN+G
Sbjct: 181  KCIAFVSGPLIDDINGKVRLAGYKEGLKHNKLDFKEGLVFEANYSYKEGFELAQRVINSG 240

Query: 241  ATAAYVAEDELAAGLLNGLFEAGKRVPEDFEIITSNDSPIAQYTRPNLTSISQPVYDLGA 300
            ATAAYVAEDELAAGLLNGLFEAGKRVPEDFEIITSNDSP+ QYTRPNL+SISQPVYDLGA
Sbjct: 241  ATAAYVAEDELAAGLLNGLFEAGKRVPEDFEIITSNDSPVVQYTRPNLSSISQPVYDLGA 300

Query: 301  VSMRMLTKIMHKEELEEKEIVLNHGIVKRGTT 332
            VSMRMLTKIM+KEELEEKEI+LNHGI KRGTT
Sbjct: 301  VSMRMLTKIMNKEELEEKEILLNHGIKKRGTT 332
```

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

Example 376

A DNA sequence (GBSx0407) was identified in *S.agalactiae* <SEQ ID 1229> which encodes the amino acid sequence <SEQ ID 1230>. This protein is predicted to be PepQ (pepQ-2). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1118(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:AAC46293 GB:AF014460 PepQ [Streptococcus mutans]
 Identities = 257/359 (71%), Positives = 304/359 (84%)

Query: 1    MSKLNRIRHHLHSVQAELAVFSDPVTVNYLTGFFCDPHERQMFLFVYEDRDPILFVPALE 60
             MSKL +I   L     E AV SDPV++NYLTGF+ DPHER MFLF++ D++ +LF+P L+
Sbjct: 1    MSKLAQIVQKLKKQGIEAAVLSDPVSINYLTGFYSDPHERLMFLFLFADQETLLFLPELD 60

Query: 61   VSRAKQSVPFPVFGYIDSENPWQKIASNLPSFSVSKVLAEFDNLNVTKFQGLQTVFDGHF 120
               RAK  +   V GY+D ENP +KI + LP  + SK+  EFDNLNVTKF+GL+T+F G F
Sbjct: 61   ALRAKSILDISVTGYLDFENPLEKIKTLLPKTNYSKIALEFDNLNVTKFKGLETIFSGQF 120

Query: 121  ENLTPYIQNMRLIKSRDEIEKMLVAGEFADKAVQVGFDNISLNNTETDIIAQIEFEMKKQ 180
             NLTP I  MRLIKS DEI+K+L+AGE ADKAVQ+GFD+ISLN TETDIIAQIEFEMKK
Sbjct: 121  TNLTPLINRMRLIKSADEIQKLLIAGELADKAVQIGFDSISLNATETDIIAQIEFEMKKL 180

Query: 181  GINKMSFDTMVLTGNNAANPHGIPGTNKIENNALLLFDLGVETLGYTSDMTRTVAVGKPD 240
             G++KMSF+TMVLTG+NAANPHG+P ++KIENN LLLFDLGVE+ GY SDMTRTVAVG+PD
Sbjct: 181  GVDKMSFETMVLTGSNAANPHGLPASHKIENNHLLLLFDLGVESTGYVSDMTRTVAVGQPD 240

Query: 241  QFKKDIYHLCLEAHQAAIDFIKPGVLASEVDAAARNVIEKAGYGQYFNHRLGHGLGMDVH 300
             QFKKDIY++CLEA   A+DFIKPGV A++VDAAAR+VIEKAGYG YFNHRLGHG+GM +H
Sbjct: 241  QFKKDIYNICLEAQLTALDFIKPGVSAAQVDAAARSVIEKAGYGDYFNHRLGHGIGMGLH 300

Query: 301  EFPSIMAGNDMEIQEGMCFSVEPGIYIPDKVGVRIEDCGYVTKTGFEVFTKTPKELLYF 359
            EFPSIMAGNDM ++EGMCFSVEPGIYIP+KVGVRIEDCG+VTK GFEVFT+TPKELLYF
```

```
Sbjct: 301 EFPSIMAGNDMLLEEGMCFSVEPGIYIPEKVGVRIEDCGHVTKNGFEVFTQTPKELLYF 359
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1231> which encodes the amino acid sequence <SEQ ID 1232>. Analysis of this protein sequence reveals the following:

```
      Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.90   Transmembrane   42 -  58 (  42 -  59)

----- Final Results -----
               bacterial membrane --- Certainty=0.1362(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAC46293 GB:AF014460 PepQ [Streptococcus mutans]
 Identities = 264/359 (73%), Positives = 304/359 (84%)

Query: 1   MTKLDQIRLYLDQKGAELAIFSDPVTINYLTGFFCDPHERQLFLFVYHDLAPVLFVPALE 60
           M+KL QI    L ++G E A+ SDPV+INYLTGF+ DPHER +FLF++ D   +LF+P L+
Sbjct: 1   MSKLAQIVQKLKKQGIEAAVLSDPVSINYLTGFYSDPHERLMFLFLFADQETLLFLPELD 60

Query: 61  VARASQAISFPVFGYVDSENPWEKIKAVLPNTAAKTIYAEFDHLNVNKFHGLQTIFSGQF 120
             RA   +   V GY+D ENP EKIK +LP T    I  EFD+LNV KF GL+TIFSGQF
Sbjct: 61  ALRAKSILDISVTGYLDFENPLEKIKTLLPKTNYSKIALEFDNLNVTKFKGLETIFSGQF 120

Query: 121 NNLTPYVQGMRLVKSADEINKMMIAGQFADKAVQVGFDNISLDATETDVIAQIEFEMKKQ 180
            NLTP +  MRL+KSADEI K++IAG+ ADKAVQ+GFD+ISL+ATETD+IAQIEFEMKK
Sbjct: 121 TNLTPLINRMRLIKSADEIQKLLIAGELADKAVQIGFDSISLNATETDIIAQIEFEMKKL 180

Query: 181 GIHKMSFDTMVLTGNNAANPHGIPGTNNIENNALLLFDLGVETLGYTSDMTRTVAVGQPD 240
           G+ KMSF+TMVLTG+NAANPHG+P ++ IENN LLLFDLGVE+ GY SDMTRTVAVGQPD
Sbjct: 181 GVDKMSFETMVLTGSNAANPHGLPASHKIENNHLLLFDLGVESTGYVSDMTRTVAVGQPD 240

Query: 241 QFKIDIYNLCLEAQLAAIDFIKPGVTAAQVDAAARQVIEKAGYGEYFNHRLGHGIGMDVH 300
           QFK DIYN+CLEAQL A+DFIKPGV+AAQVDAAAR VIEKAGYG+YFNHRLGHGIGM +H
Sbjct: 241 QFKKDIYNICLEAQLTALDFIKPGVSAAQVDAAARSVIEKAGYGDYFNHRLGHGIGMGLH 300

Query: 301 EFPSIMAGNDLVLEEGMCFSVEPGIYIPGKVGVRIEDCGHVTKNGFEVFTHTPKELLYF 359
           EFPSIMAGND++LEEGMCFSVEPGIYIP KVGVRIEDCGHVTKNGFEVFT TPKELLYF
Sbjct: 301 EFPSIMAGNDMLLEEGMCFSVEPGIYIPEKVGVRIEDCGHVTKNGFEVFTQTPKELLYF 359
```

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

```
 Identities = 288/361 (79%), Positives = 325/361 (89%)

Query: 1   MSKLNRIRHHLHSVQAELAVFSDPVTVNYLTGFFCDPHERQMFLFVYEDRDPILFVPALE 60
           M+KL++IR +L    AELA+FSDPVT+NYLTGFFCDPHERQ+FLFVY D  P+LFVPALE
Sbjct: 1   MTKLDQIRLYLDQKGAELAIFSDPVTINYLTGFFCDPHERQLFLFVYHDLAPVLFVPALE 60

Query: 61  VSRAKQSVPFPVFGYIDSENPWQKIASNLPSFSVSKVLAEFDNLNVTKFQGLQTVFDGHF 120
           V+RA Q++ FPVFGY+DSENPW+KI + LP+ +   + AEFD+LNV KF GLQT+F G F
Sbjct: 61  VARASQAISFPVFGYVDSENPWEKIKAVLPNTAAKTIYAEFDHLNVNKFHGLQTIFSGQF 120

Query: 121 ENLTPYIQNMRLIKSRDEIEKMLVAGEFADKAVQVGFDNISLNNTETDIIAQIEFEMKKQ 180
            NLTPY+Q MRL+KS DEI KM++AG+FADKAVQVGFDNISL+ TETD+IAQIEFEMKKQ
Sbjct: 121 NNLTPYVQGMRLVKSADEINKMMIAGQFADKAVQVGFDNISLDATETDVIAQIEFEMKKQ 180

Query: 181 GINKMSFDTMVLTGNNAANPHGIPGTNKIENNALLLFDLGVETLGYTSDMTRTVAVGKPD 240
           GI+KMSFDTMVLTGNNAANPHGIPGTN IENNALLLFDLGVETLGYTSDMTRTVAVG+PD
Sbjct: 181 GIHKMSFDTMVLTGNNAANPHGIPGTNNIENNALLLFDLGVETLGYTSDMTRTVAVGQPD 240

Query: 241 QFKKDIYHLCLEAHQAAIDFIKPGVLASEVDAAARNVIEKAGYGQYFNHRLGHGLGMDVH 300
           QFK DIY+LCLEA  AAIDFIKPGV A++VDAAAR VIEKAGYG+YFNHRLGHG+GMDVH
Sbjct: 241 QFKIDIYNLCLEAQLAAIDFIKPGVTAAQVDAAARQVIEKAGYGEYFNHRLGHGIGMDVH 300
```

```
Query: 301 EFPSIMAGNDMEIQEGMCFSVEPGIYIPDKVGVRIEDCGYVTKIGFEVFTKTPKELLYFEG 361
           EFPSIMAGND+ ++EGMCFSVEPGIYIP KVGVRIEDCG+VTK GFEVFT TPKELLYFEG
Sbjct: 301 EFPSIMAGNDLVLEEGMCFSVEPGIYIPGKVGVRIEDCGHVTKNGFEVFTHTPKELLYFEG 361
```

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

Example 377

A DNA sequence (GBSx0408) was identified in *S.agalactiae* <SEQ ID 1233> which encodes the amino acid sequence <SEQ ID 1234>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3629(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 378

A DNA sequence (GBSx0409) was identified in *S.agalactiae* <SEQ ID 1235> which encodes the amino acid sequence <SEQ ID 1236>. This protein is predicted to be beta-hexosamidase A precursor. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3279(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:CAB11942 GB:Z99104 alternate gene name: yzbA~similar to
          beta-hexosaminidase [Bacillus subtilis]
 Identities = 151/602 (25%), Positives = 268/602 (44%), Gaps = 69/602 (11%)

Query: 26  INEMTLDEKIGQLF------FNMGASRSEEYLTDVLDRYHIAAVRYNRGSSSEIYDQNL~ 78
           +N M+LDEK+GQ+       +      S + LT + D       +Y  G    ++ +N+
Sbjct: 39  VNRMSLDEKLGQMLMPDFRNWQKEGESSPQALTKMNDEVASLVKKYQFGGII~LFAENVK 97

Query: 79  -----------ILQTKSKLPMLIAANTEAGGDGAVTDGTKVGDEIKVAATNDPKYAYEMG 127
                      +   K+P+++ + E G    + +GT     + + A      AY+ G
Sbjct: 98  TTKQTVQLTDDYQKASPKIPLMLSIDQEGGIVTRLGEGTNFPGNMALGAARSRINAYQTG 157

Query: 128 RIAGMEASAVGCNASFSPIVDLTRNWRNPIIASRNWGANVDQIISLSKEYMKGIMQYNIV 187
            I G E SA+G N  FSP+VD+  N  NP+I  R++ +N +    L    MKG+ + +I
Sbjct: 158 SIIGKELSALGINTDFSPVVDINNNPDNPVIGVRSFSSNRELTSRLGLYTMKGLQRQDIA 217

Query: 188 PFAKHFPGDGIDERDHHLSFASNPMSKEEWMSTFGRIYGELADAGLPGVMAGHIHLPNVE 247
              KHFPG G  + D H          +E          + +  DAG   VM  H+  P  +
Sbjct: 218 SALKHFPGHGDTDVDSHYGLPLVSHGQERLREVELYPFQKAIDAGADMVMTAHVQFPAFD 277
```

```
Query: 248 KEMHPER--DLDDMLPASLNKTLLDELLRGELGYNGAIVTDASHMVGMTASMARRDLLPT 305
              +  +    D ++PA+L+K ++  LLR E+G+NG IVTDA +M  +     + + +
Sbjct: 278 DTTYKSKLDGSDILVPATLSKKVMTGLLRQEMGFNGVIVTDALNMKAIADHFGQEEAVVM 337

Query: 306 AIEAGCDLFLF---FNDPDED------IQWMKEGYEKGILTEERLHDALRRTLGLKAKLG 356
            A++AG D+ L          E+      IQ +KE  + G + E+++++++ R + LK K G
Sbjct: 338 AVKAGVDIALMPASVTSLKEEQKFARVIQALKEAVKNGDIPEQQINNSVERIISLKIKRG 397

Query: 357 LHNYEGRRQELFMPK-DKAMALINTLESQKIADEVADKAVTLVKDKQKDIFPVNPERYRH 415
            +  Y  R  +    K  KA  ++ + +  K   ++A+KAVT++K++Q  + P  P++
Sbjct: 398 M--YPARNSDSTKEKIAKAKKIVGSKQHLKAEKKLAEKAVTVLKNEQHTL-PFKPKKGSR 454

Query: 416 ILLVNVEGYKGGFGAMIAGNKQRASDYMKE------LLEARGHEVTVWESTEERIMKLPQ 469
            IL+V        +    A  +Q   D +K         L         V+++  E+ +K
Sbjct: 455 ILIV------APYEEQTASIEQTIHDLIKRKKIKPVSLSKMNFASQVFKTEHEKQVK--- 505

Query: 470 EERAAAIANVYAQK-QPIANLTEHYDLIINLVDVNAGGTTQRIIWPAAKGTPDQPFYVHE 528
             E    I   Y  K  P+ N      D +I+  D  +  +    ++P A     +  H
Sbjct: 506 -EADYIITGSYVVKNDPVVN-----DGVID--DTISDSSKWATVFPRA---VMKAALQHN 554

Query: 529 IPSIVISVQHAFALADMPQVGTYINAYD--------GLPSTISAVVAKLAGESEFTGVSP 580
             P +++S+++ +  A+  +    I  Y          L   I A V  + G+++  G  P
Sbjct: 555 KPFVLMSLRNPYDAANFEEAKALIAVYGFKGYANGRYLQPNIPAGVMAIFGQAKPKGTLP 614

Query: 581 VD 582
           VD
Sbjct: 615 VD 616
```

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

A related GBS gene <SEQ ID 8565> and protein <SEQ ID 8566> were also identified. Analysis of this protein sequence reveals the following homology to a lipoprotein, with homology with the following sequences in the databases:

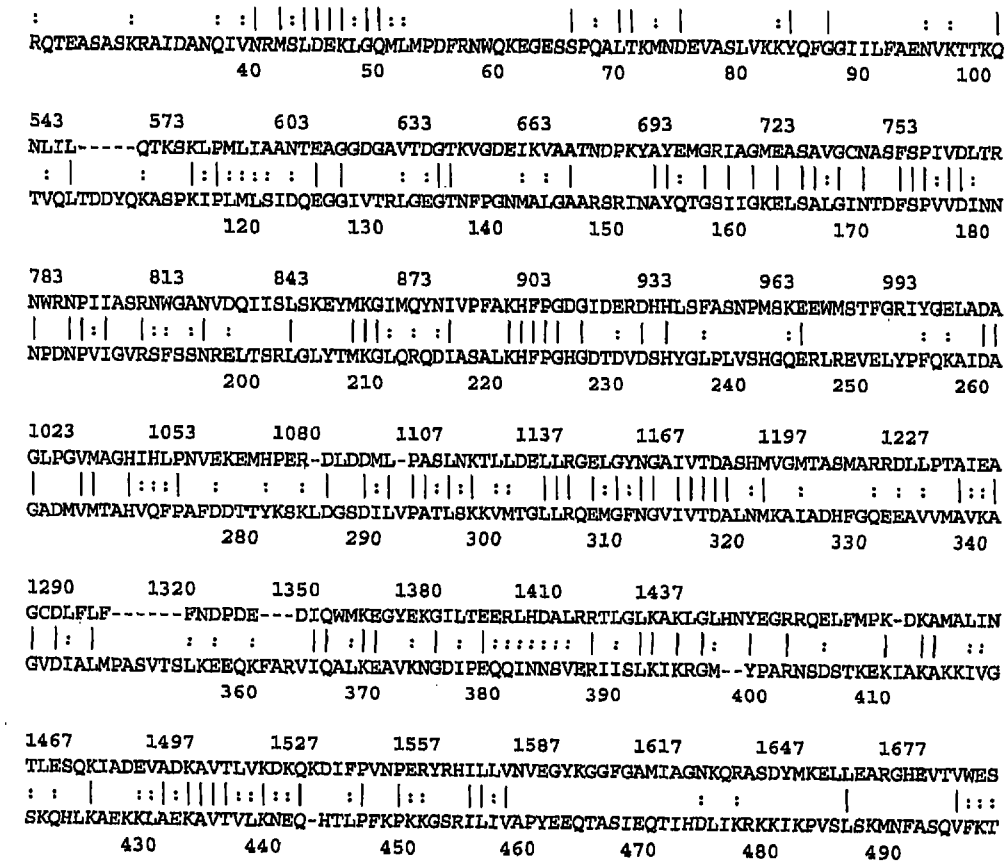
```
29.5/52.3% over 422aa
                                                          Bacillus subtilis
  EGAD|20114| hypothetical 70.6 kd protein in feua 5'region precursor Insert characterized
   SP|P40406|YBBD_BACSU HYPOTHETICAL 70.6 KDA LIPOPROTEIN IN FEUA-SIGW INTERGENIC REGION
PRECURSOR (ORF1). Insert
 characterized
  GP|1944006|dbj|BAA19499.1||AB002150 YbbD Insert characterized
   GP|438455|gb|AAA64351.1||L19954 possible N-terminal signal sequence; mature protein may
be membrane-anchored and start at Cys-17. 17.5% identity
 over 354-aa overlap with Candida pelliculosa beta-glucosidase.; putative Insert
characterized
  GP|2632433|emb Insert characterized

ORF00431(367 - 1557 of 2388)
EGAD|20114|BS0166(36 - 458 of 642) hypothetical 70.6 kd protein in feua 5'region precursor
(Bacillus subtilis) SP|P40406|YBBD_BACSU HYPOTHETICAL 70.6 KDA LIPOPROTEIN IN FEUA-SIGW
INTERGENIC REGION PRECURSOR (ORF1). GP|1944006|dbj|BAA19499.1||AB002150 YbbD (Bacillus
subtilis) GP|438455|gb|AAA64351.1||L19954 possible N-terminal signal sequence; mature
protein may be membrane-anchored and start at Cys-17. 17.5% identity over 354-aa overlap
with Candida pelliculosa beta-glucosidase.; putative (Bacillus subtilis) GP|2632433|emb
%Match = 9.6
%Identity = 29.5  %Similarity = 52.2
Matches = 119  Mismatches = 183  Conservative Sub.s = 92

114       144       174       204       234       264       294       324
LMVGDSLGDLAAAEQNGIAFYPVLVGKEVKSWEILREDIGEAFAKGQFEQQRQKESINTFWANLDN**KG*AMTHLVDLT

                                                        MRPVFPLILSAVLFLSCFFGA
                                                                10        20

354       384       414       426       456       486                 528
KKPFNLNQEAIEWIEKTINEMTLDEKIGQLFF------NMGASRSEEYLTDVLDRYHIAAVRYNRGS------SSEIYDQ
```



SEQ ID 1236 (GBS50) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 8; MW 69.2kDa).

GBS50-His was purified as shown in Figure 192, lane 5.

The GBS50-His fusion product was purified (Figure 192, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 264), 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 379

A DNA sequence (GBSx0410) was identified in *S.agalactiae* <SEQ ID 1237> which encodes the amino acid sequence <SEQ ID 1238>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

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

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

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

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

Example 380

A DNA sequence (GBSx0411) was identified in *S.agalactiae* <SEQ ID 1239> which encodes the amino acid sequence <SEQ ID 1240>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2279(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 9705> which encodes amino acid sequence <SEQ ID 9706> was also identified.

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

```
>GP:AAC21726 GB:U32690 oxidoreductase [Haemophilus influenzae Rd]
 Identities = 197/271 (72%), Positives = 229/271 (83%)

Query: 26  NKVVVITGAGGVLCGYMAKEFAKAGAKVALLDLNQEAAQTFADEIVEEGGIAKAYKANVL 85
           NK+++ITGAGGVLC ++AK+ A   A +ALLDLN EAA   A EI + GG AKAYK NVL
Sbjct: 15  NKLIIITGAGGVLCSFLAKQLAYTKANIALLDLNFEAADKVAKEINQSGGKAKAYKTNVL 74

Query: 86  SKENLEEVHQAVLEDLGPTDILVNGAGGNNPKATTDNEFHELDLPSETKTFFELDEAGIS 145
             EN++EV   +  D G  DIL+NGAGGNNPKATTDNEFH+ DL   T+TFF+LD++GI
Sbjct: 75  ELENIKEVRNQIETDFGTCDILINGAGGNNPKATTDNEFHQFDLNETTRTFFDLDKSGIE 134

Query: 146 FVFNLNYLGTLLPTQVFAQDMVGREGANIINISSMNAFTPLTKIPAYSGAKAAISNFTQW 205
           FVFNLNYLG+LLPTQVFA+DM+G++GANIINISSMNAFTPLTKIPAYSGAKAAISNFTQW
Sbjct: 135 FVFNLNYLGSLLPTQVFAKDMLGKQGANIINISSMNAFTPLTKIPAYSGAKAAISNFTQW 194

Query: 206 LAVHFSKVGIRCNAIAPGFLVTNQNRSLLFTEDGQPTARAEKILNNTPMGRFGEASELIG 265
           LAV+FSKVGIRCNAIAPGFLV+NQN +LLF  +G+PT RA KIL NTPMGRFGE+ EL+G
Sbjct: 195 LAVYFSKVGIRCNAIAPGFLVSNQNLALLFDTEGKPTDRANKILTNTPMGRFGESEELLG 254

Query: 266 GLFFLADEKSSSFVNGVVLPIDGGFAAYSGV 296
            L FL DE  S+FVNGVVLP+DGGF+AYSGV
Sbjct: 255 ALLFLIDENYSAFVNGVVLPVDGGFSAYSGV 285
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1241> which encodes the amino acid sequence <SEQ ID 1242>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0358(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 = 77/279 (27%), Positives = 125/279 (44%), Gaps = 19/279 (6%)

Query: 18  MSKTITFTNKVVVITGAGGVLCGYMAKEFAKAGAKVALLDLNQEAAQTFADEIVEEGGIA 77
           M    +   K+ +ITGA   +   +AK +A+AGA +   D+ QE          E G  A
Sbjct: 1   MENMFSLQGKIALITGASYGIGFEIAKAYAQAGATIVFNDIKQELVDKGLAAYRELGIEA 60

Query: 78  KAYKANVLSKENLEEVHQAVLEDLGPTDILVNGAGGNNPKATTDNEFHELDLPSETKTFF 137
```

```
              Y  +V  +  ++++   + +++G  DILVN AG
Sbjct: 61  HGYVCDVTDEAGIQQMVSQIEDEVGAIDILVNNAG----------------IIRRTPML 103

Query: 138 ELDEAGISFVFNLNYLGTLLPTQVFAQDMVGREGANIINISSMNAFTPLTKIPAYSGAKA 197
           E+       V +++     + ++     M+ +    IINI SM +      + AY+ AK
Sbjct: 104 EMAAEDFRQVIDIDLNAPFIVSKAVLPSMIAKGHGKIINICSMMSELGRETVSAYAAAKG 163

Query: 198 AISNFTQWLAVHFSKVGIRCNAIAPGFLVTNQNRSLLFTE-DGQPTARAEKILNNTPMGR 256
            +   T+ +A  F +  I+CN I PG++ T Q    L    + DG       + I+  TP  R
Sbjct: 164 GLKMLTKNIASEFGEANIQCNGIGPGYIATPQTAPLRERQADGSRHPFDQFIIAKTPAAR 223

Query: 257 FGEASELIGGLFFLADEKSSSFVNGVVLPIDGGFAAYSG 295
           +G   +L G   FLA + +S+FVNG +L +DGG  AY G
Sbjct: 224 WGTTEDLAGPAVFLASD-ASNFVNGHILYVDGGILAYIG 261
```

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

Example 381

A DNA sequence (GBSx0412) was identified in *S.agalactiae* <SEQ ID 1243> which encodes the amino acid sequence <SEQ ID 1244>. This protein is predicted to be D-mannonate dehydrolase (uxuA). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3188(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:BAB04425 GB:AP001509 D-mannonate dehydrolase [Bacillus halodurans]
 Identities = 202/343 (58%), Positives = 261/343 (75%)

Query: 1   MEMSFRWYGEDDPVTLENIGQIPTMKGIVTAIYDVPVGEVWSRERIQQLKEKVEAAGLKI 60
           M ++ RW+G  D V LE I QIP MKGIV+AIYDV VG VW +E+I  LK  +E  GL +
Sbjct: 1   MRLTMRWFGPSDKVKLEYIKQIPGMKGIVSAIYDVAVGGVWPKEKILALKNNIERHGLTL 60

Query: 61  SVIESVPVHEDIKLGRPTRDLLIDNYIQTVKNLAAEGIDTICYNFMPVFDWTRTDLAYQY 120
            VIESVPVHEDIKLG+PTRD  I+NY QT+++LA  GIDT+CYNFMPVFDWTR+ L ++
Sbjct: 61  DVIESVPVHEDIKLGKPTRDRYIENYKQTLRHLAECGIDTVCYNFMPVFDWTRSQLDFKL 120

Query: 121 PDGSTALIFDETVSKKMDPVNGELSLPGWDASYSKEEMKAIMDAYAEIDEEKLWENLTYF 180
            DGS ALI++E V  + +P++GEL LPGWD SY  E +K ++ AY +I EE LW++LTYF
Sbjct: 121 EDGSEALIYEEDVISRTNPLSGELELPGWDTSYENESLKGVLQAYKKISEEDLWDHLTYF 180

Query: 181 IKRIIPEAEAVGVKMAIHPDDPPYSIFGLPRIITGLEAIERFVKLYDSKSNGITLCVGSY 240
           ++ I+P A+ VG+KMAIHPDDPP+SIFGLPRI+T    +ER + LYDS ++GIT+C GS
Sbjct: 181 VQAIMPVADEVGIKMAIHPDDPPWSIFGLPRIVTNKANLERLLSLYDSPNHGITMCSGSL 240

Query: 241 ASDPQNDVLEISRRAFELDRVNFVHARNIKLGDGKSFKESAHPSEYGSIDMYEVIKLCHE 300
           ++  ND+ E+ R      R++F HARNIK    +SF+ESAH SE GS++M  ++K  H+
Sbjct: 241 GANEANDLPEMIRHFGGQGRIHFAHARNIKRTGPRSFQESAHLSEAGSVNMVAMLKAYHD 300

Query: 301 FGFEGAIRPDHGRMIWGETGRPGYGLYDRALGATYVSGLYEAV 343
           GF G +RPDHGRMIWGE GRPGYGLYDRALGATY++G++EAV
Sbjct: 301 IGFTGPLRPDHGRMIWGEKGRPGYGLYDRALGATYLNGIWEAV 343
```

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 382

A DNA sequence (GBSx0413) was identified in *S.agalactiae* <SEQ ID 1245> which encodes the amino acid sequence <SEQ ID 1246>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2447(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 383

A DNA sequence (GBSx0414) was identified in *S.agalactiae* <SEQ ID 1247> which encodes the amino acid sequence <SEQ ID 1248>. This protein is predicted to be uronate isomerase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3066(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:BAB04424 GB:AP001509 uronate isomerase [Bacillus halodurans]
 Identities = 215/465 (46%), Positives = 294/465 (62%), Gaps = 7/465 (1%)

Query: 3   FNTETFMLKNQAAIQLYEE-VKRQPIFDYHCHLDPKDIFEDHIFDNIVDLWLGGDHYKWR 61
           F +E F+L N+    +LY     K  PI DYHCHL P++I+E+  F+N+    WLGGDHYKWR
Sbjct: 4   FLSEDFLLMNEYDRELYYTFAKNMPICDYHCHLSPQEIWENKPFENMTKAWLGGDHYKWR 63

Query: 62  LMRANGISEAEITGPASNLEKFKAFARTLERAYGNPVYHWSAMELKNVFGVNEILTESNA 121
            MR NG+ E  ITG A + EKF A+A+T+ +  GNP+YHW+ MELK   F  ++ L E+N
Sbjct: 64  AMRLNGVREEFITGGAPDKEKFLAWAKTVPKTIGNPLYHWTHMELKTYFHFHQPLDETNG 123

Query: 122 EEIYHRLNHFLKEHKISPRRLIADSKVMFIGTTDHPLDTLEWHKKLAADESFKTVVAPTF 181
           E ++   N  L++   +PR LI  S V  IGTTD P D+L +H+KL AD++F   V PTF
Sbjct: 124 ENVWDACNRLLQQEAFTPRALIERSNVRAIGTTDDPTDSLLYHQKLQADDTFHVKVIPTF 183

Query: 182 RPDEAF-IEHRHFVDFITKLGDITQKEITDFSTFIAAMEERIAYFAQNGCRASDISFTEI 240
           RPD A  IE    F D++ KL D+T + +      F+ A++ER+ +F ++GCR+SD   TE+
Sbjct: 184 RPDGALKIEQDSFADWVAKLSDVTGESLDTLDAFLHALKERLTFFDEHGCRSSDHDMTEV 243

Query: 241 VFEQTDELELNDLFNKVCEGYIPNQSEISKWQTAVFMELCRLYKKYGFVTQVHFGALRNN 300
            F + +E E    +F K          + E  K++T +    L + Y   G+V Q H G +RNN
Sbjct: 244 PFVEVNEQEAQHIFRKRLANEGLTKVENEKYKTFLMTWLGKEYAARGWVMQWHIGVMRNN 303

Query: 301 HSTIFEKLGADVGVDSLGD-QVALTVNMNRLLDSLVKKDSLPKMIWYNLNPAYNIAVANT 359
```

```
            +S +  KLG D G DS+GD Q+A      +LLD L K+ +LPK I Y +NP  N  +A+
Sbjct: 304 NSRMLHKLGPDTGFDSIGDGQIAHAT--AKLLDLLDKQGALPKTILYCVNPNANYILASM 361

Query: 360 LANFQANELGVRSYLQFGAGWWFADTKLGMISQMNALAEQGMLANFIGMLTDSRSFLSYQ 419
           + NF   E GVR  +QFG+ WWF D   GM  Q+  LA  G+L+NFIGMLTDSRSFLSY
Sbjct: 362 IGNF--TESGVRGKVQFGSAWWFNDHIDGMRRQLTDLASVGLLSNFIGMLTDSRSFLSYP 419

Query: 420 RHDYFRRILCTYLGEWIEEGEVPEDYQALGSMAKDIAYQNAVNYF 464
           RHDYFRRILC  +G WI+EG++P D +  G + +DI Y N V+YF
Sbjct: 420 RHDYFRRILCQLIGSWIKEGQLPPDMERWGQIVQDICYNNVVDYF 464
```

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 384

A DNA sequence (GBSx0415) was identified in *S.agalactiae* <SEQ ID 1249> which encodes the amino acid sequence <SEQ ID 1250>. This protein is predicted to be 2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate al. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3883(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 9703> which encodes amino acid sequence <SEQ ID 9704> was also identified.

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

```
>GP:AAD35160 GB:AE001693 2-dehydro-3-deoxyphosphogluconate
          aldolase/4-hydroxy-2-oxoglutarate aldolase [Thermotoga maritima]
 Identities = 93/199 (46%), Positives = 125/199 (62%), Gaps = 6/199 (3%)

Query: 37  KNNYFFAVIRGKSSEDALEIAKHAILGGIRNIEVTFSTPEASKVIKQLSDDFKNNKEIIV 96
           K +   AV+R  S E+A E A      GG+  IE+TF+ P+A  VIK+LS  F   K  I+
Sbjct: 8   KKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELS--FLKEKGAII 65

Query: 97  GAGTVMTTELAKEAIDAGAKFLVSPHFDSDIANLANENKVYYFPGCATATEIVVARKYKC 156
           GAGTV + E  ++A+++GA+F+VSPH D +I+    E  V+Y PG  T TE+V A K
Sbjct: 66  GAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGH 125

Query: 157 QIIKLFPGGVVGPGFIKDIHGPIPDVDLMPSGGVSVSNVVEWRKAGAVAVGVGSALSSKV 216
            I+KLFPG VVGP F+K + GP P+V  +P+GGV++ NV EW KAG +AVGVGSAL
Sbjct: 126 TILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGT 185

Query: 217 ATEGYDSVTKIAKQFVSAL 235
                D V + AK FV  +
Sbjct: 186 P----DEVREKAKAFVEKI 200
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1251> which encodes the amino acid sequence <SEQ ID 1252>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1039(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 = 82/204 (40%), Positives = 132/204 (64%)

Query: 32  MLNQLKNNYFFAVIRGKSSEDALEIAKHAILGGIRNIEVTFSTPEASKVIKQLSDDFKNN 91
           +L +LK N    V+RG+SSE+AL  +  +I GGI+ IEVT++ P AS+VI QL++ FK +
Sbjct: 6   ILTKLKANRLVLVVRGESSEEALACSLASIEGGIKTIEVTYTNPFASEVIGQLAERFKED 65

Query: 92  KEIIVGAGTVMTTELAKEAIDAGAKFLVSPHFDSDIANLANENKVYYFPGCATATEIVVA 151
            E+++GAGTV+    A++AI AGA+F+V P+F+  +A + +   + Y PGC T  E+V A
Sbjct: 66  PEVLIGAGTVLDDVTARQAILAGAQFIVGPNFNRAVALICHRYSIPYLPGCMTVNEVVTA 125

Query: 152 RKYKCQIIKLFPGGVVGPGFIKDIHGPIPDVDLMPSGGVSVSNVVEWRKAGAVAVGVGSA 211
           +    ++K+FPG  VG  FI+ I  P+P V++M +GGVS  N+ +W  AG    +G+G
Sbjct: 126 LESGVDMVKIFPGSTVGISFIRAIKSPLPQVEVMVTGGVSSDNLKDWLAAGVDVLGIGGE 185

Query: 212 LSSKVATEGYDSVTKIAKQFVSAL 235
           +   + + Y+ +TK A  ++ +L
Sbjct: 186 FNQLASQKQYNLITKKAAHYIKSL 209
```

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

Example 385

A DNA sequence (GBSx0416) was identified in *S.agalactiae* <SEQ ID 1253> which encodes the amino acid sequence <SEQ ID 1254>. This protein is predicted to be pyruvate dehydrogenase complex repressor. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2827(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:CAB12044 GB:Z99105 similar to transcriptional regulator (GntR
          family) [Bacillus subtilis]
 Identities = 67/225 (29%), Positives = 119/225 (52%), Gaps = 17/225 (7%)

Query: 3   RPLVEQTADRLLHLILEREYPVGAKLPNEYELAEDLDVGRSTIREAVRSLATRNILEVRQ 62
           + L +Q  +R++HL+   +   G KLP E EL + L V R  +REA+ SL T  ++  +
Sbjct: 16  KTLAKQVIERIVHLLSSGQLRAGDKLPTEMELMDILHVSRPVLREALSSLETLGVITRKT 75

Query: 63  GSGTYISSKKGVSEDPLGFSLIKDTDRLTSDLFELRLLLEPRIAELVAYRITDDQLQLLE 122
            GTY + K G+   P    L   TD L + + E R+ LE  +  + A +I +++LQ L+
Sbjct: 76  RGGTYFNDKIGM--QPFSVMLALATDNLPA-IIEARMALELGLVTIAAEKINEEELQRLQ 132

Query: 123 KLVGDIEDAV--HAGDPKHLLLDVEFHSMLAKYSGNIAMDSLLPVINQSIHLINANYTNR 180
           K + DI ++   H G+      D EFH ++A  + N  ++ ++    QS+ + +A  ++
Sbjct: 133 KTIDDIANSTDNHYGE-----ADKEFHRIIALSANNPVVEGMI----QSLLITHAKIDSQ 183

Query: 181 ---QMKSDSLEAHREIIKAIREKNPVAAHDAMLMHIMSVRRSALK 222
              + +  ++E H++I  A+ +++P  AH  M  H+  VR   LK
Sbjct: 184 IPYRERDVTVEYHKKIYDALAQRDPYKAHYHMYEHLKFVRDKILK 228
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1255> which encodes the amino acid sequence <SEQ ID 1256>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2161(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 = 24/51 (47%), Positives = 35/51 (68%)

Query: 22 YPVGAKLPNEYELAEDLDVGRSTIREAVRSLATRNILEVRQGSGTYISSKK 72
          +P+G++LP+E  LAE    V R T+R+A+  L     ILE R GSGTY++S +
Sbjct: 30 WPIGSRLPSERHLAEHFTVSRMTLRQAITLLVEEGILERRIGSGTYVASHR 80
```

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

Example 386

A DNA sequence (GBSx0417) was identified in *S.agalactiae* <SEQ ID 1257> which encodes the amino acid sequence <SEQ ID 1258>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2178(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 9701> which encodes amino acid sequence <SEQ ID 9702> was also identified.

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

```
>GP:CAA58911 GB:X84105 gluceronidase [synthetic construct]
 Identities = 258/602 (42%), Positives = 357/602 (58%), Gaps = 31/602 (5%)

Query: 23  MLYPLLTKTRNTYDLGGIWNFKLGEHNPN-------ELLPSDEVMVIPTSFNDLMVSKEK 75
           ML P+ T TR     L G+W F L   N           L     + +P SFND     +
Sbjct: 1   MLRPVETPTREIKKLDGLWAFSLDRENCGIDQRWWESALQESRAIAVPGSFNDQFADADI 60

Query: 76  RDYIGDFWYEKVIEVPKVSEDEEMVLRFGSVTHQAKIYVDGVLVGEHKGGFTPFEVLVPE 135
           R+Y G+ WY++ + +PK     + +VLRF +VTH  K++V+   V EH+GG+TPFE  V
Sbjct: 61  RNYAGNVWYQREVFIPKGWAGQRIVLRFDAVTHYGKVWVNNQEVMEHQGGYTPFEADVTP 120

Query: 136 CKYNNEKIKVSICANNVLDYTTLPVGNYSEIIQEDGSIKKKVRENFDFFNYAGVHRPLKL 195
                + +++++C NN L++ T+P G     I  E+G  KKK      DFFNYAG+HR + L
Sbjct: 121 YVIAGKSVRITVCVNNELNWQTIPPGMV--ITDENG--KKKQSYFHDFFNYAGIHRSVML 176

Query: 196 MIRPKNHIFDITITSRLSDDLQSADLHFLVETNQKVDEVRISVFDEDNKLV--GETKDSR 253
              P   + DIT+ + ++ D   A + + V  N     +V + + D D ++V  G+
Sbjct: 177 YTTPNTWVDDITVVTHVAQDCNHASVDWQVVAN---GDVSVELRDADQQVVATGQGTSGT 233

Query: 254 LFLSDVHLWEVLNAYLYTARVEIFVDNQLQDVYEENFGLREIEVTNGQFLLNRKPIYFKG 313
           L + + HLW+     YLY   V      +  D+Y    G+R + V   QFL+N KP YF G
Sbjct: 234 LQVVNPHLWQPGEGYLYELCVTAKSQTEC-DIYPLRVGIRSVAVKGEQFLINHKPFYFTG 292

Query: 314 FGKHEDTFINGRGLNEAANLMDLNLLKDMGANSFRTSHYPYSEEMMRLADRMGVLVIDEV 373
           FG+HED  + G+G +     + D  L+  +GANS+RTSHYPY+EEM+  AD  G++VIDE
Sbjct: 293 FGRHEDADLRGKGFDNVLMVHDHALMDWIGANSYRTSHYPYAEEMLDWADEHGIVVIDET 352

Query: 374 PAVGLFQNFNASLDLS------PKDNGTWNLM--QTKAAHEQAIQELVKRDKNHPSVVMW 425
```

```
             AVG    FN SL +       PK+  +   +  +T+ AH QAI+EL+ RDKNHPSVVMW
Sbjct: 353 AAVG----FNLSLGIGFEAGNKPKELYSEEAVNGETQQAHLQAIKELIARDKNHPSVVMW 408

Query: 426 VVANEPASHEAGAHDYFEPLVKLYKDLDPQKRPVTLVNILMATPDRDQVMDLVDVVCLNR 485
            +ANEP +   GA +YF PL +  + LDP  RP+T VN++      D + DL DV+CLNR
Sbjct: 409 SIANEPDTRPQGAREYFAPLAEATRKLDPT-RPITCVNVMFCDAHTDTISDLFDVLCLNR 467

Query: 486 YYGWYVDHGDLTNAEVGIRKELLEWQDKFPDKPIIITEYGADTLPGLHSTWNIPYTEEFQ 545
            YYGWYV  GDL  AE  + KELL WQ+K   +PIIITEYG DTL GLHS +   ++EE+Q
Sbjct: 468 YYGWYVQSGDLETAEKVLEKELLAWQEKL-HQPIIITEYGVDTLAGLHSMYTDMWSEEYQ 526

Query: 546 CDFYEMSHRVFDGIPNLVGEQVWNFADFETNLMILRVQGNHKGLFSRNRQPKQVVKEFKK 605
            C + +M HRVFD +  +VGEQVWNFADF T+  ILRV GN KG+F+R+R+PK      +K
Sbjct: 527 CAWLDMYHRVFDRVSAVVGEQVWNFADFATSQGILRVGGNKKGIFTRDRKPKSAAFLLQK 586

Query: 606 RW 607
           RW
Sbjct: 587 RW 588
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1259> which encodes the amino acid sequence <SEQ ID 1260>. Analysis of this protein sequence reveals the following:

```
      Possible site: 23

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.04   Transmembrane 1131 -1147 (1130 -1147)

----- Final Results -----
               bacterial membrane --- Certainty=0.2614(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAF97242 GB:AF282987 beta-galactosidase precursor [Streptococcus pneumoniae]
 Identities = 303/921 (32%), Positives = 463/921 (49%), Gaps = 86/921 (9%)

Query: 5   QKSSEIVT----RTITKPSRATSNVKQEIDMTPDSKEQTVTGYQYHYIDQ--EGRKQPFN 58
           +K  E VT    +   KP      ++ +        ++         ++Q  E RK  FN
Sbjct: 96  KKEDEAVTPKEEKVSAKPEEKAPRIESQASNQEKPLKEDAKAVTNEEVNQMIEDRKVDFN 155

Query: 59  QGWRF-LMADVACAQDPSFDDSNWQVIHLPHDFSLTQPYTRNGEA--ESAYKLGGVGWYR 115
           Q W F L A+   A  P  D S W+ + LP+D+S+   +    A  E     GG  WYR
Sbjct: 156 QNWYFKLNANSKEAIKPDADVSTWKKLDLPYDWSIFNDFDHESPAQNEGGQLNGGEAWYR 215

Query: 116 HYLVLDEVLAGCHVAITFEGSYMETEIYVNGQFIGKHLNGYQEFTYDISDVVTF-GAENL 174
               LDE     +V +TF+G YM++++YVNGQ +G + NGY +F+YDI+  +   G EN+
Sbjct: 216 KTFKLDEKDLKKNVRLTFDGVYMDSQVYVNGQLVGHYPNGYNQFSYDITKYLQKDGRENV 275

Query: 175 LAVRVENKVPSSRWYSGSGLYREVSLSVLPQLHFVADQVAMTLADTAVQEKGQQKVDLRF 234
           +AV   NK PSSRWYSGSG+YR+V+L V  ++H   +   +       Q+ G+ +  +
Sbjct: 276 IAVHAVNKQPSSRWYSGSGIYRDVTLQVTDKVHVEKNGTTILTPKLEEQQHGKVETHVTS 335

Query: 235 ALNQSIQTCHYQLSLCLWEQSHCSKDKKLLYQETEVPLADLAFQRQYGLT--LSLEELQL 292
            + +   H ++   E     +      L     L +E  +L
Sbjct: 336 KIVNTDDKDHELVA----EYQIVERGGHAVTGLVRTASRTLKAHESTSLDAILEVERPKL 391

Query: 293 WSP--DNPHLYDLELTLYYQGQVIDCFCLETGFRQLTFMANQGLFVNGRAVKLKGVCLHH 350
           W+   D P LY+L   +Y  GQ++D      G+R   + N+G  +NG  +K FHGVSLHH
Sbjct: 392 WTVLNDKPALYELITRVYRDGQLVDAKKDLFGYRYYHWTPNEGFSLNGERIKFHGVSLHH 451

Query: 351 DQGGLGACAYEDALARQLVLLKDMGANTIRSTHNPSSPKLRQLANRLGFFVIEEAFDTWT 410
           D G LGA     A  R+L  +K+MG N+IR+THNP+S +  Q+A  LG  V EEAFDTW
Sbjct: 452 DHGALGAEENYKAEYRRLKQMKEMGVNSIRTTHNPASEQTLQIAAELGLLVQEEAFDTWY 511

Query: 411 YAKNGNVNDFSNYFHQTIGTENANYLQRVRSPETSWAQYSIEAMVWSAKNDPSVLMWSIG 470
             K     D+  +F +      A   ++       W+ + +  MV   KN+P++ MWSIG
```

```
Sbjct: 512 GGK--KPYDYGRFFEKDATHPEARKGEK-------WSDFDLRTMVERGKNNPAIFMWSIG 562

Query: 471 NELMEGFSADVSHYPELTRQMCQWITAIDTSRPITFGDNKLKEADFC-WHEEVSQMATLL 529
           NE+  G +   +H     +++ + I  +D +R +T G +K +  +    HE+++
Sbjct: 563 NEI--GEANGDAHSLATVKRLVKVIKDVDKTRYVTMGADKFRFGNGSGGHEKIA------ 614

Query: 530 SQLDHPQGLIGLNYADGKDYDRLHEEHSDWLLYGSETVSAITSR-AYYKETKKVLDS--- 585
            +LD     +G NY++  +Y  L  +H  WL+YGSET SA  +R +YY+  +++  S
Sbjct: 615 DELD----AVGFNYSE-DNYKALRAKHPKWLIYGSETSSATRTRGSYYRPERELKHSNGP 669

Query: 586 --GYHLTSYDHAKVDWGAFASQAWYDTITRDFV--AGECVWTGFDYLGEPTPWNKIDSGV 641
             Y  + Y + +V WG  A+ +W  T  RD    AG+ +WTG DY+GEPTPW+  +
Sbjct: 670 ERNYEQSDYGNDRVGWGKTATASW--TFDRDNAGYAGQFIWTGTDYIGEPTPWHNQNQTP 727

Query: 642 VGLWPSPKNAYFGILDTAGFPKDSYYFYQSQW--AQGQTTLHLLPVWQKD-----QLCFD 694
           V      K++YFGI+DTAG PK  +Y YQSQW   + +  +HLLP W  +         D
Sbjct: 728 V------KSSYFGIVDTAGIPKHDFYLYQSQWVSVKKKPMVHLLPHWNWENKELASKVAD 781

Query: 695 EQGLVEVVVYSNAASVQLMFEDEQGNLTDYGRKAFHTYSTPTGHTYQLYQGADAAKNPHE 754
            +G + V  YSNA+SV+L       N    G K F+   T  G TYQ  +GA+A
Sbjct: 782 SEGKIPVRAYSNASSVELFL-----NGKSLGLKTFNKKQTSDGRTYQ--EGANA-----N 829

Query: 755 NLYLTWRVPYQKGLLRAVAYDISGKSIPKTSGRSQVRTYGSVAKLSWKAFEAPIDAPW-E 813
           LYL W+V YQ G L A+A D SGK I     R ++ T G  A +     +  I A   +
Sbjct: 830 ELYLEWKVAYQPGTLEAIARDESGKEI----ARDKITTAGKPAAVRLIKEDHAIAADGKD 885

Query: 814 LLYLDLSLLDSRGELVSHAQDWLQVQVEGPARLLALDNGNPTDHTPYQEP-----LRQAY 868
           L Y+   ++DS+G +V  A + ++ Q+ G  +L+ +DNG       Y+       +R+A+
Sbjct: 886 LTYIYYEIVDSQGNVVPTANNLVRFQLHGQGQLVGVDNGEQASRERYKAQADGSWIRKAF 945

Query: 869 GGKLLAILALTGEAGHIKVTA 889
            GK +AI+  T +AG   +TA
Sbjct: 946 NGKGVAIVKSTEQAGKFTLTA 966
```

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

```
Identities = 98/414 (23%), Positives = 175/414 (41%), Gaps = 64/414 (15%)

Query: 54  LPSDEVMVIPTSFNDLMVSKEKRDYIGDFWYEKVIEVPKVSEDEEMVLRFGSVTHQAKIY 113
           LP D  +  P + N    S  K   +G  WY   + + +V     + + F     + +IY
Sbjct: 86  LPHDFSLTQPYTRNGEAESAYKLGGVG--WYRHYLVLDEVLAGCHVAITFEGSYMETEIY 143

Query: 114 VDGVLVGEHKGGFTPFEVLVPECKYNNEKIKVSICANNVLDYTTLPVGNYSEIIQEDGSI 173
           V+G  +G+H  G+  F   + +          V+  A N+L                 +
Sbjct: 144 VNGQFIGKHLNGYQEFTYDISDV--------VTFGAENLLAVR----------------V 179

Query: 174 KKKVRENFDFFNYAGVHRPLKLMIRPKNHIFDITITSRLSDDL------QSADLHFLVET 227
           + KV  +  +++ +G++R + L + P+ H     +   L+D        Q  DL F +
Sbjct: 180 ENKVPSS-RWYSGSGLYREVSLSVLPQLHFVADQVAMTLADTAVQEKGQQKVDLRFALNQ 238

Query: 228 NQKVDEVRISVF-------DEDNKLVGETKDS-------------RLFLSDVHLWEVLNA 267
           + +    ++S+          +D KL+ +  +               L L ++ LW   N
Sbjct: 239 SIQTCHYQLSLCLWEQSHCSKDKKLLYQETEVPLADLAFQRQYGLTLSLEELQLWSPDNP 298

Query: 268 YLYTARVEIFVDNQLQDVYEENFGLREIE-VTNGQFLLNRKPIYFKGFGKHEDTFINGRG 326
           +LY   + ++   Q+ D +   G R++  + N   +N + +  KG   H D    G
Sbjct: 299 HLYDLELTLYYQGQVIDCFCLETGFRQLTFMANQGLFVNGRAVKLKGVCLHHDQGGLGAC 358

Query: 327 LNEAANLMDLNLLKDMGANSFRTSHYPYSEEMMRLADRMGVLVIDEVPAVGLFQ---NFN 383
            E A    L LLKDMGAN+ R++H P S ++ +LA+R+G  VI+E       +   N N
Sbjct: 359 AYEDALARQLVLLKDMGANTIRSTHNPSSPKLRQLANRLGFFVIEEAFDTWTYAKNGNVN 418

Query: 384 ASLDLSPKDNGTWN---LMQTKAAH----EQAIQELVKRDKNHPSVVMWVVANE 430
             +  GT N   L + ++      + +I+ +V   KN PSV+MW + NE
Sbjct: 419 DFSNYFHQTIGTENANYLQRVRSPETSWAQYSIEAMVWSAKNDPSVLMWSIGNE 472
```

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

Example 387

A DNA sequence (GBSx0418) was identified in *S.agalactiae* <SEQ ID 1261> which encodes the amino acid sequence <SEQ ID 1262>. This protein is predicted to be 2-keto-3-deoxygluconate kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.53   Transmembrane  197 - 213 ( 197 - 213)

----- Final Results -----
               bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9699> which encodes amino acid sequence <SEQ ID 9700> was also identified.

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

```
>GP:AAD35161 GB:AE001693 2-keto-3-deoxygluconate kinase [Thermotoga maritima]
 Identities = 115/342 (33%), Positives = 180/342 (52%), Gaps = 16/342 (4%)

Query: 14  KIISLGEVLLRLSPPQVHTLMQANHLKCQFGGSELNVLASLAQLGYHVGLVSALPDNDLG 73
           K+++ GE++LRLSPP +  + Q +      +GG+E NV A LAQ+G      V+ LP+N LG
Sbjct: 2   KVVTFGEIMLRLSPPDHKRIFQTDSFDVTYGGAEANVAAFLAQMGLDAYFVTKLPNNPLG 61

Query: 74  KMASQFILSQQISPAAIIKKEGRLGIYYYEQGFSVRTNKVIYDRNYSSFWESTLSDYDFT 133
             A+  +    +    I +   R+GIY+ E G S R +KV+YDR +S+  E+   D+D+
Sbjct: 62  DAAAGHLRKFGVKTDYIARGGNRIGIYFLEIGASQRPSKVVYDRAHSAISEAKREDFDWE 121

Query: 134 SIFKGVDWFHVSGITPALTKDLYEVTRFLMTKAKEGGVKVSIDLNFRESLWSSFQEAREQ 193
            I  G  WFH SGITP L K+L  +    + A E GV VS DLN+R   LW+  +EA++
Sbjct: 122 KILDGARWFHFSGITPPLGKELPLILEDALKVANEKGVTVSCDLNYRARLWTK-EEAQKV 180

Query: 194 LSPLLGLLDVCFGLEPIYLAGESEDLKDELGLSRPYLDI-------ELLEKITQKIVQEY 246
           + P +  +DV        L    ED++  LG+S   LD+       E   KI +++ ++Y
Sbjct: 181 MIPFMEYVDV--------LIANEEDIEKVLGISVEGLDLKTGKLNREAYAKIAEEVTRKY 232

Query: 247 GLDYIAFTQREMEYTNQYMLKSYLYHNNMLYQTDKTGVEVLDRVGTGDAFAAGLIHALLE 306
             + T RE             ++ N   + +++  + ++DRVG GD+FA  LI+  L
Sbjct: 233 NFKTVGITLRESISATVNYWSVMVFENGQPHFSNRYEIHIVDRVGAGDSFAGALIYGSLM 292

Query: 307 KETPQRALEIAMATFKYKHTIQGDINIMTRDDIAYLIEKETN 348
             Q+  E A A    KHTI GD  +++ ++I  L    T+
Sbjct: 293 GFDSQKKAEFAAAASCLKHTIPGDFVVLSIEEIEKLASGATS 334
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1263> which encodes the amino acid sequence <SEQ ID 1264>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0708(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 = 111/319 (34%), Positives = 168/319 (51%), Gaps = 7/319 (2%)
```

```
Query: 12  MAKIISLGEVLLRLSPPQYHTLMQANHLKCQFGGSELNVLASLAQLGYHVGLVSALPDND 71
           M+K++ +GE L+R+SP Q+  L  A   +  FGGSE+N+  +L   G    L +ALPDN
Sbjct: 14  MSKLLLVGEPLIRVSPNQFQPLTNACEAQLFFGGSEVNIARTLGGFGLEARLFTALPDNP 73

Query: 72  LGKMASQFILSQQISPAAIIKKEGRLGIYYYEQGFSVRTNKVIYDRNYSSFWESTLSDYD 131
           +G    QF+    +  +    + R+G+YY E GF  R ++V YDR  SSF        D
Sbjct: 74  VGHAFHQFLKQSGVDMSLTAWQGHRVGLYYLENGFGCRASQVYYDRCGSSFSALDKDSLD 133

Query: 132 FTSIFKGVDWFHVSGITPALTKDLYEVTRFLMTKAKEGGVKVSIDLNFRESLWSSFQEAR 191
             +IF+G+  FH SGI+ AL K   ++   L+ +AK+  + +S DLNFR S+  +  +A+
Sbjct: 134 LAAIFEGISHFHFSGISLALGKKTQDLIEVLVREAKKRDICISFDLNFRSSM-IAVADAK 192

Query: 192 EQLSPLLGLLDVCFGLEPIYLAGESEDLKDELGLSRPYLDIELLEKITQKIVQEYGLDYI 251
             S      D+ FG+EP+ L  +  D+ D      R   D   + +     + Q Y L  I
Sbjct: 193 RLFSHFAQYADIIFGMEPLLLDSDDFDMFD-----RKKADTTTIRERLAGLYQRYQLQAI 247

Query: 252 AFTQREMEYTNQYMLKSYLYHNNMLYQTDKTGVEVLDRVGTGDAFAAGLIHALLEKETPQ 311
             T+R  +       K+Y Y +   Y++ +    VL RVG+GDAF AGL++ LLE    Q
Sbjct: 248 YHTERSNDAQGSNHFKAYAY-DRQFYESCEVTTPVLQRVGSGDAFVAGLLYQLLEGNEKQ 306

Query: 312 RALEIAMATFKYKHTIQGD 330
           R L+ A+AT    K T+  D
Sbjct: 307 RNLDFAVATASLKCTVAED 325
```

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

Example 388

A DNA sequence (GBSx0419) was identified in *S.agalactiae* <SEQ ID 1265> which encodes the amino acid sequence <SEQ ID 1266>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.17   Transmembrane    5 -  21 (   5 -  21)

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

The protein has 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 389

A DNA sequence (GBSx0420) was identified in *S.agalactiae* <SEQ ID 1267> which encodes the amino acid sequence <SEQ ID 1268>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.05   Transmembrane  198 - 214 ( 191 - 220)
   INTEGRAL    Likelihood =-11.68   Transmembrane  446 - 462 ( 437 - 467)
   INTEGRAL    Likelihood = -9.55   Transmembrane   94 - 110 (  91 - 116)
   INTEGRAL    Likelihood = -7.43   Transmembrane  291 - 307 ( 283 - 309)
   INTEGRAL    Likelihood = -4.88   Transmembrane  265 - 281 ( 257 - 282)
   INTEGRAL    Likelihood = -4.62   Transmembrane  321 - 337 ( 318 - 339)
   INTEGRAL    Likelihood = -3.93   Transmembrane  406 - 422 ( 405 - 426)
   INTEGRAL    Likelihood = -1.59   Transmembrane  121 - 137 ( 121 - 137)
```

```
    INTEGRAL    Likelihood = -1.12   Transmembrane  345 - 361 ( 345 - 362)
    INTEGRAL    Likelihood = -0.48   Transmembrane   43 -  59 (  43 -  59)

----- Final Results -----
               bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB13641 GB:Z99113 similar to H+-symporter [Bacillus subtilis]
 Identities = 105/452 (23%), Positives = 182/452 (40%), Gaps = 37/452 (8%)

Query: 36  IYLFTFMFVTYFSTGVLGSAAIFVSQIMGYIRIFDGFIDPAIGIMIDKTDTKFGKYRPIL 95
           IY     ++ +F T V G +A     +   +RI D   DP IG ++D+T+++F ++RP L
Sbjct: 27  IYATVSTYLLFFYTDVFGLSAAAAGTMFLVVRIIDALADPFIGTIVDRTNSRFARFRPYL 86

Query: 96  IIGNVITALSLIFLLALRGVDENIRFPLFILVLIIHKIGYSMQQTITKAGQTALTNDPKQ 155
           + G    A   + L  L     +      ++   I  +G S+  T       ALT+
Sbjct: 87  LFG----AFPFVILAILCFTTPDFSDMGKLIYAYITYVGLSLTYTTINVPYGALTS-AMT 141

Query: 156 RPIFNIVDAVMTTSLMTGGQFVVSVFLVPKFGNFTPQFFNVLIFGTILISAILAIV--AI 213
           R    +V       L      +V  F VP    +          G  L   IL ++   +
Sbjct: 142 RNNQEVVSITSVRMLFANLGGLVVAFFVPLLAAYLSDTSGNESLGWQLTMGILGMIGGCL 201

Query: 214 IGIWAKDRKEFFGLGENTQKTALKDYWKVLKGNKPLQILSIAAALVKFAIQFFGDSV-VM 272
           +    K  KE   L ++ +K    D ++  + N+PL +LSI    ++ F +    +SV +
Sbjct: 202 LIFCFKSTKERVTLQKSEEKIKFTDIFEQFRVNRPLVVLSIFFIII-FGVNSISNSVGIY 260

Query: 273 VLLFGI----LFGNYALSGQFSLLFIVPGVIINILFSTIARKKGLRFSYVRAIQIGMIGL 328
            + + +    L   Y L G    L I+P   I  L   + +KK L +    A+ + +IGL
Sbjct: 261 YVTYNLEREDLVKWYGLIGSLPALVILP--FIPRLHQFLGKKKLLNY----ALLLNIIGL 314

Query: 329 LAFGAVLYVGKPGDLSLTSLNLYTILFIVTNIIARYASQAPASLVLTMGADISDYETSES 388
           LA            L +   N+Y IL  V  +IA   S      +  +  +  +Y    +
Sbjct: 315 LAL-----------LFVPPSNVYLIL--VCRLIAAAGSLTAGGYMWALIPETIEYGEYRT 361

Query: 389 GRYVSGMIGTIFSLTDSIASSFAPMVVGFVLAGIGFSKSFPTIETPLPPDLKMAAISILV 448
           G+ + G+I  I          +   +V G VL   G+             P   M  +
Sbjct: 362 GKRMGGLIYAIIGFFFKFGMALGGVVPGLVLDKFGY-----VANQAQTPAALMGILITTT 416

Query: 449 AIPFIALSIALLLMKFYKLDKEEMVRIQEKIQ 480
            IP   L +AL+ + FY LD+++   +  +++
Sbjct: 417 IIPVFLLVLALIDINFYNLDEKKYKNMVRELE 448
```

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 390

A DNA sequence (GBSx0422) was identified in *S.agalactiae* <SEQ ID 1269> which encodes the amino acid sequence <SEQ ID 1270>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3375(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:AAB17663 GB:U31175 D-specific D-2-hydroxyacid dehydrogenase [S. aureus]
 Identities = 165/331 (49%), Positives = 231/331 (68%), Gaps = 1/331 (0%)

Query: 1   MMKLKVFNVREEEATLAQDWANRNHVELSMSEGPLTLETVNEVEGFDGIANAQIEPLDDA 60
            M K+  F  R+ E  +A +W  +N+VE++ S+  L+  TV++++ +DG+   Q   L++
Sbjct: 1   MTKIMFFGTRDYEKEMALNWGKKNNVEVTTSKELLSSATVDQLKDYDGVTTMQFGKLEND 60

Query: 61  IYPLLKEMGIKQIAQRSAGVDMYNLELAKQHGIIISNVPSYSPESIAEFTVTIALNLIRK 120
            +YP L+  GIKQIAQR+AG DMY+L+LAK+H I+ISNVPSYSPE+IAE++V+IAL L+R+
Sbjct: 61  VYPKLESYGIKQIAQRTAGFDMYDLDLAKKHNIVISNVPSYSPETIAEYSVSIALQLVRR 120

Query: 121 VELIRANVREQNFSWTLPIRGRVLGNMTVAIIGTGRIGLATAKIFKGFGCRVIGYDIYHN 180
               I   V+  +F+W   I  + + NMTVAIIGTGRIG ATAKI+ GFG  +  YD Y N
Sbjct: 121 FPDIERRVQAHDFTWQAEIMSKPVKNMTVAIIGTGRIGAATAKIYAGFGATITAYDAYPN 180

Query: 181 PMADGILEYVNSVEEAVEEADLVSLHMPPTAENTHLFNLDMFKQFKKGAILMNMARGALV 240
               D  L Y +SV+EA+++AD++SLH+P   E+ HLF+  MF   KKGAIL+N ARGA++
Sbjct: 181 KDLD-FLTYKDSVKEAIKDADIISLHVPANKESYHLFDKAMFDHVKKGAILVNAARGAVI 239

Query: 241 ETKDLLEALDQGLLEGAGIDTYEFEGPYIPKNCQGQDISDKDFLRLINHPKVIYTPHAAY 300
             T DL+ A++ G L GA IDTYE E  Y   +   +DI DK  L LI H +++ TPH A+
Sbjct: 240 NTPDLIAAVNDGTLLGAAIDTYENEAAYFTNDWTNKDIDDKTLLELIEHERILVTPHIAF 299

Query: 301 YTDEAVKNLVEGALNACVEVIETGTTTTKVN 331
           ++DEAV+NLVEG LNA + VI TGT  T++N
Sbjct: 300 FSDEAVQNLVEGGLNAALSVINTGTCETRLN 330
```

There is also homology to SEQ ID 124.

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

Example 391

A DNA sequence (GBSx0423) was identified in *S.agalactiae* <SEQ ID 1271> which encodes the amino acid sequence <SEQ ID 1272>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2364(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 392

A DNA sequence (GBSx0424) was identified in *S.agalactiae* <SEQ ID 1273> which encodes the amino acid sequence <SEQ ID 1274>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.90   Transmembrane   64 -  80 (  53 -  89)

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

```
                bacterial membrane --- Certainty=0.6158(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 9325> which encodes amino acid sequence <SEQ ID 9326> was also identified.

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

```
>GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema
          pallidum]
 Identities = 33/91 (36%), Positives = 55/91 (60%), Gaps = 1/91 (1%)

Query: 1   MANVLAKPKIMLPMISSAAILGILGALFNIQGTPASAGFGISGLIGPINALNLAKGGWSV 60
           M N +  P + +P++ +  + G+L  LFN+QGTPASAGFG  GL+GPINA  L      V
Sbjct: 250 MPNWIRYPILNIPLLLNGLVCGVLAWLFNLQGTPASAGFGFIGLVGPINAYRLMAYTPMV 309

Query: 61  MNMLLIIIIFVAAPIILNFIFNYLFIKVLKI 91
              +L ++ FV +  +  ++ +++ +  LK+
Sbjct: 310 RAGILFLVYFVLS-FLAAYLIDFILVDRLKL 339
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1275> which encodes the amino acid sequence <SEQ ID 1276>. Analysis of this protein sequence reveals the following:

```
     Possible site: 51
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.31   Transmembrane  141 - 157 ( 133 - 166)
   INTEGRAL    Likelihood = -6.00   Transmembrane   92 - 108 (  88 - 112)

----- Final Results -----
               bacterial membrane --- Certainty=0.5925(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:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema
          pallidum]
 Identities = 63/178 (35%), Positives = 107/178 (59%), Gaps = 10/178 (5%)

Query: 2   IGQGIASLLGLQPILMSLLIAMIFCFLIVSPITTVGIALAINLSGIGSGAASFG------ 55
           +G+ IA+ + LQP+LMS+L++M F  +I+SP+++V + +A+ L+G+ SGAA+ G
Sbjct: 164 VGRVIATFIALQPLLMSILLSMSFSLIIISPVSSVAVGIAVGLTGLASGAANIGVSSCAM 223

Query: 56  -LCLAGWAVNSKGTSLAHVLRSPKISMANVLSKPKIMLPMLCSAAVLGVIGAIFNIQGTP 114
            L +   VN  G  LA    + K+ M N +  P + +P+L +  V GV+  +FN+QGTP
Sbjct: 224 TLIVGTMRVNKIGVPLAMFAGAMKMLMPNWIRYPILNIPLLLNGLVCGVLAWLFNLQGTP 283

Query: 115 ASAGFGISGLIGPINALNLAKGGWCP-VNILLIIIIFVGAPIVLNMIFNYLFIKVLKV 171
           ASAGFG  GL+GPINA  L    + P V   ++ +++     +   + +++ +  LK+
Sbjct: 284 ASAGFGFIGLVGPINAYRLM--AYTPMVRAGILFLVYFVLSFLAAYLIDFILVDRLKL 339
```

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

```
 Identities = 86/101 (85%), Positives = 96/101 (94%)

Query: 1   MANVLAKPKIMLPMISSAAILGILGALFNIQGTPASAGFGISGLIGPINALNLAKGGWSV 60
           MANVL+KPKIMLPM+ SAA+LG++GA+FNIQGTPASAGFGISGLIGPINALNLAKGGW
Sbjct: 81  MANVLSKPKIMLPMLCSAAVLGVIGAIFNIQGTPASAGFGISGLIGPINALNLAKGGWCP 140

Query: 61  MNMLLIIIIFVAAPIILNFIFNYLFIKVLKIIDPMDYKLDI 101
           +N+LLIIIIFV API+LN IFNYLFIKVLK+IDPMDYKLDI
Sbjct: 141 VNILLIIIIFVGAPIVLNMIFNYLFIKVLKVIDPMDYKLDI 181
```

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

Example 393

A DNA sequence (GBSx0426) was identified in *S.agalactiae* <SEQ ID 1277> which encodes the amino acid sequence <SEQ ID 1278>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.58   Transmembrane  148 - 164 ( 145 - 169)
   INTEGRAL    Likelihood = -5.26   Transmembrane   33 -  49 (  25 -  52)
   INTEGRAL    Likelihood = -4.73   Transmembrane   70 -  86 (  62 -  88)
   INTEGRAL    Likelihood = -3.45   Transmembrane  124 - 140 ( 122 - 143)
   INTEGRAL    Likelihood = -1.33   Transmembrane   96 - 112 (  96 - 112)

----- Final Results -----
              bacterial membrane --- Certainty=0.3633(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 9735> which encodes amino acid sequence <SEQ ID 9736> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9697> which encodes amino acid sequence <SEQ ID 9698> was also identified.

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

```
>GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema
          pallidum]
 Identities = 61/158 (38%), Positives = 92/158 (57%)

Query: 24  KSFIMNVLNGLALGTVIVLIPGAILGELMKALLPMWSGFATLIAATAVATSMMGLVIGIM 83
           + F+M +LNG + G VI L+P AI GEL +AL P+    FA L           +  +IG +
Sbjct: 9   RQFMMKILNGSSAGIVIGLVPPAIAGELFRALAPLSPLFAALYHVVLPIQFSVPALIGTL 68

Query: 84  VGLNFKFNPIQSASLGLAVMFAGGAATFLKGAIMLKGTGDIINMGITAALGVLLIQFLSD 143
           VGL F   +  A+L     + A G   T   GA ++ G GD+IN+ + +AL ++L++ L
Sbjct: 69  VGLQFHCSAPEVATLAFVSVIASGNVTLQNGAWLITGIGDVINVMLISALAIILVRALRG 128

Query: 144 KTKSFTLIVIPTVTLLLVGGVGHVLLPYVKMITTMIGQ 181
           K  S T+I +P +  ++ GGVG   LPYVKMIT  +G+
Sbjct: 129 KLGSLTIIALPVIVAVVAGGVGSFSLPYVKMITLFVGR 166
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1279> which encodes the amino acid sequence <SEQ ID 1280>. Analysis of this protein sequence reveals the following:

```
     Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.06   Transmembrane  314 - 330 ( 301 - 335)
   INTEGRAL    Likelihood =-11.30   Transmembrane  185 - 201 ( 178 - 215)
   INTEGRAL    Likelihood = -8.01   Transmembrane   22 -  38 (  11 -  42)
   INTEGRAL    Likelihood = -3.29   Transmembrane  266 - 282 ( 265 - 285)
   INTEGRAL    Likelihood = -2.66   Transmembrane  141 - 157 ( 141 - 159)
   INTEGRAL    Likelihood = -2.13   Transmembrane   53 -  69 (  53 -  69)
   INTEGRAL    Likelihood = -1.33   Transmembrane  114 - 130 ( 113 - 131)
   INTEGRAL    Likelihood = -0.80   Transmembrane  206 - 222 ( 206 - 222)

----- Final Results -----
              bacterial membrane --- Certainty=0.6222(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:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema
         pallidum]
 Identities = 137/346 (39%), Positives = 217/346 (62%), Gaps = 14/346 (4%)

Query: 12  FMNKVLAGTAIAIVVALIPNAILATFLKPLLP-NMAAAEFLHIVQVFQFFTPIMAGFLIG 70
           FM K+L G++  IV+ L+P AI     + L P +   A   H+V   QF  P + G L+G
Sbjct: 11  FMMKILNGSSAGIVIGLVPPAIAGELFRALAPLSPLFAALYHVVLPIQFSVPALIGTLVG 70

Query: 71  QQFKFNPMQQLAVGGAAYIGSGAWAYTEVIQKGVATGTFQLRGIGDLINMMITASLAVLA 130
            QF  +  +   + I SG           +  G + + GIGD+IN+M+ ++LA++
Sbjct: 71  LQFHCSAPEVATLAFVSVIASG--------NVTLQNGAWLITGIGDVINVMLISALAIIL 122

Query: 131 VKYFGNKFGSLTIILLPITIGTGVGYIGWKFLPYVSYVTTLIGQGINSFTTLQPILMSIL 190
           V+    K GSLTII LP+ +    G +G   LPYV  +T  +G+ I +F  LQP+LMSIL
Sbjct: 123 VRALRGKLGSLTIIALPVIVAVVAGGVGSFSLPYVKMITLFVGRVIATFIALQPLLMSIL 182

Query: 191 IAVAFSLIIVSPISTVAIGLAIGLNGMAAGAASMGIASTAAVLVWATLKVNKSGVPIAIA 250
           ++++FSLII+SP+S+VA+G+A+GL G+A+GAA++G++S A  L+  T++VNK GVP+A+
Sbjct: 183 LSMSFSLIIISPVSSVAVGIAVGLTGLASGAANIGVSSCAMTLIVGTMRVNKIGVPLAMF 242

Query: 251 LGAMKMMMPNFLKHPIMAIPMVFTAAISSLTVPLFNLVGTPASSGFGLVGAVGPIAS--L 308
            GAMKM+MPN++++PI+ IP++    +  +   LFNL GTPAS+GFG +G VGPI +  L
Sbjct: 243 AGAMKMLMPNWIRYPILNIPLLLNGLVCGVLAWLFNLQGTPASAGFGFIGLVGPINAYRL 302

Query: 309 AGGSSIL---IIILAWIIVPFAVAFAAHKVSKDILKLYKEDIFVFE 351
              + ++   I+ L + ++ F  A+    + D LKLY+ ++F+ E
Sbjct: 303 MAYTPMVRAGILFLVYFVLSFLAAYLIDFILVDRLKLYRRELFIPE 348
```

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

```
 Identities = 65/172 (37%), Positives = 95/172 (54%), Gaps = 9/172 (5%)

Query: 19  EKQTTKSFIMNVLNGLALGTVIVLIPGAILGELMKALLPMWSGFATLIAATAVATSMMGL 78
           +K+T  SF+  VL G A+  V+ LIP AIL   +K LLP  +  A  +    V      +
Sbjct: 5   DKETFSSFMNKVLAGTAIAIVVALIPNAILATFLKPLLPNMAA-AEFLHIVQVFQFFTPI 63

Query: 79  VIGIMVGLNFKFNPIQSASLGLAVMFAGGAATFLK--------GAIMLKGTGDIINMGIT 130
           + G ++G  FKFNP+Q  ++G A    GA  + +         G   L+G GD+INM IT
Sbjct: 64  MAGFLIGQQFKFNPMQQLAVGGAAYIGSGAWAYTEVIQKGVATGTFQLRGIGDLINMMIT 123

Query: 131 AALGVLLIQFLSDKTKSFTLIVIPTVTLLLVGGVGHVLLPYVKMITTMIGQG 182
           A+L VL +++  +K  S T+I++P      VG +G   LPYV  +TT+IGQG
Sbjct: 124 ASLAVLAVKYFGNKFGSLTIILLPITIGTGVGYIGWKFLPYVSYVTTLIGQG 175
```

A related GBS gene <SEQ ID 8567> and protein <SEQ ID 8568> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:    -13.49
GvH: Signal Score (-7.5): -5.82
     Possible site: 48
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 5 value:  -6.58 threshold:  0.0
   INTEGRAL    Likelihood = -6.58   Transmembrane  148 - 164 ( 145 - 169)
   INTEGRAL    Likelihood = -5.26   Transmembrane   33 -  49 (  25 -  52)
   INTEGRAL    Likelihood = -4.73   Transmembrane   70 -  86 (  62 -  88)
   INTEGRAL    Likelihood = -3.45   Transmembrane  124 - 140 ( 122 - 143)
   INTEGRAL    Likelihood = -1.33   Transmembrane   96 - 112 (  96 - 112)
  PERIPHERAL   Likelihood =  1.85       51
modified ALOM score:   1.82

*** Reasoning Step: 3
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.3633(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:

```
ORF01226(352 - 843 of 1218)
EGAD|138195|TP0038(3 - 166 of 350) regulatory protein (Treponema pallidum) OMNI|TP0038
regulatory protein (pfoS/R) GP|3322295|gb|AAC65034.1||AE001189 regulatory protein (pfoS/R)
(Treponema pallidum) PIR|E71373|E71373 probable regulatory protein (pfoS/R) - syphilis
spirochete
%Match = 13.6
%Identity = 37.2  %Similarity = 59.1
Matches = 61  Mismatches = 67  Conservative Sub.s = 36

273       303       333       363       393       423       453       483
I*FFPIFLLQIAMI*LI*LVKSQTIIISRRHLMSDVVEKQTTKSFIMNVLNGLALGTVIVLIPGAILGELMKALLPMWSG
                                          :  : : |:| :||| : | || |:| || ||| :|| |:
                                        MHTQSLSPRQFMMKILNGSSAGIVIGLVPPAIAGELFRALAPLSPL
20                                                10        20        30        40

513       543       573       603       633       663       693       723
FATLIAATAVATSMMGLVIGIMVGLNFKFNPIQSASLGLAVMFAGGAATFLKGAIMLKGTGDIINMGITAALGVLLIQFL
|| |          :  :|| :||| |  :  : |:| :  : | |  |:  || :: | ||:||: : :|| ::|:: |
FAALYHVVLPIQFSVPALIGTLVGLQFHCSAPEVATLAFVSVIASGNVTLQNGAWLITGIGDVINVMLISALAIILVRAL
             60        70        80        90        100       110       120

753       783       813       843       873       903       933       963
SDKTKSFTLIVIPTVTLLLVGGVGHVLLPYVKMITTMIGQGTRRTHENFLFILLCPDINFEKIPF*INDLLSLFLQIIGL
  |  |:|:| :| :  :: ||||   |||||||| :|:
RGKLGSSLTIIALPVIVAVVAGGVGSFSLPYVKMITLFVGRVIATFIALQPLLMSILLSMSFSLIIISPVSSVAVGIAVGL
            140       150       160       170       180       190       200
```

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

Example 394

A DNA sequence (GBSx0428) was identified in *S.agalactiae* <SEQ ID 1281> which encodes the amino acid sequence <SEQ ID 1282>. This protein is predicted to be cyn operon transcriptional activator. Analysis of this protein sequence reveals the following:

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

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

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

```
>GP:CAB15857 GB:Z99123 alternate gene name: ipa-24d~similar to
          transcriptional regulator (LysR family) [Bacillus subtilis]
 Identities = 87/282 (30%), Positives = 152/282 (53%), Gaps = 5/282 (1%)

Query: 1   MDIRQLTYFIAVAEAKNYSRAAKSLFVTQPTLSQSIKKLEAELNTILFLQNGRQLALTEA 60
           MDIR LTYF+ VA  K++++A++SL+V+QPT+S+ IK LE EL   LF +NGRQ+ LT+A
Sbjct: 1   MDIRHLTYFLEVARLKSFTKASQSLYVSQPTISKMIKNLEEELGIELFYRNGRQVELTDA 60

Query: 61  GEILYEKGQLLMTNVNQMVTEIQQLNQEKKEGIRVGLTSLFAIQFMKQI-STFMATHSNV 119
           G  +Y + Q ++ +   + +E+  + + KK  +R+GL  +    F  ++    F  + NV
Sbjct: 61  GHSMYVQAQEIIKSFQNLTSELNDIMEVKKGHVRIGLPPMIGSGFFPRVLGDFRENYPNV 120
```

```
Query: 120 EVSLIQDGSRKLQELLAKGKIDIGLLSFPSTRNDITIEPLQTSTKGYKVSIVMPKSHPLA 179
           L++DGS K+QE +  G +DIG++  P+  +      +   T    + +V+  SH LA
Sbjct: 121 TFQLVEDGSIKVQEGVGDGSLDIGVVVLPANEDIFHSFTIVKET----LMLVVHPSHRLA 176

Query: 180 TLPEIELNDLRDYKVASLNEHYMLGEMLPRKCRALGFDPHIVFKHNDWEVLIHSLQDLNA 239
             E +L +L+D       E ++L   +  +C   GF PHI+++ + W+ +   +
Sbjct: 177 DEKECQLRELKDEPFIFFREDFVLHNRIMTECIKAGFRPHIIYETSQWDFISEMVSANLG 236

Query: 240 VTILPSEFESISQVQDLCWVPLKDKNNFYPIGIAYRNDTSFS 281
           + +LP         + +  +PL D    + + I +R D   S
Sbjct: 237 IGLLPERICRGLDPEKVKVIPLVDPVIPWHLAIIWRKDRYLS 278
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1283> which encodes the amino acid sequence <SEQ ID 1284>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1101(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 = 125/160 (78%), Positives = 144/160 (89%)

Query: 135 LAKGKIDIGLLSFPSTRNDITIEPLQTSTKGYKVSIVMPKSHPLATLPEIELNDLRDYKV 194
           L++GKIDIGLLSF S R DITIE LQTSTKGYKVSIV+ K HPLA  P+++L DL+ YK+
Sbjct: 1   LSQGKIDIGLLSFLSIRKDITIELLQTSTKGYKVSIVLLKQHPLAQHPQLKLKDLKGYKI 60

Query: 195 ASLNEHYMLGEMLPRKCRALGFDPHIVFKHNDWEVLIHSLQDLNAVTILPSEFESISQVQ 254
           ASLN+HYMLGEMLPRKCRALGF+P IVFKHNDWEVLIHSL DLN +TILPS+FES++QV
Sbjct: 61  ASLNDHYMLGEMLPRKCRALGFEPDIVFKHNDWEVLIHSLHDLNTLTILPSDFESLNQVD 120

Query: 255 DLCWVPLKDKNNFYPIGIAYRNDTSFSPMIEEFLSLLKTN 294
           +L W+PL+DKNNFYPIGIAYR+D SFSP+IEEFLSLLKTN
Sbjct: 121 NLVWIPLQDKNNFYPIGIAYRDDASFSPVIEEFLSLLKTN 160
```

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

Example 395

A DNA sequence (GBSx0429) was identified in *S.agalactiae* <SEQ ID 1285> which encodes the amino acid sequence <SEQ ID 1286>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1833(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

Signal peptide: 1-21
```

A related GBS nucleic acid sequence <SEQ ID 8569> which encodes amino acid sequence <SEQ ID 8570> 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 8570 (GBS271) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 8; MW 31.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 6; MW 56.3kDa) and in Figure 62 (lane 10; MW 56.3kDa).

GBS271-GST was purified as shown in Figure 210, lane 8.

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

Example 396

A DNA sequence (GBSx0430) was identified in *S.agalactiae* <SEQ ID 1287> which encodes the amino acid sequence <SEQ ID 1288>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.74   Transmembrane    9 -  25 (   5 -  28)
   INTEGRAL    Likelihood = -5.84   Transmembrane   97 - 113 (  92 - 122)
   INTEGRAL    Likelihood = -5.47   Transmembrane   37 -  53 (  35 -  61)
   INTEGRAL    Likelihood = -2.55   Transmembrane  220 - 236 ( 220 - 238)
   INTEGRAL    Likelihood = -1.65   Transmembrane   64 -  80 (  63 -  81)
   INTEGRAL    Likelihood = -1.28   Transmembrane  193 - 209 ( 192 - 209)
   INTEGRAL    Likelihood = -0.53   Transmembrane  125 - 141 ( 125 - 141)

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

```
>GP:AAC73593 GB:AE000155 putative metal resistance protein
          [Escherichia coli K12]
 Identities = 128/252 (50%), Positives = 186/252 (73%)

Query: 5   NSISLMSLLMASSLVLITLFFSYWQKLNLEKEVIISAIRAVIQLLAVGFLLDYIFGYQNP 64
           ++I+  SL +A  LV++ +  S+ +KL LEK+++ S  RA+IQL+ VG++L YIF   +
Sbjct: 13  HNITNESLALALMLVVVAILISHKEKLALEKDILWSVGRAIIQLIIVGYVLKYIFSVDDA 72

Query: 65  IFTALLMLFMIINASYNAAKRGKGINKGFVISFIAIGSGTIITLSVLIFSGILKFVPNQM 124
            T L++LF+  NA++NA KR K I K F+ SFIAI  G  ITL+VLI SG ++F+P Q+
Sbjct: 73  SLTLLMVLFICFNAAWNAQKRSKYIAKAFISSFIAITVGAGITLAVLILSGSIEFIPMQV 132

Query: 125 IPVGGMIISNSMVAIGLCYKQLLSEFRSKQEEVETKLALGADILPASIDIIRDVIKTGMV 184
           IP+ GMI  N+MVA+GLCY  L     S+Q++++ KL+LGA    AS  +IRD I+  ++
Sbjct: 133 IPIAGMIAGNAMVAVGLCYNNLGQRVISEQQQIQEKLSLGATPKQASAILIRDSIRAALI 192

Query: 185 PTIDSAKTLGIVSLPGMMTGLILAGTSPIQAVKYQMMVTFMLLATTSIASFVATYLAYKI 244
           PT+DSAKT+G+VSLPGMM+GLI AG  P++A+KYQ+MVTFMLL+T S+++ +A YL Y+
Sbjct: 193 PTVDSAKTVGLVSLPGMMSGLIFAGIDPVKAIKYQIMVTFMLLSTASLSTIIACYLTYRK 252

Query: 245 FFNNRKQLVVTK 256
           F+N+R QLVVT+
Sbjct: 253 FYNSRHQLVVTQ 264
```

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 397

A DNA sequence (GBSx0431) was identified in *S.agalactiae* <SEQ ID 1289> which encodes the amino acid sequence <SEQ ID 1290>. This protein is predicted to be SUGAR TRANSPORT ATP-BINDING PROTEIN. (b0490). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1903(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:AAC73592 GB:AE000155 putative ATP-binding component of a
          transport system [Escherichia coli K12]
 Identities = 95/202 (47%), Positives = 142/202 (70%), Gaps = 2/202 (0%)

Query: 4   LTFKHVDFKTDDKLVLNDINFAIDEGDFVSIVGPSGSGKSTVLKLASGLMSPTAGHIFFD 63
           L  ++V +   D  +LN+INF++  G+F  I GPSG GKST+LK+ + L+SPT+G + F+
Sbjct: 8   LQLQNVGYLAGDAKILNNINFSLRAGEFKLITGPSGCGKSTLLKIVASLISPTSGTLLFE 67

Query: 64  GKDLNQLEPIESRKMISYCFQTPHLFGNTVEDNISFPYHIRHEKVDYRRVDDLFQRFEMD 123
           G+D++ L+P   R+ +SYC QTP LFG+TV DN+ FP+ IR+ + D     D  +RF +
Sbjct: 68  GEDVSTLKPEIYRQQVSYCAQTPTLFGDTVYDNLIFPWQIRNRQPDPAIFLDFLERFALP 127

Query: 124 QSYLKQDVKKLSGGEKQRIALIRQLLFEPKVLLLDEVTSALDNHNKAIVEKVI-KSLHDK 182
            S L +++ +LSGGEKQRI+LIR L F PKVLLLDE+TSALD  NK  V ++I + + ++
Sbjct: 128 DSILTKNIAELSGGEKQRISLIRNLQFMPKVLLLDEITSALDESNKHNVNEMIHRYVREQ 187

Query: 183 GITILWITHDEEQSRRFANKVL 204
           I +LW+THD+++    A+KV+
Sbjct: 188 NIAVLWVTHDKDEINH-ADKVI 208
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1291> which encodes the amino acid sequence <SEQ ID 1292>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2053(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 = 73/214 (34%), Positives = 133/214 (62%), Gaps = 9/214 (4%)

Query: 4   LTFKHVD--FKTDDKLVLNDINFAIDEGDFVSIVGPSGSGKSTVLKLASGLMSPTAGHIF 61
           +TF +V   F+     VL +INF ++EG F +++G SGSGKST+L + +GL+  ++G I+
Sbjct: 6   ITFNNVSKTFEDSGTQVLKNINFDLEEGKFYTLLGASGSGKSTILNIMAGLLDASSGDIY 65

Query: 62  FDGKDLNQLEPIESRKMISYCFQTPHLFGN-TVEDNISFPYHIR--HEKVDYRRVDDLFQ 118
            DG+ +N L PI  R  I   FQ   LF + TV +N++F   ++   +K   +RV +  +
Sbjct: 66  LDGERINDL-PINKRD-IHTVFQNYALFPHMTVFENVAFALKLKKVDKKEIAKRVKETLK 123

Query: 119 RFEMDQSYLKQDVKKLSGGEKQRIALIRQLLFEPKVLLLDEVTSALDNHNKAIVEKVIKS 178
             ++ + +  + ++KLSGG++QR+A+ R ++ +P+V+LLDE  SALD   +  ++  ++
Sbjct: 124 MVQL-EGFENRSIQKLSGGQRQRVAIARAIINQPRVVLLDEPLSALDLKLRTEMQYELRE 182

Query: 179 LHDK-GITILWITHDEEQSRRFANKVLKVVNGSI 211
           L  + GIT +++THD+E++   ++ +  + G I
Sbjct: 183 LQQRLGITFVFVTHDQEEALAMSDWIFVMNEGEI 216
```

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

Example 398

A DNA sequence (GBSx0432) was identified in *S.agalactiae* <SEQ ID 1293> which encodes the amino acid sequence <SEQ ID 1294>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0658(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 399

A DNA sequence (GBSx0434) was identified in *S.agalactiae* <SEQ ID 1295> which encodes the amino acid sequence <SEQ ID 1296>. This protein is predicted to be deda protein (dedA). Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.05   Transmembrane  186 - 202 ( 178 - 208)
   INTEGRAL    Likelihood = -8.81   Transmembrane   65 -  81 (  61 -  89)
   INTEGRAL    Likelihood = -7.54   Transmembrane   26 -  42 (  24 -  47)
   INTEGRAL    Likelihood = -0.37   Transmembrane  152 - 168 ( 152 - 168)

----- Final Results -----
             bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAC75377 GB:AE000320 orf, hypothetical protein [Escherichia coli K12]
 Identities = 91/211 (43%), Positives = 131/211 (61%), Gaps = 7/211 (3%)

Query: 2   FLIDFILHIDTHIYAMANTVGNWTYLLLFLVIFVETGAVIFPFLPGDSLLFAAGALAANP 61
           FLIDFILHID H+  +    G W Y +LFL++F ETG V+ PFLPGDSLLF AGALA+
Sbjct: 6   FLIDFILHIDVHLAELVAEYGVWVYAILFLILFCETGLVVTPFLPGDSLLFVAGALASLE 65

Query: 62  KMSFNIVTFLIIFFIAAFIGDSCNFLIGRTFGYRFIKHP---FFRRFIKEKNIRDAELYF 118
              N+   +++  IAA +GD+ N+ IGR FG +   +P    FRR   +K       ++
Sbjct: 66  TNDLNVHMMVVLMLIAAIVGDAVNYTIGRLFGEKLFSNPNSKIFRRSYLDK----THQFY 121

Query: 119 EKKGTAAIILGRYIPIIRTFVPFVAGISQLPPKVFIKRAFIAALSWSVIATGSGFLFGNI 178
           EK G   IIL R++PI+RTF PFVAG+  +  + F     I AL W ++ T +G+ FG I
Sbjct: 122 EKHGGKTIILARFVPIVRTFAPFVAGMGHMSYRHFAAYNVIGALLWVLLFTYAGYFFGTI 181

Query: 179 PFVKQHFSLIILGIVFVTLIPVLISGVKSYR 209
           P V+ +  L+I+GI+ V+++P +I  ++  R
Sbjct: 182 PMVQDNLKLLIVGIIVVSILPGVIEIIRHKR 212
```

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 400

A DNA sequence (GBSx0435) was identified in *S.agalactiae* <SEQ ID 1297> which encodes the amino acid sequence <SEQ ID 1298>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3100(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 401

A DNA sequence (GBSx0436) was identified in *S.agalactiae* <SEQ ID 1299> which encodes the amino acid sequence <SEQ ID 1300>. This protein is predicted to be DNA-entry nuclease. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3990(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 9323> which encodes amino acid sequence <SEQ ID 9324> was also identified.

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

```
>GP:CAA38134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]
 Identities = 87/157 (55%), Positives = 110/157 (69%), Gaps = 1/157 (0%)

Query: 1   MLDRTIRQYQNRRDTTLPDANWKPLGWHQVAT-NDHYGHAVDKGHLIAYALAGNFKGWDA 59
           +L +  RQY+NR++T      +W P GWHQV       Y HAVD+GHL+ YAL G   G+DA
Sbjct: 116 LLSKATRQYKNRKETGNGSTSWTPPGWHQVKNLKGSYTHAVDRGHLLGYALIGGLDGFDA 175

Query: 60  SVSNPQNVVTQTAHSNQSNQKINRGQNYYESLVRKAVDQNKRVRYRVTPLYRNDTDLVPF 119
           S SNP+N+  QTA +NQ+  + + GQNYYES VRKA+DQNKRVRYRVT  Y ++ DLVP
Sbjct: 176 STSNPKNIAVQTAWANQAQAEYSTGQNYYESKVRKALDQNKRVRYRVTLYYASNEDLVPS 235

Query: 120 AMHLEAKSQDGTLEFNVAIPNTQASYTMDYATGEITL 156
           A  +EAKS DG LEFNV +PN Q     +DY TGE+T+
Sbjct: 236 ASQIEAKSSDGELEFNVLVPNVQKGLQLDYRTGEVTV 272
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1301> which encodes the amino acid sequence <SEQ ID 1302>. Analysis of this protein sequence reveals the following:

```
     Possible site: 42
>>> Seems to have a cleavable N-term signal seq.

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

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

```
>GP:CAA38134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]
 Identities = 89/135 (65%), Positives = 104/135 (76%), Gaps = 1/135 (0%)

Query: 25  SPAGWHRLHHLKGSYDHAVDRGHLLGYALVGGLKGFDASTGNPDNIATQLSWANQANKPY 84
           +P GWH++ +LKGSY HAVDRGHLLGYAL+GGL GFDAST NP NIA Q +WANQA   Y
Sbjct: 138 TPPGWHQVKNLKGSYTHAVDRGHLLGYALIGGLDGFDASTSNPKNIAVQTAWANQAQAEY 197

Query: 85  LTGQNYYEGLVRRALDKGHRVRYRVTLLY-DGDNLLASGSHLEAKSSDDSLTFNVFVPNV 143
            TGQNYYE  VR+ALD+  RVRYRVTL Y   ++L+ S S +EAKSSD  L FNV VPNV
Sbjct: 198 STGQNYYESKVRKALDQNKRVRYRVTLYYASNEDLVPSASQIEAKSSDGELEFNVLVPNV 257

Query: 144 QAGLTADYRTGQIAI 158
           Q GL  DYRTG++ +
Sbjct: 258 QKGLQLDYRTGEVTV 272
```

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

```
 Identities = 73/135 (54%), Positives = 92/135 (68%), Gaps = 2/135 (1%)

Query: 24  PLGWHQVA-TNDHYGHAVDKGHLIAYALAGNFKGWDASVSNPQNVVTQTAHSNQSNQKIN 82
           P GWH++      Y HAVD+GHL+ YAL G  KG+DAS  NP N+ TQ + +NQ+N+
Sbjct: 26  PAGWHRLHHLKGSYDHAVDRGHLLGYALVGGLKGFDASTGNPDNIATQLSWANQANKPYL 85

Query: 83  RGQNYYESLVRKAVDQNKRVRYRVTPLYRNDTDLVPFAMHLEAKSQDGTLEFNVAIPNTQ 142
            GQNYYE LVR+A+D+  RVRYRVT LY  D +L+     HLEAKS D +L FNV +PN Q
Sbjct: 86  TGQNYYEGLVRRALDKGHRVRYRVTLLYDGD-NLLASGSHLEAKSSDDSLTFNVFVPNVQ 144

Query: 143 ASYTMDYATGEITLN 157
           A  T DY TG+I +N
Sbjct: 145 AGLTADYRTGQIAIN 159
```

SEQ ID 9324 (GBS656) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 186 (lane 10; MW 57kDa).

GBS656-GST was purified as shown in Figure 236, lane 4.

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

Example 402

A DNA sequence (GBSx0437) was identified in *S.agalactiae* <SEQ ID 1303> which encodes the amino acid sequence <SEQ ID 1304>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have a cleavable N-term signal seq.

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

A related GBS nucleic acid sequence <SEQ ID 9321> which encodes amino acid sequence <SEQ ID 9322> 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 1305> which encodes the amino acid sequence <SEQ ID 1306>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5350(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 = 24/73 (32%), Positives = 37/73 (49%), Gaps = 2/73 (2%)

Query: 1   MFYMKLANRLSLAATIVNEANANSPFGIIIHSDKAENVEWNDFETQFPDLFNSPKKEESP 60
           + YMKLA    L  TI+ E +  SPF  I+H+D A N++    E       N    +++P
Sbjct: 80  ILYMKLAKENHLPVTIITETHMTSPFAFILHTDHAINLKETRLEVILKQTKNDQLSKQTP 139

Query: 61  K--KSLWQHFFSQ 71
           +  KS W+ F  +
Sbjct: 140 EKTKSFWKRFLKK 152
```

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

Example 403

A DNA sequence (GBSx0438) was identified in *S.agalactiae* <SEQ ID 1307> which encodes the amino acid sequence <SEQ ID 1308>. This protein is predicted to be Isopentenyl-diphosphate delta-isomerase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1649(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:AAG20030 GB:AE005083 isopentenyl pyrophosphate isomerase; Idi
           [Halobacterium sp. NRC-1]
 Identities = 24/77 (31%), Positives = 40/77 (51%)

Query: 14  TGLTLNRDQNIPQGLFHLVVDVILFHEDGDVLMMKRHPKKKAFPAYFEATAGGSALKGEN 73
           TGL    D +   G+ H      +LF EDG VL+ +R  +K+ +  +++ T     ++G++
Sbjct: 42  TGLANRLDAHTGDGVRHRAFTCLLFDEDGRVLLAQRADRKRLWDTHWDGTVASHPIEGQS 101

Query: 74  AKQAILRELKEETGIVP 90
             A  + L EE GI P
Sbjct: 102 QVDATRQRLAEELGIEP 118
```

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 404

A DNA sequence (GBSx0439) was identified in *S.agalactiae* <SEQ ID 1309> which encodes the amino acid sequence <SEQ ID 1310>. This protein is predicted to be phosphoserine phosphatase (serB). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0613(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:CAB50876 GB:AL096844 putative phosphoserine phosphatase
          [Streptomyces coelicolor A3(2)]
 Identities = 96/193 (49%), Positives = 132/193 (67%)

Query: 5   LLVMDVDSTLIMEEAIDLLAIEAGVGKQVAALTDAAMRGELDFEEALKKRVALLKGLPVT 64
           L+VMDVDSTLI +E I+L A  AG    +VA +T AAMRGELDFE++L  RVALL GL  +
Sbjct: 183 LVVMDVDSTLIQDEVIELFAAHAGCEDEVAEVTAAAMRGELDFEQSLHARVALLAGLDAS 242

Query: 65  ILTDILSSIHFTPGAYELIKECHKRQMKVGLVSGGFHETIDILAKQLQVDYVKANRLGVK 124
           ++  + + +  TPGA  LI+   +   +VG+VSGGF +  D L +QL +D+ +AN L +
Sbjct: 243 VVDKVRAEVRLTPGARTLIRTLKRLGYQVGVVSGGFTQVTDALQEQLGLDFAQANTLEIV 302

Query: 125 GGFLTGEVEGEIVTKEVKKIKLKEWASENHLDLSQTIAMGDGANDLPMIKSAGVGIAFCA 184
            G LTG V GEIV +  K   L+ +A+   + LSQT+A+GDGANDL M+ +AG+G+AF A
Sbjct: 303 DGRLTGRVTGEIVDRAGKARLLRRFAAAAGVPLSQTVAIGDGANDLDMLNAAGLGVAFNA 362

Query: 185 KPIVREEAAYQIN 197
           KP+VRE A   +N
Sbjct: 363 KPVVREAAHTAVN 375
```

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 405

A DNA sequence (GBSx0440) was identified in *S.agalactiae* <SEQ ID 1311> which encodes the amino acid sequence <SEQ ID 1312>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-17.88   Transmembrane    5 -  21 (   1 -  29)

----- Final Results -----
             bacterial membrane --- Certainty=0.8153(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:BAB06924 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 122/553 (22%), Positives = 265/553 (47%), Gaps = 12/553 (2%)
```

```
Query: 7   LLLVAIVLLVIIAYVVGVVIRKRNDTLIANLETRKQELVDLPVQEEIEQVKLLHLIGQSQ 66
           +++ ++++L +  +V G + RK     +  LE  K +++  P+ +EI +VK L + G+++
Sbjct: 3   IVVFSLLVLTVTFFVYGALRRKAFYKRVDKLEDWKNDILQRPIPDEIGKVKGLTMSGETE 62

Query: 67  STFREWNQKWTDLSTNSFKDIDFHLVEAENLNDSFNFVRAKHEIDNVDSQLTIIEEDIVS 126
             F  W   W D+       +++  L + E+  + + F +AK  +D ++ +L  IEE +
Sbjct: 63  EKFEVWRSDWDDIVGVILPNVEEQLFDVEDFANKYRFQKAKALLDTIEQRLHSIEEQLKI 122

Query: 127 IREALEVLKEQEEKNSARVTHALDLYETLQKSISEKEDNYGTTMPEIEKQLKNIEAEFSH 186
           + + ++VL + EE+N   +    +L + L K    +  +  ++     +++L+
Sbjct: 123 MVDDIQVLVQSEEQNRTEIGSVRELQQKLIKEAITRRGSLSSSAKVFDEKLEKANELLQA 182

Query: 187 FVTLNSTGDPIEASEVLNKAEEHTIALGQITEQIPAIVAKLEDDFPDQLDDLETGYRRLL 246
           F      G+ I+ASEVL +A+E    +  + + +P +  +L+ + P +L +L+ G R +
Sbjct: 183 FDERTEKGNYIQASEVLEEAKELLGQIEHLLKIVPGLFVELQTNIPAELTNLKNGLRDME 242

Query: 247 EENYHFPEKDIEQRFQEVREAIRSNSDGLVSLDLDRARDENEHIQEKIDKLYDIFEREIA 306
           E  +      I+ + + + E      + L  L+ +   +E   I+E +++++++ E+E+
Sbjct: 243 EAGFFLETFAIDSQMERLEEKRVELLEQLTVLECNGMEEEINFIEESMEQMFELLEKEVE 302

Query: 307 AYKVAHKDSKIIPQFLAHAKSNNEQLGH---EIKRLSAKYILNENESLSLRSFTNDLEEI 363
           A    ++ + ++P          E+L H   E + +   Y L E E +  +    +L+E+
Sbjct: 303 A---KNEITILLPNLREDLTKTEEKLTHLKEETESVQLSYRLAEEELVFQQKLGKELKEL 359

Query: 364 ETKVLPSVENFGQEASPYTHLQILFERTLKTLTTVEENQMEVFEAVKTIESVETRARQNM 423
             ++    E   ++   ++ ++ + E   + LT  +    +  E++ ++   E +A++ +
Sbjct: 360 RQQLQVIDEVTEEQKQTFSSVRSMLEEWREGLTACQNKIEQAQESLNSLRKDELKAKEEL 419

Query: 424 DKYVNKLHMIKRFMEKRNLPGIPQDFLSTFFTTSSQIEALINELSRGRIDIEAVSRLNDV 483
             +   KL   KR ++K N+PG+P+  L          ++   I +LS   +++  V+ L D
Sbjct: 420 KQLKEKLLEDKRLVQKSNIPGLPETLLHRLEDGEQKLAQAIAKLSDVPLEMGRVTALVDE 479

Query: 484 TTNAIANLEQATYLVVQDATLTEQLLQYSNRYRSFEQNVQKSFEQALYLFEVEHNYKASF 543
             T  I      +  ++ A L E ++QY NRYRS    V+K    A  LF      +
Sbjct: 480 AQGLIHENSSILHETIEKARLAEHVIQYGNRYRSRSAEVKKRLSNAEELFRA-----FEY 534

Query: 544 DE-ISYALETVEP 555
           DE I  A++ +EP
Sbjct: 535 DEAIEMAVQAIEP 547
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1313> which encodes the amino acid sequence <SEQ ID 1314>. Analysis of this protein sequence reveals the following:

```
     Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-18.04   Transmembrane    5 -  21 (   1 -  29)

----- Final Results -----
              bacterial membrane --- Certainty=0.8217(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:BAB06924 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 131/555 (23%), Positives = 269/555 (47%), Gaps = 16/555 (2%)

Query: 7   LLIVAIVLLVIIAYLVGVIIRKRNDSLITSLEERKQALFALPVNDEIEEVKSLHLIGQSQ 66
           +++ ++++L +  ++ G + RK     +  LE+ K  +   P+ DEI +VK L + G+++
Sbjct: 3   IVVFSLLVLTVTFFVYGALRRKAFYKRVDKLEDWKNDILQRPIPDEIGKVKGLTMSGETE 62

Query: 67  TSFREWNQKWVDLTVNSFADIENHIFEAENLNDTFNFIRAKHEINSVESQLNLVEEDIAS 126
             F  W   W D+       ++E  +F+ E+  + + F +AK  +++++ +L  +EE +
Sbjct: 63  EKFEVWRSDWDDIVGVILPNVEEQLFDVEDFANKYRFQKAKALLDTIEQRLHSIEEQLKI 122

Query: 127 IREALNILKEQEEKNSARVTHALDLYEKLQASISENEDNFGSTMPEIDKQMKNIETEFSQ 186
           + + + +L + EE+N   +    +L +KL          +  S+    D++++
Sbjct: 123 MVDDIQVLVQSEEQNRTEIGSVRELQQKLIKEAITRRGSLSSSAKVFDEKLEKANELLQA 182
```

```
Sbjct: 123 MVDDIQVLVQSEEQNRTEIGSVRELQQKLIKEAITRRGSLSSSAKVFDEKLEKANELLQA 182

Query: 187 FVALNSSGDPVEASEVLDRAEEHTIALGQITEQIPAIVAKLEDDFPDQLDDLETGYRRLL 246
           F      G+ ++ASEVL+ A+E    +  + + +P +  +L+ + P +L +L+ G R +
Sbjct: 183 FDERTEKGNYIQASEVLEEAKELLGQIEHLLKIVPGLFVELQTNIPAELTNLKNGLRDME 242

Query: 247 EENYHFPEKNIEARFQEIRESIRANSSELVTLDLDRAREENTHIQERIDSLYEVFEREIA 306
           E  +      I+++ + + E       +L  L+ +   EE   I+E ++ ++E+ E+E
Sbjct: 243 EAGFFLETFAIDSQMERLEEKRVELLEQLTVLECNGMEEEINFIEESMEQMFELLEKE-- 300

Query: 307 AYKVAAKN--SKMLPRYLEHVKRNNEQ---LKDEIARLSRKYILSETESLTVKAFEKDIK 361
              V AKN  + +LP   E + +  E+   LK+E   +   Y L+E E +  +   K++K
Sbjct: 301 ---VEAKNEITILLPNLREDLTKTEEKLTHLKEETESVQLSYRLAEEELVFQQKLGKELK 357

Query: 362 EIEDSTLAVAEQFGLQEKPFSELQVTFERSIKTLTNVESGQMDVFAAVKDIEKIESQARH 421
           E+      + E    Q++ FS ++   E   + LT  ++       ++  + K E +A+
Sbjct: 358 ELRQQLQVIDEVTEEQKQTFSSVRSMLEEWREGLTACQNKIEQAQESLNSLRKDELKAKE 417

Query: 422 NLDVYVTQLHMIKRYMEKRHLPGIPQDFLSAFFTTSSQLEALMDELSRGRINIEAVSRLS 481
            L     +L   KR ++K ++PG+P+  L         +L   + +LS   + +  V+ L
Sbjct: 418 ELKQLKEKLLEDKRLVQKSNIPGLPETLLHRLEDGEQKLAQAIAKLSDVPLEMGRVTALV 477

Query: 482 EVATVAIANLEDLTYQVVQNATLTEQLLQYSNRYRSFEAGVQSSFEHALRLFEVENDYQA 541
           + A   I     + ++ ++ A L E ++QY NRYRS  A V+     +A  LF
Sbjct: 478 DEAQGLIHENSSILHETIEKARLAEHVIQYGNRYRSRSAEVKKRLSNAEELFRA-----F 532

Query: 542 SFDE-ISYALETVEP 555
            +DE I  A++ +EP
Sbjct: 533 EYDEAIEMAVQAIEP 547
```

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

```
 Identities = 429/574 (74%), Positives = 503/574 (86%)

Query: 1   MSSGIILLLVAIVLLVIIAYVVGVVIRKRNDTLIANLETRKQELVDLPVQEEIEQVKLLH 60
           MSSGIILL+VAIVLLVIIAY+VGV+IRKRND+LI +LE RKQ L  LPV +EIE+VK LH
Sbjct: 1   MSSGIILLIVAIVLLVIIAYLVGVIIRKRNDSLITSLEERKQALFALPVNDEIEEVKSLH 60

Query: 61  LIGQSQSTFREWNQKWTDLSTNSFKDIDFHLVEAENLNDSFNFVRAKHEIDNVDSQLTII 120
           LIGQSQ++FREWNQKW DL+ NSF DI+ H+ EAENLND+FNF+RAKHEI++V+SQL ++
Sbjct: 61  LIGQSQSTSFREWNQKWVDLTVNSFADIENHIFEAENLNDTFNFIRAKHEINSVESQLNLV 120

Query: 121 EEDIVSIREALEVLKEQEEKNSARVTHALDLYETLQKSISEKEDNYGTTMPEIEKQLKNI 180
           EEDI SIREAL +LKEQEEKNSARVTHALDLYE LQ SISE EDN+G+TMPEI+KQ+KNI
Sbjct: 121 EEDIASIREALNILKEQEEKNSARVTHALDLYEKLQASISENEDNFGSTMPEIDKQMKNI 180

Query: 181 EAEFSHFVTLNSTGDPIEASEVLNKAEEHTIALGQITEQIPAIVAKLEDDFPDQLDDLET 240
           E EFS FV LNS+GDP+EASEVL++AEEHTIALGQITEQIPAIVAKLEDDFPDQLDDLET
Sbjct: 181 ETEFSQFVALNSSGDPVEASEVLDRAEEHTIALGQITEQIPAIVAKLEDDFPDQLDDLET 240

Query: 241 GYRRLLEENYHFPEKDIEQRFQEVREAIRSNSDGLVSLDLDRARDENEHIQEKIDKLYDI 300
           GYRRLLEENYHFPEK+IE RFQE+RE+IR+NS  LV+LDLDRAR+EN HIQE+ID LY++
Sbjct: 241 GYRRLLEENYHFPEKNIEARFQEIRESIRANSSELVTLDLDRAREENTHIQERIDSLYEV 300

Query: 301 FEREIAAYKVAHKDSKIIPQFLAHAKSNNEQLGHEIKRLSAKYILNENESLSLRSFTNDL 360
           FEREIAAYKVA K+SK++P++L H K NNEQL  EI RLS KYIL+E ESL++++F  D+
Sbjct: 301 FEREIAAYKVAAKNSKMLPRYLEHVKRNNEQLKDEIARLSRKYILSETESLTVKAFEKDI 360

Query: 361 EEIETKVLPSVENFGQEASPYTHLQILFERTLKTLTTVEENQMEVFEAVKTIESVETRAR 420
           +EIE    L   E FG +  P++ LQ+ FER++KTLT VE  QM+VF AVK IE +E++AR
Sbjct: 361 KEIEDSTLAVAEQFGLQEKPFSELQVTFERSIKTLTNVESGQMDVFAAVKDIEKIESQAR 420

Query: 421 QNMDKYVNKLHMIKRFMEKRNLPGIPQDFLSTFFTTSSQIEALINELSRGRIDIEAVSRL 480
            N+D YV +LHMIKR+MEKR+LPGIPQDFLS FFTTSSQ+EAL++ELSRGRI+IEAVSRL
Sbjct: 421 HNLDVYVTQLHMIKRYMEKRHLPGIPQDFLSAFFTTSSQLEALMDELSRGRINIEAVSRL 480

Query: 481 NDVTTNAIANLEQATYLVVQDATLTEQLLQYSNRYRSFEQNVQKSFEQALYLFEVEHNYK 540
           ++V T AIANLE  TY VVQ+ATLTEQLLQYSNRYRSFE  VQ SFE AL LFEVE++Y+
```

```
Sbjct: 481 SEVATVAIANLEDLTYQVVQNATLTEQLLQYSNRYRSFEAGVQSSFEHALRLFEVENDYQ 540

Query: 541 ASFDEISYALETVEPGVTDRFVTSYEKTQERIRF 574
           ASFDEISYALETVEPGVTDRFV SYEKT+E IRF
Sbjct: 541 ASFDEISYALETVEPGVTDRFVNSYEKTREHIRF 574
```

SEQ ID 1312 (GBS642) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 142 (lane 2-4; MW 27kDa).

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

Example 406

A DNA sequence (GBSx0441) was identified in *S.agalactiae* <SEQ ID 1315> which encodes the amino acid sequence <SEQ ID 1316>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2471(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 9671> which encodes amino acid sequence <SEQ ID 9672> was also identified.

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

```
>GP:CAA91553 GB:Z67740 DNA gyrase [Streptococcus pneumoniae]
 Identities = 574/650 (88%), Positives = 618/650 (94%), Gaps = 2/650 (0%)

Query: 1   MTEETKNMEQRAQEYDASQIQVLEGLEAVRMRPGMYIGSTSKEGLHHLVWEIVDNSIDEA 60
           MTEE KN++  AQ+YDASQIQVLEGLEAVRMRPGMYIGSTSKEGLHHLVWEIVDNSIDEA
Sbjct: 1   MTEEIKNLQ--AQDYDASQIQVLEGLEAVRMRPGMYIGSTSKEGLHHLVWEIVDNSIDEA 58

Query: 61  LAGFAGHIKVYIEPDNSITVVDDGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGGYKV 120
           LAGFA HI+V+IEPD+SITVVDDGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGGYKV
Sbjct: 59  LAGFASHIQVFIEPDDSITVVDDGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGGYKV 118

Query: 121 SGGLHGVGSSVVNALSTQLDVKVYKNGKVHYQEYQRGVVVNDLEIIGDTDLSGTTVHFTP 180
           SGGLHGVGSSVVNALSTQLDV V+KNGK+HYQEY+RG VV DLE++GDTD +GTTVHFTP
Sbjct: 119 SGGLHGVGSSVVNALSTQLDVHVHKNGKIHYQEYRRGHVVADLEVVGDTDRTGTTVHFTP 178

Query: 181 DPEIFTETTVFDFDKLAKRIQELAFLNRGLRISISDKREGQEVEKEYHYEGGIGSYVEFI 240
           DPEIFTETT+FDFDKL KRIQELAFLNRGL+ISI+DKR+G E  K YHYEGGI SYVE+I
Sbjct: 179 DPEIFTETTIFDFDKLNKRIQELAFLNRGLQISITDKRQGLEQTKHYHYEGGIASYVEYI 238

Query: 241 NENKEVIFENPIYTDGELDGISVEVAMQYTTGYQETVMSFANNIHTHEGGTHEQGFRTAL 300
           NENK+VIF+ PIYTDGE+D I+VEVAMQYTTGY E VMSFANNIHTHEGGTHEQGFRTAL
Sbjct: 239 NENKDVIFDTPIYTDGEMDDITVEVAMQYTTGYHENVMSFANNIHTHEGGTHEQGFRTAL 298

Query: 301 TRVINDYAKKNKILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 360
           TRVINDYA+KNK+LK+NEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT
Sbjct: 299 TRVINDYARKNKLLKDNEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 358

Query: 361 NRLFSEAFNRFLLENPQVAKKIVEKGILASKARIAAKRAREVTRKKSGLEISNLPGKLAD 420
           NRLFSEAF+ FL+ENPQ+AK+IVEKGILA+KAR+AAKRAREVTRKKSGLEISNLPGKLAD
Sbjct: 359 NRLFSEAFSDFLMENPQIAKRIVEKGILAAKARVAAKRAREVTRKKSGLEISNLPGKLAD 418

Query: 421 CSSNNAEMNELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS 480
           CSSNN    ELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKA+MDKILANEEIRS
Sbjct: 419 CSSNNPAETELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKASMDKILANEEIRS 478
```

```
Query: 481 LFTAMGTGFGADFDVSKVRYQKLVIMTDADVDGAHIRTLLLTLIYRFMRPVLEAGYVYIA 540
           LFTAMGTGFGA+FDVSK RYQKLV+MTDADVDGAHIRTLLLTLIYR+M+P+LEAGYVYIA
Sbjct: 479 LFTAMGTGFGAEFDVSKARYQKLVLMTDADVDGAHIRTLLLTLIYRYMKPILEAGYVYIA 538

Query: 541 QPPIYGVKVGSEIKAYIQPGVNQEEELRQALDTYSSGRSKPTVQRYKGLGEMDDHQLWET 600
           QPPIYGVKVGSEIK YIQPG +QE +L++AL  YS GR+KPT+QRYKGLGEMDDHQLWET
Sbjct: 539 QPPIYGVKVGSEIKEYIQPGADQEIKLQEALARYSEGRTKPTIQRYKGLGEMDDHQLWET 598

Query: 601 TMDPENRLMARVSVDDAAEADKIFDMLMGDRVEPRREFIEANAVYSNLDI 650
           TMDPE+RLMARVSVDDAAEADKIFDMLMGDRVEPRREFIE NAVYS LD+
Sbjct: 599 TMDPEHRLMARVSVDDAAEADKIFDMLMGDRVEPRREFIEENAVYSTLDV 648
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1317> which encodes the amino acid sequence <SEQ ID 1318>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1698(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 = 584/650 (89%), Positives = 618/650 (94%)

Query: 1   MTEETKNMEQRAQEYDASQIQVLEGLEAVRMRPGMYIGSTSKEGLHHLVWEIVDNSIDEA 60
           M EE K+ E++ QEYDASQIQVLEGLEAVRMRPGMYIGST+KEGLHHLVWEIVDNSIDEA
Sbjct: 1   MIEENKHFEKKMQEYDASQIQVLEGLEAVRMRPGMYIGSTAKEGLHHLVWEIVDNSIDEA 60

Query: 61  LAGFAGHIKVYIEPDNSITVVDDGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGGYKV 120
           LAGFA HIKV+IE DNSITVVDDGRGIPVDIQ KTGRPAVETVFTVLHAGGKFGGGGYKV
Sbjct: 61  LAGFASHIKVFIEADNSITVVDDGRGIPVDIQAKTGRPAVETVFTVLHAGGKFGGGGYKV 120

Query: 121 SGGLHGVGSSVVNALSTQLDVKVYKNGKVHYQEYQRGVVVNDLEIIGDTDLSGTTVHFTP 180
           SGGLHGVGSSVVNALSTQLDV+VYKNG++HYQE++RG VV DLE+IG TD++GTTVHFTP
Sbjct: 121 SGGLHGVGSSVVNALSTQLDVRVYKNGQIHYQEFKRGAVVADLEVIGTTDVTGTTVHFTP 180

Query: 181 DPEIFTETTVFDFDKLAKRIQELAFLNRGLRISISDKREGQEVEKEYHYEGGIGSYVEFI 240
           DPEIFTETT FD+  LAKRIQELAFLNRGL+ISI+DKR G E E+ + YEGGIGSYVEF+
Sbjct: 181 DPEIFTETTQFDYSVLAKRIQELAFLNRGLKISITDKRSGMEQEEHFLYEGGIGSYVEFL 240

Query: 241 NENKEVIFENPIYTDGELDGISVEVAMQYTTGYQETVMSFANNIHTHEGGTHEQGFRTAL 300
           N+ K+VIFE PIYTDGEL+GI+VEVAMQYTT YQETVMSFANNIHTHEGGTHEQGFR AL
Sbjct: 241 NDKKDVIFETPIYTDGELEGIAVEVAMQYTTSYQETVMSFANNIHTHEGGTHEQGFRAAL 300

Query: 301 TRVINDYAKKNKILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 360
           TRVINDYAKKNKILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT
Sbjct: 301 TRVINDYAKKNKILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 360

Query: 361 NRLFSEAFNRFLLENPQVAKKIVEKGILASKARIAAKRAREVTRKKSGLEISNLPGKLAD 420
           NRLFSEAF RFLLENPQVA+KIVEKGILASKARIAAKRAREVTRKKSGLEISNLPGKLAD
Sbjct: 361 NRLFSEAFQRFLLENPQVARKIVEKGILASKARIAAKRAREVTRKKSGLEISNLPGKLAD 420

Query: 421 CSSNNAEMNELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS 480
           CSSN+A  NELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS
Sbjct: 421 CSSNDANQNELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS 480

Query: 481 LFTAMGTGFGADFDVSKVRYQKLVIMTDADVDGAHIRTLLLTLIYRFMRPVLEAGYVYIA 540
           LFTAMGTGFGADFDVSK RYQKLVIMTDADVDGAHIRTLLLTLIYRFMRPVLEAGYVYIA
Sbjct: 481 LFTAMGTGFGADFDVSKARYQKLVIMTDADVDGAHIRTLLLTLIYRFMRPVLEAGYVYIA 540

Query: 541 QPPIYGVKVGSEIKAYIQPGVNQEEELRQALDTYSSGRSKPTVQRYKGLGEMDDHQLWET 600
           QPPIYGVKVGSEIK YIQPG++QE++L+ AL+ YS GRSKPTVQRYKGLGEMDDHQLWET
Sbjct: 541 QPPIYGVKVGSEIKEYIQPGIDQEDQLKTALEKYSIGRSKPTVQRYKGLGEMDDHQLWET 600
```

```
Query: 601 TMDPENRLMARVSVDDAAEADKIFDMLMGDRVEPRREFIEANAVYSNLDI 650
           TMDPENRLMARV+VDDAAEADK+FDMLMGDRVEPRR+FIE NAVYS LDI
Sbjct: 601 TMDPENRLMARVTVDDAAEADKVFDMLMGDRVEPRRDFIEENAVYSTLDI 650
```

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

Example 407

A DNA sequence (GBSx0442) was identified in *S.agalactiae* <SEQ ID 1319> which encodes the amino acid sequence <SEQ ID 1320>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3186(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:CAA91552 GB:Z67740 unidentified [Streptococcus pneumoniae]
 Identities = 82/142 (57%), Positives = 105/142 (73%)

Query: 45  LKESTADAIAYFIPEEADFLKEYKANEAKVLETPILFQGAKELLAKIQRQGSRNFLVSHR 104
           LK ST  AI  F P   +FL++YK NEA+ LE PILF+G  +LL  I  QG R+FLVSHR
Sbjct: 2   LKVSTPFAIETFAPNLENFLEKYKENEARELEHPILFEGVSDLLEDILNQGGRHFLVSHR 61

Query: 105 DNQVIVILEKTEIIDYFTEVVTADNGFSRKPSPESMLYLKEKYQIDNCLVIGDRDIDKQA 164
           ++QV+ ILEKT I  YFTEVVT+ +GF RKP+PESMLYL+EKYQI + LVIGDR ID +A
Sbjct: 62  NDQVLEILEKTSIAAYFTEVVTSSSGFKRKPNPESMLYLREKYQISSGLVIGDRPIDIEA 121

Query: 165 GESAGFDTLLVDGSKSLMEIIE 186
           G++AG DT L     +L ++++
Sbjct: 122 GQAAGLDTHLFTSIVNLRQVLD 143
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1321> which encodes the amino acid sequence <SEQ ID 1322>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2472(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 = 122/185 (65%), Positives = 145/185 (77%)

Query: 1   MNYHDYIWDLGGTLLDNYESSTRAFVETLKEFGYQADHDSVYQKLKESTADAIAYFIPEE 60
           MNY DYIWDLGGTLLDNYE ST+AFV+TL  F    DHD+VYQKLKESTA A+A F P E
Sbjct: 4   MNYQDYIWDLGGTLLDNYELSTQAFVQTLAFFSLPGDHDAVYQKLKESTAIAVAMFAPNE 63

Query: 61  ADFLKEYKANEAKVLETPILFQGAKELLAKIQRQGSRNFLVSHRDNQVIVILEKTEIIDY 120
            +FL  Y+  EA  L  PI   GAKE+L KI   GSRNFL+SHRD QV  +LE+  ++ Y
Sbjct: 64  PEFLHVYRLREADKLAQPIWCLGAKEILGKIATSGSRNFLISHRDCQVNQLLEQAGLLIY 123

Query: 121 FTEVVTADNGFSRKPSPESMLYLKEKYQIDNCLVIGDRDIDKQAGESAGFDTLLVDGSKS 180
           FTEVVTA NGF+RKP+PES+ YLKEKY I++ LVIGDR IDKQAG++AGF+TLLVDG K+
Sbjct: 124 FTEVVTASNGFARKPNPESLFYLKEKYDINSGLVIGDRLIDKQAGQAAGFNTLLVDGRKN 183
```

```
Query: 181 LMEII 185
          L+EI+
Sbjct: 184 LLEIV 188
```

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

Example 408

A DNA sequence (GBSx0443) was identified in *S.agalactiae* <SEQ ID 1323> which encodes the amino acid sequence <SEQ ID 1324>. This protein is predicted to be stage V sporulation protein E (rodA). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-11.15   Transmembrane  206 - 222 ( 177 - 226)
   INTEGRAL    Likelihood =-10.14   Transmembrane   58 -  74 (  50 -  82)
   INTEGRAL    Likelihood = -9.34   Transmembrane  182 - 198 ( 177 - 205)
   INTEGRAL    Likelihood = -8.55   Transmembrane  158 - 174 ( 156 - 177)
   INTEGRAL    Likelihood = -8.12   Transmembrane  300 - 316 ( 299 - 324)
   INTEGRAL    Likelihood = -2.66   Transmembrane   86 - 102 (  83 - 102)
   INTEGRAL    Likelihood = -2.34   Transmembrane  338 - 354 ( 338 - 357)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9669> which encodes amino acid sequence <SEQ ID 9670> was also identified.

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

```
>GP:CAB15838 GB:Z99123 alternate gene name: ipa-42d-similar to
          cell-division protein [Bacillus subtilis]
 Identities = 142/392 (36%), Positives = 237/392 (60%), Gaps = 23/392 (5%)

Query: 10  QKSNYFKGQIDYAVVIPVFFLLMIGLASIYVA-TMNDYPSNIYIAMFQQVSWIIMGCIIA 68
           Q+S +++G  D   +  VFF+  I + SIY A     Y +  +I   QQ+ + ++G +
Sbjct: 7   QQSPFYQG--DLIFIFGVFFI--ISVVSIYAAGQFGQYGNTDWI---QQIVFYLLGAVAI 59

Query: 69  FVVMLFSTEFLWKATPYLYALGLTLMVLPLIFYSPQLFAAT--GAKNWVTIGSVTLFQPS 126
            V++ F  E L K + Y++ +G+  +++  SP+  A    GAK+W  IG +T+ QPS
Sbjct: 60  TVLLYFDLEQLEKLSLYIFIIGILSLIILKI--SPESIAPVIKGAKSWFRIGRITI-QPS 116

Query: 127 EFMKISYILMLSRITVSFHQKNRKTFQDDWKLL-GLFGLVTLPVMILLMLQKDLGTALVF 185
           EFMK+  I+ML+ +    + K  +T +DD  LL  + G+  +PV ++LM  +D GTA +
Sbjct: 117 EFMKVGLIMMLASVIGKANPKGVRTLRDDIHLLLKIAGVAVIPVGLILM--QDAGTAGIC 174

Query: 186 LAILSGLILLSGISWWIILPILSTIVLFIASFLMIFISPNGKEWFYNLGMDTYQINRLSA 245
           + I+  ++ +SGI+W +I  I  + +L I+  L++ I  N  +   ++G+  YQI R+++
Sbjct: 175 MFIVLVMVFMSGINWKLIAIIAGSGILLISLILLVMI--NFPDVAKSVGIQDYQIKRVTS 232

Query: 246 WIDPFSFAD---SIAYQQTQGMVSIGSGGVTGKGFNILELSVPVRESDMIFTVIAENFGF 302
           W+   +       + ++Q  Q +++IGSGG+ G G + L++ VP   +D IF++I E+FGF
Sbjct: 233 WVSASNETQEDSNDSWQVDQAIMAIGSGGILGNGISNLKVYVPESTTDFIFSIIGESFGF 292

Query: 303 IGSAIVLGLYLIIIYRMLRIT--IESNNQFYTFISTGFIMMIVFHVFENIGAAVGILPLT 360
           IG AIV+ ++  +IYR++ +   I   N+F +F   G+  +IV H F+NIG  +GI+P+T
Sbjct: 293 IGCAIVVIMFFFLIYRLVVLIDKIHPFNRFASFFCVGYTALIVIHTFQNIGMNIGIMPVT 352

Query: 361 GIPLPFISQGGSSLLSNLIGIGLVLSMSYQNT 392
           GIPL F+S GGSS LS LIG G+V + S Q T
Sbjct: 353 GIPLLFVSYGGSSTLSTLIGFGIVYNASVQLT 384
```

There is also homology to SEQ ID 1028.

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

Example 409

A DNA sequence (GBSx0444) was identified in *S.agalactiae* <SEQ ID 1325> which encodes the amino acid sequence <SEQ ID 1326>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3195(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 1327> which encodes the amino acid sequence <SEQ ID 1328>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2735(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 = 38/55 (69%), Positives = 48/55 (87%)

Query: 8  DEFKEAIDKGYISGNTVAIVRKNGKIFDYVLLHEEVREEEVVTVERVLDVLRKLS 62
          DEFK+AID GYI+G+TVAIVRK+G+IFDYVL HE+V+  EVVT E+V +VL +LS
Sbjct: 5  DEFKQAIDNGYIAGDTVAIVRKDGQIFDYVLPHEKVKNGEVVTKEKVEEVLVELS 59
```

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

Example 410

A DNA sequence (GBSx0445) was identified in *S.agalactiae* <SEQ ID 1329> which encodes the amino acid sequence <SEQ ID 1330>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4241(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1331> which encodes the amino acid sequence <SEQ ID 1332>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4551(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 = 57/66 (86%), Positives = 63/66 (95%)

Query: 1  MSQEKLKSKLDQAKGGAKEGFGKITGDKELEAKGFIEKTIAKGKELADDAKDAVEGAVDA 60
          MS+EKLKSK++QA GG KEG GK+TGDKELEAKGF+EKTIAKGKELADDAK+AVEGAVDA
Sbjct: 1  MSEEKLKSKIEQASGGLKEGAGKLTGDKELEAKGFVEKTIAKGKELADDAKEAVEGAVDA 60

Query: 61 VKEKLK 66
          VKEKLK
Sbjct: 61 VKEKLK 66
```

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

Example 411

A DNA sequence (GBSx0447) was identified in *S.agalactiae* <SEQ ID 1333> which encodes the amino acid sequence <SEQ ID 1334>. This protein is predicted to be TnpA (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=0.3961(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 9667> which encodes amino acid sequence <SEQ ID 9668> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1335> which encodes the amino acid sequence <SEQ ID 1336>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3365(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 = 152/160 (95%), Positives = 154/160 (96%)

Query: 1   MKNMALPKMATVKTKTALKKTQKTYPQNLLNQKFNPDKPNQVWSTDFTYISIGYKKYVYL 60
           MKNMALPKMATVK KTALK+TQKTYPQNLLNQKFNPDKPNQVWSTDFTYISIGYKKYVYL
Sbjct: 194 MKNMALPKMATVKPKTALKRTQKTYPQNLLNQKFNPDKPNQVWSTDFTYISIGYKKYVYL 253

Query: 61  CAIIDLYSRKYIAWKLSHRMDAKLACDTLELALNKRKIEGTLLFHSDQGSQFKAREFRKI 120
           CAI+DLYSRK IAWKLSHRMDAKLACDTLELALNKRKIEGTLLFHSDQGSQFKARE RKI
Sbjct: 254 CAILDLYSRKCIAWKLSHRMDAKLACDTLELALNKRKIEGTLLFHSDQGSQFKARELRKI 313
```

```
Query: 121 IDDNNIMHSFSKPRYPYDNAVTEAFFKYLKHRQINQKNYQ 160
           IDDN IMHSFSKP YPYDNAVTEAFFKYLKHRQINQK YQ
Sbjct: 314 IDDNTIMHSFSKPGYPYDNAVTEAFFKYLKHRQINQKKYQ 353
```

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

Example 412

A DNA sequence (GBSx0448) was identified in *S.agalactiae* <SEQ ID 1337> which encodes the amino acid sequence <SEQ ID 1338>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1090(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 413

A DNA sequence (GBSx0449) was identified in *S.agalactiae* <SEQ ID 1339> which encodes the amino acid sequence <SEQ ID 1340>. This protein is predicted to be histidine kinase (resE). Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.57   Transmembrane   17 -  33 (   6 -  38)
   INTEGRAL    Likelihood = -4.67   Transmembrane  147 - 163 ( 142 - 166)

----- Final Results -----
             bacterial membrane --- Certainty=0.5628(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:AAD25109 GB:AF140356 VncS [Streptococcus pneumoniae]
 Identities = 178/435 (40%), Positives = 281/435 (63%), Gaps = 1/435 (0%)

Query: 1   MKKLKIFPKMFIQIFSILGILIILVHSLFFFIFPKTYLETRKVKIHIMADEISKNMNGKE 60
           MK+  +F K+FI  FSI  +L+I +H   +F+FP TYL  R+  I   A  I++++ GK+
Sbjct: 1   MKRTGLFAKIFIYTFSIFSVLVICLHLAIYFLFPSTYLSHRQETIGQKATAIAQSLEGKD 60

Query: 61  LKYLDQTLELYSKSSDIKVFIKKNNNKNELQINDNINVNVKSDSNSLIIEEREIKLHDGK 120
            + ++Q L+LYS++SDIK  +K    +++L++ D++ ++    + SL IEERE+K  DG
Sbjct: 61  RQSIEQVLDLYSQTSDIKGTVKGEMTEDKLEVKDSLPLDTDRQTTSLFIEEREVKTQDGG 120

Query: 121 KIHLQFVSTADMQKDAKDLSLKFLPYSLSISFLFSIVISLIYAKSIKNNIQEITMVTDKM 180
            + LQF+++ D+QK+A+ +SL+FLPY+L  SFL S++++ IYA++I   I EI  VT +M
Sbjct: 121 TMILQFLASMDLQKEAEQISLQFLPYTLLASFLISLLVAYIYARTIVAPILEIKRVTRRM 180

Query: 181 IKLDKETRLKISSNDEIGQLKQQINDLYCALLNTINDLEFKNKEILKLEKLKYDFFKGAS 240
           + LD + RL++ S DEIG LK+QIN LY  LL  I DL  KN+ IL+LEK+K +F +GAS
Sbjct: 181 MDLDSQVRLRVDSKDEIGNLKEQINSLYQHLLTVIADLHEKNEAILQLEKMKVEFLRGAS 240
```

```
Query: 241 HELKTPLSSLKILLENMKYNIGKYKDRDFYISECINIVDNLTKNVSQILSFYSIKDLNND 300
           HELKTPL+SLKIL+ENM+ NIG+YKDRD Y+   + IVD L  +V QILS  S+++L +D
Sbjct: 241 HELKTPLASLKILIENMRENIGRYKDRDQYLGVALGIVDELNHHVLQILSLSSVQELRDD 300

Query: 301 EEYLNVGDTLDEVLEKYSILVNQKKININKELLDYNIYIGKTALNIVFSNLISNAVKYTN 360
            E +++      +++ Y++L  ++++ I+  L     Y+  + + ++ SNLISNA+K++
Sbjct: 301 RETIDLLQMTQNLVKDYALLAKERELQIDNSLTHQQAYLNPSVMKLILSNLISNAIKHSV 360

Query: 361 RNGIINIKIANDWLLIENSYDKNKISKINKILDASFDLKLDNSNGLGLNIVKNILNKYNI 420
             G++ I      L IENS    +  K+ +    +   K+  S G+GL +VK++L    +
Sbjct: 361 PGGLVRIGEREGEGLFIENSCSSEEQEKLAQSFSDNASRKVKGS-GMGLFVVKSLLEHEKL 419

Query: 421 KYEILHGENYFIFKI 435
           Y     EN   F I
Sbjct: 420 AYRFEMEENSLTFFI 434
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1341> which encodes the amino acid sequence <SEQ ID 1342>. Analysis of this protein sequence reveals the following:

```
      Possible site: 37

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.83   Transmembrane   14 -  30 (   6 -  35)
   INTEGRAL    Likelihood = -2.44   Transmembrane  157 - 173 ( 156 - 174)

----- 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:AAD25109 GB:AF140356 VncS [Streptococcus pneumoniae]
 Identities = 123/455 (27%), Positives = 223/455 (48%), Gaps = 23/455 (5%)

Query: 3   LIKKTFLVINGLIIVVVTSILLVLYFAMPIYYTKVKDKEVKCEFDQTSKQIKGKTVTEIR 62
           L  K F+    +  V+V  + L +YF  P  Y   + + +  +    ++ ++GK    I
Sbjct: 6   LFAKIFIYTFSIFSVLVICLHLAIYFLFPSTYLSHRQETIGQKATAIAQSLEGKDRQSIE 65

Query: 63  DILTKKINKDNIWYSLVDSDNQLLYPSLQLLDGVSESKDSQNVNIVTTFDNSYSNVKVMS 122
            +L       +I  ++     ++    L++ D +    D Q  ++             +
Sbjct: 66  QVLDLYSQTSDIKGTV---KGEMTEDKLEVKDSLPLDTDRQTTSLF-----------IEE 111

Query: 123 QKVTLRDGKKMTLLGQSSLQPVTDASKVLLDLYPSLLIFSVTVGSIVAYLYSRTSSRRIL 182
           ++V  +DG  M L   +S+    +A ++ L   P  L+ S  +  +VAY+Y+RT    IL
Sbjct: 112 REVKTQDGGTMILQFLASMDLQKEAEQISLQFLPYTLLASFLISLLVAYIYARTIVAPIL 171

Query: 183 SMSQTAKKMVNLEPNLTCTIHGKDEIAMLASDINRLYASLSTSIKSLQKEYEKASDSERE 242
            + +  ++M++L+  +   + KDEI  L   IN LY  L T I  L ++ E     E+
Sbjct: 172 EIKRVTRRMMDLDSQVRLRVDSKDEIGNLKEQINSLYQHLLTVIADLHEKNEAILQLEKM 231

Query: 243 KSEFLRMTSHELKTPITSVIGMIDGMLYNVGDFADRDKYLRKCRDVLEGQAQLVQSILSL 302
           K EFLR  SHELKTP+ S+  +I+ M  N+G + DRD+YL     +++      V  ILSL
Sbjct: 232 KVEFLRGASHELKTPLASLKILIENMRENIGRYKDRDQYLGVALGIVDELNHHVLQILSL 291

Query: 303 SKIETLASQNQELFSLKSSLEEEMEVFLVLSELKHLKVTINLEEQFVKANKVYLLKAIKN 362
           S ++ L   ++E   L    + ++ + +L++ + L++  +L  Q    N   +  + N
Sbjct: 292 SSVQEL-RDDRETIDLLQMTQNLVKDYALLAKERELQIDNSLTHQQAYLNPSVMKLILSN 350

Query: 363 IIDNAFHYTKSGGQVMIQLKDNQLVIKNEAETLLTQQQMKQLFQPFYRPDYSRNRKDGGT 422
           +I NA  ++  GG V I  ++ +L I+N      + ++ ++L Q F      + +RK  G+
Sbjct: 351 LISNAIKHSVPGGLVRIGEREGEGLFIENSC----SSEEQEKLAQSF---SDNASRKVKGS 403

Query: 423 GLGLFITHQILDQHHLAYRFVVLDQRWMVFTIDFP 457
           G+GLF+   +L+   LAYRF  +++  + F IDFP
Sbjct: 404 GMGLFVVKSLLEHEKLAYRF-EMEENSLTFFIDFP 437
```

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

```
 Identities = 108/454 (23%), Positives = 220/454 (47%), Gaps = 22/454 (4%)

Query: 4   LKIFPKMFIQIFSILGILIILVHSLFFFIFPKTYLETRKVKIHIMADEISKNMNGKELKY 63
           +++  K F+ I  ++ +++  +  + +F  P  Y + +  ++    D+ SK + GK +
Sbjct: 1   VRLIKKTFLVINGLIIVVVTSILLVLYFAMPIYYTKVKDKEVKCEFDQTSKQIKGKTVTE 60

Query: 64  LDQTLELYSKSSDIKVFIKKNNNK-----------NELQINDNINVNVKSDSN--SLII 109
           +   L       +I   +  ++N+            +E + + N+N+    D++  ++ +
Sbjct: 61  IRDILTKKINKDNIWYSLVDSDNQLLYPSLQLLDGVSESKDSQNVNIVTTFDNSYSNVKV 120

Query: 110 EEREIKLHDGKKIHLQFVSTADMQKDAKDLSLKFLPYSLSISFLFSIVISLIYAKSIKNN 169
             +++ L DGKK+ L   S+     DA  + L   P  L  S     +++ +Y+++
Sbjct: 121 MSQKVTLRDGKKMTLLGQSSLQPVTDASKVLLDLYPSLLIFSVTVGSIVAYLYSRTSSRR 180

Query: 170 IQEITMVTDKMIKLDKETRLKISSNDEIGQLKQQINDLYCALLNTINDLEFKNKEILKLE 229
           I  ++    KM+ L+      I   DEI  L   IN LY +L  +I  L+ + ++    E
Sbjct: 181 ILSMSQTAKKMVNLEPNLTCTIHGKDEIAMLASDINRLYASLSTSIKSLQKEYEKASDSE 240

Query: 230 KLKYDFFKGASHELKTPLSSLKILLENMKYNIGKYKDRDFYISECINIVDNLTKNVSQIL 289
           + K +F +  SHELKTP++S+  +++ M YN+G + DRD Y+ +C ++++    + V  IL
Sbjct: 241 REKSEFLRMTSHELKTPITSVIGMIDGMLYNVGDFADRDKYLRKCRDVLEGQAQLVQSIL 300

Query: 290 SFYSIKDL-NNDEEYLNVGDTLDEVLEKYSILVNQKKININKELLDYNIYIGKTALNIVF 348
           S   I+ L + ++E  ++  +L+E +E + +L   K + +   L +  +   K  L
Sbjct: 301 SLSKIETLASQNQELFSLKSSLEEEMEVFLVLSELKHLKVTINLEEQFVKANKVYLLKAI 360

Query: 349 SNLISNAVKYTNRNGIINIKIANDWLLIENSYDKNKISKINKILDASF------DLKLDN 402
            N+I NA  YT   G + I++ ++ L+I+N  +     +  K L   F        + D
Sbjct: 361 KNIIDNAFHYTKSGGQVMIQLKDNQLVIKNEAETLLTQQQMKQLFQPFYRPDYSRNRKDG 420

Query: 403 SNGLGLNIVKNILNKYNIKYE-ILHGENYFIFKI 435
             GLGL I   IL+++++ Y  ++  + + +F I
Sbjct: 421 GTGLGLFITHQILDQHHLAYRFVVLDQRWMVFTI 454
```

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

Example 414

A DNA sequence (GBSx0450) was identified in *S.agalactiae* <SEQ ID 1343> which encodes the amino acid sequence <SEQ ID 1344>. This protein is predicted to be response regulator (regX3). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.80   Transmembrane   50 -  66 (  50 -  66)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9665> which encodes amino acid sequence <SEQ ID 9666> was also identified.

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

```
>GP:AAD25108 GB:AF140356 VncR [Streptococcus pneumoniae]
 Identities = 131/218 (60%), Positives = 176/218 (80%), Gaps = 1/218 (0%)

Query: 5   MKILTVEDDKLIREGISEYLSEFGYTVIQAKDGREALSKFNS-DINLVILDIQIPFINGL 63
```

```
            MKIL VED+++IREG+S+YL++ GY  I+A DG+EAL +F+S ++ LV+LDIQ+P +NGL
Sbjct: 1    MKILIVEDEEMIREGVSDYLTDCGYETIEAADGQEALEQFSSYEVALVLLDIQMPKLNGL 60

Query: 64   EVLKEIRKKSNLPILILTAFSDEEYKIDAFTNLVDGYVEKPFSLPVLKARIDSLIKKNFG 123
            EVL EIRK S +P+L+LTAF DEEYK+ AF +L DGY+EKPFSL +LK R+D++ K+ +
Sbjct: 61   EVLAEIRKTSQVPVLMLTAFQDEEYKMSAFASLADGYLEKPFSLSLLKVRVDAIFKRYYD 120

Query: 124  HLEKFEYKNLSVNFNSYTAKINDEKIDVNAKELEILKCLLDNDGQVLTRMQIIDYVWKDS 183
                 F YK+  V+F SY+A +  +++ +NAKELEIL  L+ N+G+ LTR QIID VWK +
Sbjct: 121  TGRIFSYKDTKVDFESYSASLAGQEVPINAKELEILDYLVKNEGRALTRSQIIDAVWKAT 180

Query: 184  EEIPYDRVVDVYIKELRKKLQLDCITTIRNVGYKLERK 221
            +E+P+DRV+DVYIKELRKKL LDCI T+RNVGYKLERK
Sbjct: 181  DEVPFDRVIDVYIKELRKKLDLDCILTVRNVGYKLERK 218
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1345> which encodes the amino acid sequence <SEQ ID 1346>. Analysis of this protein sequence reveals the following:

```
     Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60   Transmembrane   48 -  64 (  48 -  64)

----- 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>
```

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

```
>GP:AAF72358 GB:AF192329 VanRB [Enterococcus faecalis]
 Identities = 88/215 (40%), Positives = 128/215 (58%), Gaps = 2/215 (0%)

Query: 3    KILVVEDDDTISQVICEFLKANNYDPDCVFDGQAALDKWQTTSYDLIILDIMLPSLSGLE 62
            +IL+VEDDD I   +  FL    Y  D   DG  A  K+   +Y L+ILDIMLP ++G E
Sbjct: 4    RILLVEDDDHICNTVRGFLAEAGYQVDACTDGNEAYTKFYENTYQLVILDIMLPGMNGHE 63

Query: 63   VLKTIRKTSDVPIIMLTALDDEYTQLVSFNHLISDYVTKPFSPLILIKRIENVLRVSTPD 122
            +L+  R  +D PI+M+TAL D+  Q+ +F+    DYVTKPF   IL+KR+E +LR S
Sbjct: 64   LLREFRAKNDTPILMMTALSDDENQIRAFDAEADDYVTKPFKMQILLKRVEALLRRSGAL 123

Query: 123  EKR-QIGDLLVDETEHSVYWQGTLVKLTKKEYDIIDYLAKRHQKIVTRDQLMDDIWGYS- 180
             K  ++G L +   + +V   GT + LT+KE++I+  L +   + +T + ++  IWGY
Sbjct: 124  AKEIRVGRLTLLPEDFTVLCDGTELPLTRKEFEILLLLVQNKGRTLTHEIILSRIWGYDF 183

Query: 181  ELDTRVLDNHIKNLRKKMTGIPLKTITGMGYLLGE 215
            E D   +  HIKNLR K+    +KTI G+GY L E
Sbjct: 184  EGDGSTVHTHIKNLRAKLPENIIKTIRGVGYRLEE 218
```

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

```
 Identities = 80/214 (37%), Positives = 126/214 (58%), Gaps = 4/214 (1%)

Query: 6    KILTVEDDKLIREGISEYLSEFGYTVIQAKDGREALSKFNS-DINLVILDIQIPFINGLE 64
            KIL VEDD  I + I E+L    Y      DG+ AL K+ +   +L+ILDI +P ++GLE
Sbjct: 3    KILVVEDDDTISQVICEFLKANNYDPDCVFDGQAALDKWQTTSYDLIILDIMLPSLSGLE 62

Query: 65   VLKEIRKKSNLPILILTAFSDEEYKIDAFTNLVDGYVEKPFSLPVLKARIDSLIKKNFGH 124
            VLK IRK S++PI++LTA  DE  ++ +F +L+  YV KPFS  +L  RI+++++ +
Sbjct: 63   VLKTIRKTSDVPIIMLTALDDEYTQLVSFNHLISDYVTKPFSPLILIKRIENVLRVSTPD 122

Query: 125  LEKFEYKNLSVNFNSYTAKINDEKIDVNAKELEILKCLLDNDGQVLTRMQIIDYVWKDSE 184
             EK +  +L V+   ++       + +  KE +I+  L     +++TR Q++D +W  SE
Sbjct: 123  -EKRQIGDLLVDETEHSVYWQGTLVKLTKKEYDIIDYLAKRHQKIVTRDQLMDDIWGYSE 181

Query: 185  EIPYDRVVDVYIKELRKKLQLDCITTIRNVGYKL 218
                 RV+D +IK LRKK+    + TI  +GY L
Sbjct: 182  --LDTRVLDNHIKNLRKKMTGIPLKTITGMGYLL 213
```

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

Example 415

A DNA sequence (GBSx0451) was identified in *S.agalactiae* <SEQ ID 1347> which encodes the amino acid sequence <SEQ ID 1348>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.68   Transmembrane  423 - 439 ( 413 - 447)
   INTEGRAL    Likelihood =-10.67   Transmembrane   16 -  32 (  12 -  37)
   INTEGRAL    Likelihood = -9.77   Transmembrane  303 - 319 ( 301 - 326)
   INTEGRAL    Likelihood = -3.13   Transmembrane  343 - 359 ( 343 - 367)

----- Final Results -----
               bacterial membrane --- Certainty=0.6074(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:AAD47594 GB:AF140784 Vexp3 [Streptococcus pneumoniae]
 Identities = 280/458 (61%), Positives = 363/458 (79%), Gaps = 3/458 (0%)

Query: 1   MIKNAFAYVTRKSLKSLIIILVILSMATLSIISLSIKDATDRASKETFANITNSFSMEIN 60
            M+ NAFAYVTRK  KS++I L+IL MA+LS++ LSIK AT +AS+ETF NITNSFSM+IN
Sbjct: 1   MLHNAFAYVTRKFFKSIVIFLIILLMASLSLVGLSIKGATAKASQETFKNITNSFSMQIN 60

Query: 61  RQVNPGTPRGGGNVKGEDIKKISQTNSIDSYVKRINSVADLVDHDIIETQDTLANQSPER 120
            R+VN GTPRG GN+KGEDIKKI++  +I+SYVKRIN++ DL  +D+IET +T  N + +R
Sbjct: 61  RRVNQGTPRGAGNIKGEDIKKITENKAIESYVKRINAIGDLTGYDLIETPETKKNLTADR 120

Query: 121 AKNFKRTVMLTGVNDSAKETKFVSEAYKLVEGKHLENKDKNKILMHKDLAKKNNLKVGDK 180
            AK F  ++M+TGVNDS+KE KFVS +YKLVEG+HL N DK+KIL+HKDLA K+  KVGDK
Sbjct: 121 AKRFGSSLMITGVNDSSKEDKFVSGSYKLVEGEHLTNDDKDKILLHKDLAAKHGWKVGDK 180

Query: 181 IKIKSNLFDADNEKVANETVEVEIKGLFDGHNSGGVSAAQELYENTLITDVHSAAKVYGN 240
            +K+ SN++DADNEK A ETVEV IKGLFDGHN   V+ +QELYENT ITD+H+AAK+YG
Sbjct: 181 VKLDSNIYDADNEKGAKETVEVTIKGLFDGHNKSAVTYSQELYENTAITDIHTAAKLYGY 240

Query: 241 TEDTAVYQDATFFVKGDKNLDSVIKDL-GKLDINWREYNLIKSSSNYPALQQSISGIYSI 299
            TEDTA+Y DATFFV  DKNLD V+K+L G   INW+ Y L+KSSSNYPAL+QSISG+Y +
Sbjct: 241 TEDTAIYGDATFFVTADKNLDDVMKELNGISGINWKSYTLVKSSSNYPALEQSISGMYKM 300

Query: 300 SNKLFVGSLIFAGVVVSLLLFLWMNARKKEIAVLLSLGISKLEIFGQFIIEMVFISIPAL 359
            +N LF GSL F+ ++++LLL LW+NAR+KE+ +LLS+G+ +  I GQFI E + I+IPAL
Sbjct: 301 ANLLFWGSLSFSVLLLALLLSLWINARRKEVGILLSIGLKQASILGQFITESILIAIPAL 360

Query: 360 LGSYFLAQYTADKLGNNILNKVTGDIAKQIARQSASSQLGGGAEAEGFNKTLSGLDINV- 418
            + +YFLA YTA  +GN +L  VT  +AKQ ++ + +S LGGGAE +GF+KTLS LDI++
Sbjct: 361 VSAYFLANYTARAIGNTVLANVTSGVAKQASKAAQASNLGGGAEVDGFSKTLSSLDISIQ 420

Query: 419 LPKFIIYVVIFMSFVLLVSLILSSIYTLRKNPKELLID 456
               FII  V+ +  V+LV + L+S    LRK PKELL+D
Sbjct: 421 TSDFIIIFVLALVLVVLV-MALASSNLLRKQPKELLLD 457
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1349> which encodes the amino acid sequence <SEQ ID 1350>. Analysis of this protein sequence reveals the following:

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

```
   INTEGRAL    Likelihood =-12.90   Transmembrane   19 -  35 (  16 -  43)
   INTEGRAL    Likelihood = -7.27   Transmembrane  371 - 387 ( 359 - 392)
   INTEGRAL    Likelihood = -7.01   Transmembrane  335 - 351 ( 326 - 357)
   INTEGRAL    Likelihood = -6.21   Transmembrane  282 - 298 ( 276 - 308)

----- Final Results -----
              bacterial membrane --- Certainty=0.6158(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:AAC24912 GB:AF012285 YknZ [Bacillus subtilis]
 Identities = 176/408 (43%), Positives = 250/408 (61%), Gaps = 16/408 (3%)

Query: 1   MENWKFALSSIWGHKMRSILTMLGIIIGVAAVVIIMGLGNAMKNSVTSTFSSKQKDIQLY 60
           +EN + ALSS+  HKMRSILTMLGIIIGV +V++++ +G   +  +  + S     ++LY
Sbjct: 4   LENIRMALSSVLAHKMRSILTMLGIIIGVGSVIVVVAVGQGGEQMLKQSISGPGNTVELY 63

Query: 61  FQEKGEE--EDLYAGLHTHENNHEVKPEWLEQIVKDIDGIDSYYFTNSATSTISYEKKKV 118
           +    EE   + A   +    +++K       +K I+GI     + S +    Y +++
Sbjct: 64  YMPSDEELASNPNAAAESTFTENDIKG------LKGIEGIKQVVASTSESMKARYHEEET 117

Query: 119 DNASIIGVSKDYFNIKNYDIVAGRTLTDNDYSNFSRIILLDTVLADDLFGKGNYKSALNK 178
           D A++ G++  Y N+ +  I +GRT TDND+   +R+ ++   +A +LF K    S L +
Sbjct: 118 D-ATVNGINDGYMNVNSLKIESGRTFTDNDFLAGNRVGIISQKMAKELFDK---TSPLGE 173

Query: 179 VVSLSDKDYLVIGVYKTDQTPVSFDGLSGGAVMANTQVASEFGTKEIGSIYIHVNDIQNS 238
           VV ++ +   +IGV K     +SFD LS   V  N  + S FGT +  ++ + V    +
Sbjct: 174 VVWINGQPVEIIGVLKKVTGLLSFD-LSEMYVPFN-MMKSSFGTSDFSNVSLQVESADDI 231

Query: 239 MNLGNQAADMLTNISHIKDGQYAVPDNSKIVEEINSQFSIMTTVIGSIAAISLLVGGIGV 298
            + G +AA  L N +H  +  Y V +  +I   I    +IMTT+IGSIA ISLLVGGIGV
Sbjct: 232 KSAGKEAAQ-LVNDNHGTEDSYQVMNMEEIAAGIGKVTAIMTTIIGSIAGISLLVGGIGV 290

Query: 299 MNIMLVSVTERTREIGLRKALGATRLKILSQFLIESVVLTVLGGLIGLLLAQLSVGALGN 358
           MNIMLVSVTERTREIG+RK+LGATR +IL+QFLIESVVLT++GGL+G+  +      AL +
Sbjct: 291 MNIMLVSVTERTREIGIRKSLGATRGQILTQFLIESVVLTLIGGLVGIGIG-YGGAALVS 349

Query: 359 AMTLKGACISLDVALIAVLFSASIGVFFGMLPANKASKLDPIEALRYE 406
           A+    + IS  V    VLFS  IGV FGMLPANKA+KLDPIEALRYE
Sbjct: 350 AIAGWPSLISWQVVCGGVLFSMLIGVIFGMLPANKAAKLDPIEALRYE 397
```

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

```
 Identities = 56/247 (22%), Positives = 101/247 (40%), Gaps = 42/247 (17%)

Query: 147 YKLVEGKHLENKDKNKI--------LMHKDLAKKNNLK--------VGDKIKIKSNLFDA 190
           Y +V G+ L + D +          ++  DL  K N K        + DK  +  ++
Sbjct: 136 YDIVAGRTLTDNDYSNFSRIILLDTVLADDLFGKGNYKSALNKVVSLSDKDYLVIGVYKT 195

Query: 191 DNEKVANETVEVEIKGLFDGHNSGGVSAAQELYENTLITDVHSAAKVYGNTEDTAVYQDA 250
           D   V+           FDG + G V A      NT +      A  +G  E  ++Y
Sbjct: 196 DQTPVS-----------FDGLSGGAVMA------NTQV------ASEFGTKEIGSIYIHV 232

Query: 251 TFFVKGDKNLDSVIKDL--GKLDINWREYNLIKSSSNYPALQQSISGIYSISNKLFVGSL 308
              ++   NL +   D+      I   +Y +  +S     +   S + ++   +  SL
Sbjct: 233 ND-IQNSMNLGNQAADMLTNISHIKDGQYAVPDNSKIVEEINSQFSIMTTVIGSIAAISL 291

Query: 309 IFAGVVVSLLLFLWMNARKKEIAVLLSLGISKLEIFGQFIIEMVFISIPALLGSYFLAQY 368
           + G+ V  ++ + +  R +EI +  +LG ++L+I  QF+IE V +++   L    LAQ
Sbjct: 292 LVGGIGVMNIMLVSVTERTREIGLRKALGATRLKILSQFLIESVVLTVLGGLIGLLLAQL 351

Query: 369 TADKLGN 375
           +  LGN
Sbjct: 352 SVGALGN 358
```

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

Example 416

A DNA sequence (GBSx0452) was identified in *S.agalactiae* <SEQ ID 1351> which encodes the amino acid sequence <SEQ ID 1352>. This protein is predicted to be Vexp2 (b0879). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3194(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:AAD47593 GB:AF140784 Vexp2 [Streptococcus pneumoniae]
 Identities = 142/207 (68%), Positives = 169/207 (81%)

Query: 1   MDILEIKNVNYSYANSKEKVLSGVNQKFELGKFYAIVGKSGTGKSTLLSLLAGLDKVQTG 60
           M +L++++V Y Y N+ E VL  +N  FE GKFY+I+G+SG GKSTLLSLLAGLD    G
Sbjct: 1   MTLLQLQDVTYRYKNTAEAVLYQINYNFEPGKFYSIIGESGAGKSTLLSLLAGLDSPVEG 60

Query: 61  KILFKNEDIEKKGYSNHRKNNISLVFQNYNLIDYLSPIENIRLVNKSVDESILFELGLDK 120
            ILF+ EDI KKGYS HR ++ISLVFQNYNLIDYLSP+ENIRLVNK   ++ L ELGLD+
Sbjct: 61  SILFQGEDIRKKGYSYHRMHHISLVFQNYNLIDYLSPLENIRLVNKKASKNTLLELGLDE 120

Query: 121 KQIKRNVMKLSGGQQQRVAIARALVSDAPIILADEPTGNLDSVTAGEIINILKELAQDRN 180
            QIKRNV++LSGGQQQRVAIAR+LVS+AP+ILADEPTGNLD  TAG+I+ +LK LAQ
Sbjct: 121 SQIKRNVLQLSGGQQQRVAIARSLVSEAPVILADEPTGNLDPKTAGDIVELLKSLAQKTG 180

Query: 181 KCVIVVTHSKEVADSADIILELSGKKL 207
           KCVIVVTHSKEVA ++DI LEL  KKL
Sbjct: 181 KCVIVVTHSKEVAQASDITLELKDKKL 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1353> which encodes the amino acid sequence <SEQ ID 1354>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2717(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 = 83/230 (36%), Positives = 135/230 (58%), Gaps = 13/230 (5%)

Query: 1   MDILEIKNVNYSYANSKEKVLSGVNQKFEL--GKFYAIVGKSGTGKSTLLSLLAGLDKVQ 58
           M  +E+K V+ SY    +  V +      FE+  G+    I+G SG GKST+L++LG+D V
Sbjct: 5   MAFIELKQVSKSYQIGETTVFANHEVSFEINKGELVVILGASGAGKSTVLNILGGMDTVD 64

Query: 59  TGKILFKNEDIE---KKGYSNHRKNNISLVFQNYNLIDYLSPIENIRLVNKSVDES---- 111
            G+++    +DI       K  + +R+N I  VFQ YNL+  L+  EN+ L  + V ++
Sbjct: 65  AGQVIIDGKDIAHYTSKALTQYRRNAIGFVFQFYNLVPNLTAKENVELAVEIVADALDPV 124

Query: 112 -ILFELGLDKKQIKRNVMKLSGGQQQRVAIARALVSDAPIILADEPTGNLDSVTAGEIIN 170
            IL E+GL  + +      +LSGG+QQRV+IARAL  + ++L DEPTG LD  T  +I+
Sbjct: 125 TILKEVGLSHR-LDHFPAQLSGGEQQRVSIARALAKNPKLLLCDEPTGALDYQTGKQILT 183
```

```
Query: 171 ILKELAQDRNKCVIVVTHSKEVADSADIILELSGKKLKK--VNKMNLEVE 218
           +L+++AQ +    V++VTH+  +A  AD ++ +    ++ K  +NK     +E
Sbjct: 184 LLQDMAQTKGTTVVIVTHNAAIAPIADRVIFMHDAQVTKIVINKEPASIE 233
```

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

Example 417

A DNA sequence (GBSx0453) was identified in *S.agalactiae* <SEQ ID 1355> which encodes the amino acid sequence <SEQ ID 1356>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.35   Transmembrane   17 -  33 (  17 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.2338(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 418

A DNA sequence (GBSx0454) was identified in *S.agalactiae* <SEQ ID 1357> which encodes the amino acid sequence <SEQ ID 1358>. This protein is predicted to be Vexp1. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.52   Transmembrane  294 - 310 ( 285 - 312)
   INTEGRAL    Likelihood =-10.67   Transmembrane  396 - 412 ( 385 - 417)
   INTEGRAL    Likelihood = -8.76   Transmembrane   17 -  33 (  14 -  38)
   INTEGRAL    Likelihood = -4.14   Transmembrane  335 - 351 ( 333 - 357)

----- 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:AAD47592 GB:AF140784 Vexp1 [Streptococcus pneumoniae]
 Identities = 165/425 (38%), Positives = 271/425 (62%), Gaps = 4/425 (0%)

Query: 2   IKNAIAYITRKKNRTLIIFAILTIVLSCLYSCLTIMKSSNEIEKALYESSNSSISITK-K 60
           I+ + AY++RK+ R+ I+F IL ++L+ + +CLT+MKS+  +E  LY+S N+S SI K +
Sbjct: 4   IQRSWAYVSRKRLRSFILFLILLVLLAGISACLTLMKSNKTVESNLYKSLNTSFSIKKIE 63

Query: 61  DGKYFNINQFKNIEKIKEVEEKIFQYDGLAKLKDLKVVSGEQSINREDLSDEFKNVVSLE 120
           +G+ F ++    ++ KIK +E     + + +AKLKD + V+GEQS+ R+DLS     N+VSL
Sbjct: 64  NGQTFKLSDLASVSKIKGLENVSPELETVAKLKDKEAVTGEQSVERDDLSAADNNLVSLT 123

Query: 121 ATSNTKRNLLFSSGVFSFKEGKNIEENDKNSILVHEEFAKQNKLKLGDEIDLELLDTEKS 180
           A  ++ +++ F+S  F+ KEG+++++ D   IL+HEE AK+N L L D+I L+   +E S
Sbjct: 124 ALEDSSKDVTFTSSAFNLKEGRHLQKGDSKKILIHEELAKKNGLSLHDKIGLDAGQSE-S 182
```

```
Query: 181 GKIKSHKFKIIGIFSGKKQETYTGLSSDFSENMVFVDYSTSQEILNKSENNRIANKILMY 240
           GK ++ +F+IIGIFSGKKQE +TGLSSDFSEN VF DY +SQ +L  SE    A +   Y
Sbjct: 183 GKGQTVEFEIIGIFSGKKQEKFTGLSSDFSENQVFTDYESSQTLLGNSEAQVSAARF--Y 240

Query: 241 SGSLESTELALNKLKDFKIDKSKYSIKKDNKAFEESLESVSGIKHIIKIMTYSIMLGGIV 300
             + +  + + ++++  ++   Y ++K+NKAFE+  +SV+  +  + I  Y +++ G
Sbjct: 241 VENPKEMDGLMKQVENLALENQGYQVEKENKAFEQIKDSVATFQTFLTIFLYGMLIAGAG 300

Query: 301 VLSLILILWLRERIYEIGIFLSIGTTKIQIIRQFIFELIFISIPSIISSLFLGNLLLKVI 360
            L L+L LWLRER+YE+GI L++G  K  I  QF  E++ +S+ +++ +   GN +   +
Sbjct: 301 ALILVLSLWLRERVYEVGILLALGKGKSSIFLQFCLEVVLVSLGALLPAFVAGNAITTYL 360

Query: 361 VEGFINSENSMIFGGSLINKSSFMLNITTLAESYLILISIIVLSVVMASSLILFKKPKEI 420
           ++  + S +       +L   SS   +I + AESY+ L+ +  LSV +    +  K PKEI
Sbjct: 361 LQTLLASGDQASLQDTLAKASSLSTSILSFAESYVFLVLLSCLSVALCFLFLFRKSPKEI 420

Query: 421 LSKIS 425
           LS IS
Sbjct: 421 LSSIS 425
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1359> which encodes the amino acid sequence <SEQ ID 1360>. Analysis of this protein sequence reveals the following:

```
     Possible site: 15
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.57   Transmembrane   23 -  39 ( 16 -  43)
   INTEGRAL    Likelihood =-11.36   Transmembrane  371 - 387 ( 362 - 396)
   INTEGRAL    Likelihood = -8.12   Transmembrane  331 - 347 ( 324 - 360)
   INTEGRAL    Likelihood = -7.70   Transmembrane  280 - 296 ( 277 - 308)

----- Final Results -----
              bacterial membrane --- Certainty=0.5628(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:AAB97962 GB:U96166 ATP-binding cassette transporter-like protein
           [Streptococcus cristatus]
 Identities = 222/311 (71%), Positives = 278/311 (89%)

Query: 16  MRSILTMLGIIIGIGAIIAIFSIIEGNTENTKRQLIGGSNNTINIVFNKKSSIDPKFPDK 75
           MRS+LTMLGIIIGIGAIIAIFSIIEGNTENTKRQLIGGSNNTI +V++KKS+IDP  P+K
Sbjct: 1   MRSMLTMLGIIIGIGAIIAIFSIIEGNTENTKRQLIGGSNNTIKVVYDKKSAIDPSIPEK 60

Query: 76  SNAKKPDYLPFMAEEELSKIQQVKGVKNALISYGIDDKVYHLGQKSSAKISAITKNVAEV 135
           S A+KP Y+PFM E+ LSKI+++ GVKNAL++YG D+K+Y+L QKSS+K+ A++++VA++
Sbjct: 61  SQAQKPSYIPFMGEDVLSKIKEIPGVKNALMTYGADEKIYYLSQKSSSKVQAVSQSVADI 120

Query: 136 RRMTFIKGSDFSDKDFIDQKQVIYLEKSLYESLFPKDDGLGKFVEVMGNPFRVIGVFESK 195
           ++   ++G  F  + F +Q+QV YLEKSLY++LFPK DG+GK+VEV GNPF+VIGVFES
Sbjct: 121 KQQRLLEGEGFDSEAFKNQEQVAYLEKSLYDTLFPKGDGIGKYVEVKGNPFKVIGVFEST 180

Query: 196 EQSGLTSGTEKIAYIPLHQWYNINGVVDATPEITIQTYRADDLKPVAKRVSDMLNQTIPK 255
           EQSGLTSG+EK+AYIPL QW+ I   ++ +PE+T+QT++ADDLK VAK+VSD LNQ +P+
Sbjct: 181 EQSGLTSGSEKVAYIPLQQWHRIFDTINVSPEVTVQTHKADDLKKVAKKVSDYLNQQMPQ 240

Query: 256 SDYMFGVMNLKEFERQLDNLNKSNFVLLAGIASISLIVGGIGVMNIMLVSVTERTREIGI 315
           SDYMFGV+NL+EFERQLDNLN+SNFVLLAGIASISL+VGGIGVMNIMLVSVTERTREIGI
Sbjct: 241 SDYMFGVLNLQEFERQLDNLNQSNFVLLAGIASISLLVGGIGVMNIMLVSVTERTREIGI 300

Query: 316 KKALGARRKLI 326
           KKALGARRK++
Sbjct: 301 KKALGARRKIL 311
```

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

```
 Identities = 79/386 (20%), Positives = 170/386 (43%), Gaps = 38/386 (9%)

Query: 5   AIAYITRKKNRTLIIFAILTIVLSCLYSCLTIMKSSNE-IEKALYESSNSSISITKKDGK 63
           A++ I   K R+++    + I +  + +  +I++ + E  ++ L   SN++I+I
Sbjct: 7   ALSSILSHKMRSILTMLGIIIGIGAIIAIFSIIEGNTENTKRQLIGGSNNTINIV----- 61

Query: 64  YFNINQFKNIEKIKEVEEKIFQYDGLAKLKDLKVVSGEQSINREDLSDEFKNVVSLEATS 123
            FN        K   ++ K F     AK  D      E+ +++       KN +
Sbjct: 62  -FN--------KKSSIDPK-FPDKSNAKKPDYLPFMAEEELSKIQQVKGVKNALISYGID 111

Query: 124 NTKRNLLFSSGVFSFKEGKNIEENDKNSILVHEEFAKQNKLKLGDEIDLELLDTE----- 178
           +   +L   S        KN+ E  + + +   +F+ ++ +     I LE    E
Sbjct: 112 DKVYHLGQKSSAKISAITKNVAEVRRMTFIKGSDFSDKDFIDQKQVIYLEKSLYESLFPK 171

Query: 179 -----KSGKIKSHKFKIIGIFSGKKQETYTGLSSDFSENMVFVDYSTSQEILNKSENNRI 233
                K  ++  + F++IG+F  K+Q   +GL+S  +E + ++       I    +
Sbjct: 172 DDGLGKFVEVMGNPFRVIGVFESKEQ---SGLTSG-TEKIAYIPLHQWYNINGVVDATPE 227

Query: 234 ANKILMYSGSLESTELALNKLKDFKIDKSKYSIKKDN-KAFEESLESVSGIKHIIK--IM 290
                + L+     ++ + +  I KS Y     N K FE  L+++    ++  I
Sbjct: 228 ITIQTYRADDLKPVAKRVSDMLNQTIPKSDYMFGVMNLKEFERQLDNLNKSNFVLLAGIA 287

Query: 291 TYSIMLGGIVVLSLILILWLRERIYEIGIFLSIGTTKIQIIRQFIFELIFIS----IPSI 346
           + S+++GGI V++++L+  + ER  EIGI  ++G  +  I++QF+ E + ++    +  +
Sbjct: 288 SISLIVGGIGVMNIMLVS-VTERTREIGIKKALGARRKLILKQFLIEAVILTLLGGVIGV 346

Query: 347 ISSLFLGNLLLKVIVEGFINSENSMI 372
           IS +  G ++ + +   +I S  S++
Sbjct: 347 ISGMVSGLIITRSLEYPYILSLFSVV 372
```

A related GBS gene <SEQ ID 8571> and protein <SEQ ID 8572> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 10
McG: Discrim Score:      5.59
GvH: Signal Score (-7.5): -5.97
    Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 4 value: -11.52 threshold:  0.0
   INTEGRAL    Likelihood =-11.52   Transmembrane  294 - 310 ( 285 - 312)
   INTEGRAL    Likelihood =-10.67   Transmembrane  396 - 412 ( 385 - 417)
   INTEGRAL    Likelihood = -8.76   Transmembrane   17 -  33 (  14 -  38)
   INTEGRAL    Likelihood = -4.14   Transmembrane  335 - 351 ( 333 - 357)
   PERIPHERAL  Likelihood =  4.51     315
 modified ALOM score:   2.80

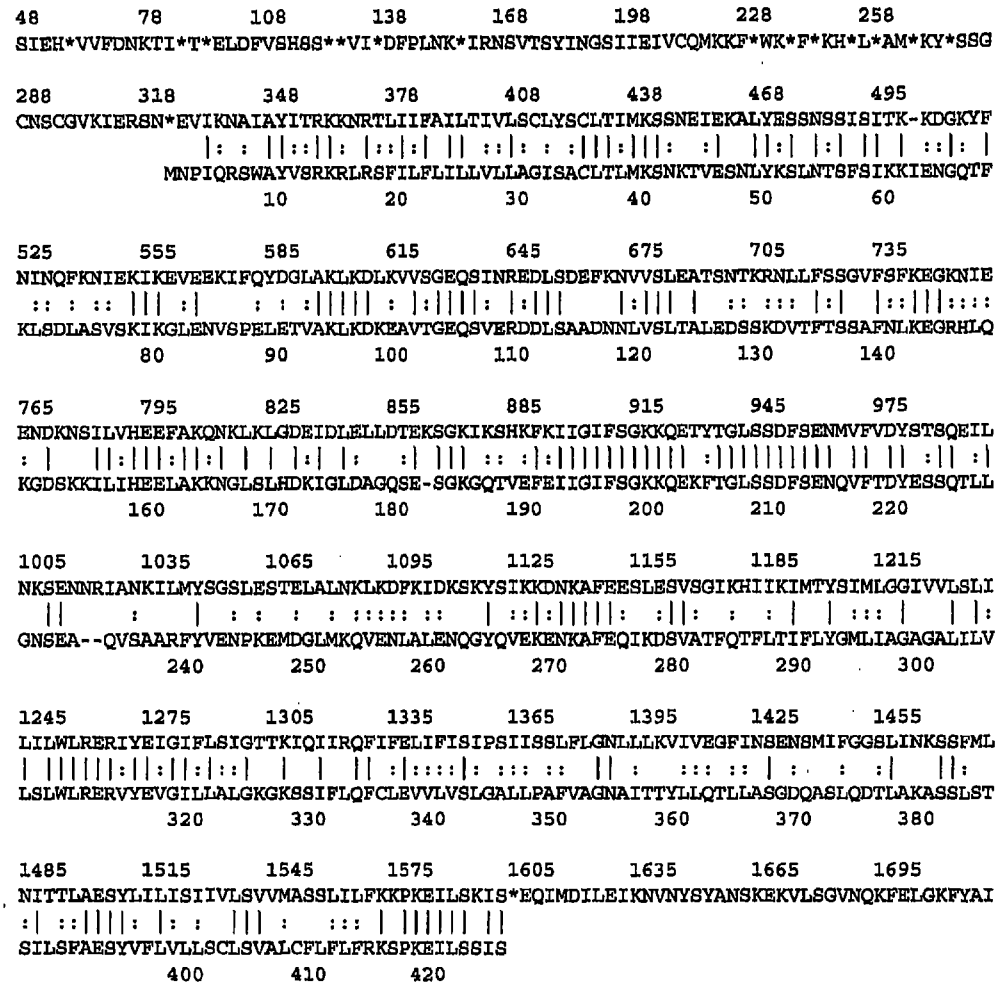
*** Reasoning Step: 3

----- 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 databases:

```
38.7/67.3% over 421aa
                                                                Streptococcus
pneumoniae
  GP|5712667| Vexp1 Insert characterized

ORF00815(304 - 1575 of 1875)
GP|5712667|gb|AAD47592.1|AF140784_1|AF140784(4   -   425   of   425)   Vexp1   (Streptococcus
pneumoniae)
%Match = 25.0
%Identity = 38.7  %Similarity = 67.2
Matches = 164  Mismatches = 136  Conservative Sub.s = 121
```



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

Example 419

A DNA sequence (GBSx0455) was identified in *S.agalactiae* <SEQ ID 1361> which encodes the amino acid sequence <SEQ ID 1362>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.04   Transmembrane   19 -  35 (  14 -  42)

----- Final Results -----
               bacterial membrane --- Certainty=0.3017(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 420

A DNA sequence (GBSx0456) was identified in *S.agalactiae* <SEQ ID 1363> which encodes the amino acid sequence <SEQ ID 1364>. 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>
```

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 421

A DNA sequence (GBSx0457) was identified in *S.agalactiae* <SEQ ID 1365> which encodes the amino acid sequence <SEQ ID 1366>. 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>
```

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

```
>GP:AAA74029 GB:U30715 ORFB [Bacillus anthracis]
 Identities = 33/76 (43%), Positives = 44/76 (57%), Gaps = 1/76 (1%)

Query: 11  IRRVSHACTKAGDRFYEENILNREFTATAHNQKWCTDVTYLQYGLGAKAYLSAIKDLYNG 70
           ++R           R   EN+LNR F A   N+KW TD+TYL +G      YL +I DLYN
Sbjct: 86  VKRKRRTWINGESRIVVENLLNRNFQANKPNEKWVTDITYLPFGT-EMLYLLSIMDLYNN 144

Query: 71  SIIAYEISHNNEIHLL 86
            IIAYEIS+  ++ L+
Sbjct: 145 EIIAYEISNRQDVTLV 160
```

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

Example 422

A DNA sequence (GBSx0458) was identified in *S.agalactiae* <SEQ ID 1367> which encodes the amino acid sequence <SEQ ID 1368>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.69   Transmembrane   10 -  26 (  10 -  26)

----- 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 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 423

A DNA sequence (GBSx0459) was identified in *S.agalactiae* <SEQ ID 1369> which encodes the amino acid sequence <SEQ ID 1370>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4170(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:AAA56999 GB:U09558 ORFA, putative Helix-Turn-Helix motif from
          amino acid 21 through 42 and from amino acid 78 through
          99 [Lactobacillus johnsonii]
 Identities = 28/116 (24%), Positives = 59/116 (50%), Gaps = 6/116 (5%)

Query: 3   YSTLAKEQGVQGYLDGKGSLRDICKWYDISSRSVLQKWIKRYTSGEDLKATSRGYSRMKQ 62
            YST  K + V  YL+ + S++ + K Y+I   +++++W+ +    + L A S  +++
Sbjct: 4   YSTELKIEIVSKYLNHEDSIKGLAKQYNIHW-TLIRRWVDK-AKCQGLAALSVKHTKTTY 61

Query: 63  GRQATFEERVEIVNYTIAHGKDYQAAIEKFGVSYQQIYSWVRKLEKNGSQGLVDRR 118
                + ++ +V Y + H          KF +S  Q+Y+W +K  + G  GL+ ++
Sbjct: 62  SS----DFKLNVVRYYLTHSIGVSKVAAKFNISDSQVYNWAKKFNEEGYAGLLPKQ 113
```

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 424

A DNA sequence (GBSx0460) was identified in *S.agalactiae* <SEQ ID 1371> which encodes the amino acid sequence <SEQ ID 1372>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.69   Transmembrane    2 -  18 (   2 -  19)

----- 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 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 425

A DNA sequence (GBSx0461) was identified in *S.agalactiae* <SEQ ID 1373> which encodes the amino acid sequence <SEQ ID 1374>. This protein is predicted to be integrase (phage-relatedpr). Analysis of this protein sequence reveals the following:

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

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

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

```
>GP:AAC79517 GB:U88974 ORF1 [Streptococcus thermophilus temperate
          bacteriophage O1205]
 Identities = 104/172 (60%), Positives = 127/172 (73%), Gaps = 11/172 (6%)

Query: 10  QHQSYAALYLIAKTGMRFAECLGLTVNDIDYTNKYLSINKTWDYHFNQRYLPTKNKSSIR 69
           ++ SYAALY+I+KTG+RFAECLGLTV+DI       LS+NKTWDY  N  ++PTK KSSIR
Sbjct: 186 EYASYAALYIISKTGIRFAECLGLTVDDIKRDTGMLSVNKTWDYKNNTGFMPTKTKSSIR 245

Query: 70  NIPIDNDTLFFLHEFTKNKNDRLFDKLSNNAVNKTIRKITGREVRVHSLRHTFASY---- 125
            IP+D++ + F+ +     + RL   LSNNAVNKT+RKI GREVRVHSLRHT+ASY
Sbjct: 246 EIPLDDEFINFIDQLPPTDDGRLLPSLSNNAVNKTLRKIVGREVRVHSLRHTYASYLIAH 305

Query: 126 ---LISISQVLDHENLNITLEVYAHQLQEQKDRNDKLNQRNLGRIWGKIALN 174
              LIS+SQVL HENLNITLEVYAHQLQEQK RND+      + ++W K   N
Sbjct: 306 DIDLISVSQVLGHENLNITLEVYAHQLQEQKSRNDE----KIKQMWTKCGQN 353
```

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 426

A DNA sequence (GBSx0462) was identified in *S.agalactiae* <SEQ ID 1375> which encodes the amino acid sequence <SEQ ID 1376>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3206(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 1328.

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

Example 427

A DNA sequence (GBSx0463) was identified in *S.agalactiae* <SEQ ID 1377> which encodes the amino acid sequence <SEQ ID 1378>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
```

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

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.6542(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:CAB52541 GB:AJ131519 hypothetical protein [Lactobacillus
          bacteriophage phi adh]
 Identities = 24/55 (43%), Positives = 36/55 (64%)

Query: 12 MDKELTPQEKANKKWAENNREHRTYLSKRSTARSFINKNATKEDLLELKQLIESK 66
          M K    + KANKKW E N+  + Y++KRSTA+SFI   AT+EDL  +++ +  +
Sbjct: 1  MAKITEARAKANKKWDEKNKARKLYINKRSTAKSFILNLATEEDLANIEEYVAER 55
```

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 428

A DNA sequence (GBSx0464) was identified in *S.agalactiae* <SEQ ID 1379> which encodes the amino acid sequence <SEQ ID 1380>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4417(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 1332.

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

Example 429

A DNA sequence (GBSx0465) was identified in *S.agalactiae* <SEQ ID 1381> which encodes the amino acid sequence <SEQ ID 1382>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have a cleavable N-term signal seq.

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

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 430

A DNA sequence (GBSx0466) was identified in *S.agalactiae* <SEQ ID 1383> which encodes the amino acid sequence <SEQ ID 1384>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.30   Transmembrane  205 - 221 ( 202 - 223)
   INTEGRAL    Likelihood = -3.56   Transmembrane  296 - 312 ( 294 - 312)

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

A related GBS nucleic acid sequence <SEQ ID 9663> which encodes amino acid sequence <SEQ ID 9664> 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*.

A related GBS gene <SEQ ID 8573> and protein <SEQ ID 8574> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     -8.80
GvH: Signal Score (-7.5): -4.03
     Possible site: 47
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value:  -4.30 threshold:  0.0
   INTEGRAL    Likelihood = -4.30   Transmembrane  205 - 221 ( 202 - 223)
   INTEGRAL    Likelihood = -3.56   Transmembrane  296 - 312 ( 294 - 312)
   PERIPHERAL  Likelihood =  2.97       20
 modified ALOM score:   1.36

*** Reasoning Step: 3

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

SEQ ID 8574 (GBS366) was expressed in *E.coli* as a GST-fusion product. The purified fusion protein (Figure 215, lane 5) was used to immunise mice. The resulting antiserum was used for FACS (Figure 281), 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 431

A DNA sequence (GBSx0467) was identified in *S.agalactiae* <SEQ ID 1385> which encodes the amino acid sequence <SEQ ID 1386>. This protein is predicted to be N-acetylmuramoyl-L-alanine amidase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1471(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 8575> which encodes amino acid sequence <SEQ ID 8576> was also identified. This has an RGD motif at residues 81-83.

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

```
>GP:CAB07986 GB:Z93946 N-acetylmuramoyl-L-alanine amidase
          [bacteriophage Dp-1]
 Identities = 99/140 (70%), Positives = 120/140 (85%)

Query: 10  MVINIEQAIAWMASRKGKVTYSMDYRNGPSSYDCSSSVYFALRSAGASDNGWAVNTEYEH 69
           M ++IE+ +AWM +RKG+V+YSMD+R+GP SYDCSSS+Y+ALRSAGAS  GWAVNTEY H
Sbjct: 1   MGVDIEKGVAWMQARKGRVSYSMDFRDGPDSYDCSSSMYYALRSAGASSAGWAVNTEYMH 60

Query: 70  DWLIKNGYVLIAENTNWNAQRGDIFIWGKRGASAGAFGHTGMFVDPDNIIHCNYGYNSIT 129
            WLI+NGY LI+EN  W+A+RGDIFIWG++GASAGA GHTGMF+D DNIIHCNY Y+ I+
Sbjct: 61  AWLIENGYELISENAPWDAKRGDIFIWGRKGASAGAGGHTGMFIDSDNIIHCNYAYDGIS 120

Query: 130 VNNHDEIWGYNGQPYVYAYR 149
           VN+HDE W Y GQPY Y YR
Sbjct: 121 VNDHDERWYYAGQPYYYVYR 140
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1387> which encodes the amino acid sequence <SEQ ID 1388>. Analysis of this protein sequence reveals the following:

```
      Possible site: 26

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.06   Transmembrane   79 -  95 (  77 -  95)

----- 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>
```

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

```
 Identities = 56/91 (61%), Positives = 68/91 (74%)

Query: 158 KVDNQSVVSKFEKELDVNTPLSNSNMPYYEATISEDYYVESKPDVNSTDKELLVAGTRVR 217
           K+D       F ++LD NT L NSN+PYYEAT+  DYYVESKP+ +S DKE + AGTRVR
Sbjct: 354 KIDKPQSQLTFNQKLDTNTKLDNSNVPYYEATLRTDYYVESKPNASSADKEFIKAGTRVR 413

Query: 218 VYEKVKGWARIGAPQSNQWVEDAYLIDATDM 248
           VYEKV GW+RI A QS+QWVED YL +AT +
Sbjct: 414 VYEKVNGWSRINASQSDQWVEDKYLSNATQV 444
```

SEQ ID 8576 (GBS301) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 9; MW 30kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 3; MW 55kDa).

The GBS301-GST fusion product was purified (Figure 205, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 300), 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 432

A DNA sequence (GBSx0468) was identified in *S.agalactiae* <SEQ ID 1389> which encodes the amino acid sequence <SEQ ID 1390>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.53   Transmembrane    8 -  24 (   3 -  25)

----- 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>
```

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 433

A DNA sequence (GBSx0469) was identified in *S.agalactiae* <SEQ ID 1391> which encodes the amino acid sequence <SEQ ID 1392>. 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>
```

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 434

A DNA sequence (GBSx0470) was identified in *S.agalactiae* <SEQ ID 1393> which encodes the amino acid sequence <SEQ ID 1394>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0120(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 435

A DNA sequence (GBSx0471) was identified in *S.agalactiae* <SEQ ID 1395> which encodes the amino acid sequence <SEQ ID 1396>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4757(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 9661> which encodes amino acid sequence <SEQ ID 9662> was also identified.

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

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

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

Example 436

A DNA sequence (GBSx0472) was identified in *S.agalactiae* <SEQ ID 1397> which encodes the amino acid sequence <SEQ ID 1398>. 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
   INTEGRAL    Likelihood = -2.39   Transmembrane  349 - 365 ( 347 - 366)

----- 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:AAF43531 GB:AF145054 ORF39 [Streptococcus thermophilus
          bacteriophage 7201]
 Identities = 212/666 (31%), Positives = 323/666 (47%), Gaps = 52/666 (7%)

Query: 10  WGNNLTLEILSAWNKP---NIASNTSTVNVQVFL-----KMSSYGYISIGETRPLKITVD 61
           W NN     +  W       +I +NTS V +++ L       + Y   +  E    ++
Sbjct: 5   WSNNDRGYRIRLWVDQVGQDIQNNTSQVRLRLSLLNTTTTFAQYSCSAFVEFNGQRLNWS 64

Query: 62  GRAETINVNPSINYGQRKLLFAKDYIVNHNSDGNKPLFNISAYYPIN--FSNYGEATANQ 119
           G    +  N +I       L  +   V H  DG+  +F + A++  +  +S       NQ
Sbjct: 65  GSPSVLGWNQTIQ------LIDQTITVRHADDGSG-VFGVHAHFNGSGGWSPGNLDIGNQ 117

Query: 120 SISLPKINRLSVSSAISGVLGNAVTITINRYSTSFTHNLKYDFKGSTGTIATGVGTSYLW 179
            I+L  I R S       G +GN V I+I+R      TH L+Y ++   G IA  VGTSY W
Sbjct: 118 QITLTTIPRGSSVRVSDGFIGNQVDISIDRKIGGATHTLRYAWENKQGKIADNVGTSYKW 177

Query: 180 TIPPTFANLLPNELTGTGNLIVETMDGSAKIGETKYTLSITIPNTATYKPKLSSITLSDT 239
           TIP  FAN +PN  +G G + V+T       I      TL+ ++  T   KP  +  TL+DT
Sbjct: 178 TIPEDFANDIPNSTSGRGTIYVDTYINGNFIQTQSTTLTASV-ITNNLKPSFTGFTLTDT 236

Query: 240 NTLTSSIVSG-NNFVRIISKVKVDFGSAIGNNGSTITSYNAEIVGKSNSIIGNGSVFDKL 298
           N  +  IV G  +FV I+S VKV F  A   +G+TI  Y AEIVG +NSI  NG V  ++
Sbjct: 237 NPTSQRIVPGQTHFVSIMSLVKVVFNGAQAKSGATIVGYYAEIVGANNSISSNGGVLREV 296
```

```
Query: 299 DFFGSA--TIRATVTDSRGLTSEPVDTKINVIDYFLPIVTSAKVVRSQQNPDILQVLPFV 356
                 T+R  V DSRG+ S+ V+TK+  + YF P +   +V RS +  DIL +  F
Sbjct: 297 SVNQDTEMTLRGRVQDSRGIWSDWVETKLTFLFYFSPAL-RFEVKRSDKKLDILTIKRFA 355

Query: 357 KIAPIIVGGIQKNQLKMSVSVAPYNTGIYAVDSGAATNTWSTISQMSGAPLNLGGTYDKS 416
           KIAP+ V GIQ+N +K++ S A      + VD+G A   WS+IS+ + +   LG +Y
Sbjct: 356 KIAPLSVNGIQRNVMKLTFSTAKVGWDNFVVDNGQAGGVWSSISEFNASDAKLGNSYPAD 415

Query: 417 KSWLVKISVSDNLMSATPIIQPVASEFVLVTKAPSGVAFGKIWEHGIIDAKGDVYVDGTI 476
            S++V   + D   S T     V ++ V++T    GV  GK  E G +D  GD      I
Sbjct: 416 TSYVVIGKLEDEFTS-TSFQATVPTDEVIMTYDRQGVGIGKYRERGALDVNGD------I 468

Query: 477 YCGDKAIQQKPLALNNGGSFRHDDTDLNSLQDTGFYCVFRGANRPAGAGPGYVTVVRHET 536
           Y  +  IQQ  L  NNG     ++   N+++D G Y +F  A  P      +  +  H +
Sbjct: 469 YANNSPIQQYQLTNNNGSPKMTNNA--NTIEDPGQYYLFSAA--PGNPSGQWGHLFHHSS 524

Query: 537 ------ANYAYQQFYDRTNKTI-----FTRLLENGVWSGWSEYVKKD--SLQTTGWITIG 583
                 A Q F+    +       ++R++++  W  W E+ + D  +L  TGW   G
Sbjct: 525 YGKGSMYKEAIQIFWSNDGRLFSRHHRWSRIIDD--WEPWKEFARNDNTNLINTGWQPAG 582

Query: 584 -NGFKYKRKGDDIDLMYNFASNGLQRWSVGNMPSGLI--PQELMFAITGWTLAPDKSIHL 640
            +G  YKR GD + + +NF   G   + + ++P  +   PQ  MF +TGW++  +K  ++
Sbjct: 583 VDGSFYKRVGDVLTIKFNFTGTG-GDFLLASVPPEIFKAPQSYMFVVTGWSVWANKQYNV 641

Query: 641 QINASG 646
           Q+N  G
Sbjct: 642 QVNEGG 647
```

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

SEQ ID 1398 (GBS365) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 82 (lane 2; MW 102kDa).

GBS365-GST was purified as shown in Figure 216, lane 11.

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

Example 437

A DNA sequence (GBSx0473) was identified in *S.agalactiae* <SEQ ID 1399> which encodes the amino acid sequence <SEQ ID 1400>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3481(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 = 504/998 (50%), Positives = 675/998 (67%), Gaps = 56/998 (5%)

Query: 1   MLTIHGPDLKPVLFLDNDKQGALNYFNHKWYRKQKTGSSVLEFSVYKKDLLGDSPLSHKY 60
           +LTIH  +L+ V ++DN+KQ  LN+FN KW R  ++G+SV EFSV+KK +  DS +   Y
Sbjct: 2   LLTIHDNNLQKVAYIDNEKQSTLNFFNDKWTRSLESGTSVFEFSVFKKSIKSDSKVEISY 61

Query: 61  HVLNDQAFVSFVHKGKVQLLNIMKIDEDEKQIDCYCENLNLELLNEYCNAYKATKAMSFE 120
```

```
              LN++AFVSF HKGK  L N+MKI+EDE+ I CYCENL+LELL EY  AYKA+K M+F+
Sbjct: 62   KYLNERAFVSFKHKGKSYLFNVMKIEEDEQIIRCYCENLSLELLLEYRGAYKASKPMTFK 121

Query: 121 EYLVQFDILSWGALTVGTNEVKDKKLTLEWTSQETKLARLLSIANNFDAEIEFETKLNFN 180
            EY   + +  + LT+G NEV D+K TLEW  QET LARL+S+A NFDAEIEF+T+L  N
Sbjct: 122 EYFDDWGMGQFAKLTLGVNEVSDQKRTLEWEGQETTLARLISLARNFDAEIEFDTRLKPN 181

Query: 181 HTFKQLIINIYKEYEEGKSYGVDRDKTDVILRYQKNISGIRKTVDKRQIYNAIRPYGKK- 239
               + ++N+YK Y+ GK+ GV R ++DVIL+Y KNI+GI+++VDK QIYN I PYG+K
Sbjct: 182 SQLDEFVLNVYKAYD-GKNQGVGRRRSDVILKYGKNINGIKRSVDKTQIYNMITPYGRKS 240

Query: 240 -TVRGERVISNPVTRKVTKTVGSNRT---YLGGDLKYYGHTIKKANVQAIINYAVQYNIL 295
             T +  + IS+PVT +     V S R    Y GGDL Y GHT+  + VQ I N  VQ N+L
Sbjct: 241 DTKKETKRISDPVTIQNPVVVPSARVEKRYAGGDLTYAGHTLSASLVQTIFNLCVQRNLL 300

Query: 296 PSGIITQLYLESFWGDSTVGKRDNNWAGMSGGAQTRPSGVKVTTGMARPANEGGTYMHYA 355
            PSG+I+QLYLESFWG S V +RDNNW+GM+GGAQTRPSGV VTTG  RPA+EGGTYMHYA
Sbjct: 301 PSGVISQLYLESFWGSSNVARRDNNWSGMTGGAQTRPSGVVVTTGSPRPASEGGTYMHYA 360

Query: 356 SVDDFLKDYTYLLAKQG-----IYNVVGKKNIADYTKGLFRAGGAKYDYAAAGYQSYTNL 410
            SVDDF+KDYTYLLA Q       +Y V GK+NI +YTKGLFR GGA YDYAAAGY  Y  L
Sbjct: 361 SVDDFMKDYTYLLADQTSGGRKMYGVKGKQNIEEYTKGLFRIGGALYDYAAAGYNHYIYL 420

Query: 411 MTNIRNGINKVTGNILNTIDKLWQTPVKPITAVNVARRATKTIQA------INEATKLKG 464
            M +IRNGIN+  GNIL+ +D LW+ P   IT  N  ++ T+T++A       +NE  LKG
Sbjct: 421 MRDIRNGINRSNGNILDKLDDLWRQPDNQITQPN--KQVTRTVKADRVIAVLNEMQGLKG 478

Query: 465 RRIGSGQCYALSGWYAKKLDGAWIDSSIGGIRGRIGGGMAAALIGTDYNWGAYGWKVDKS 524
            RR+G+GQCYAL+ WY+ KL G  + + + G  G IG GMAAA IGTDY W  +GW V +
Sbjct: 479 RRVGNGQCYALAAWYSMKLGGPGLGAGVTGKSGVIGAGMAAAKIGTDYAWDRFGWSVVRP 538

Query: 525 PNAGNLKAGGIYNVRANRGAPFYTTGWGHTGIIKSVSKTRVTVLEQNFVGRMYVVENSYD 584
             +  LK G I N++A       T+ WGH  II S + + VTVLEQN+ GR YVV+NSY
Sbjct: 539 TSVDQLKPGAIANIKAYNSY-LGTSVWGHVSIIISNNGSTVTVLEQNYAGRQYVVQNSYP 597

Query: 585 INSFASGLQTVCYPREIAQGMSVNGATTQQVSGGTQISYEEVVQEAQTESYEEEQIIYID 644
             +++   ++T+CYP E+ +G +V G T    +   ++   E+       + E    + ID
Sbjct: 598 ASAYLGAVETLCYPPELKEGKTVEGRTETVSTPNVEVQKVEIPPIDVEVTTESTAALTID 657

Query: 645 NSIYKEWKDENGKVEYYLKNGFLYAPLSRDRYPSVLTGNETRDNWIRKDMEVETDSQEVL 704
             +  +EW++ENG+VE+YL+NG LYAP+S++ YPS+LTG E  DNWIRKDME++TDS++VL
Sbjct: 658 SKRKQEWRNENGQVEFYLENGSLYAPISKELYPSILTGKENGDNWIRKDMEIDTDSEDVL 717

Query: 705 MSTGLKDLKAHAYPAITYEVDGYVDLELGDVVRIQDDGYEPPLILTARVVEQEISITNPS 764
            +ST L++L+   YPAITYEVDG++DL++GD V+IQD G+ P L+L ARV EQ+IS TNP
Sbjct: 718 ISTALRNLRKFCYPAITYEVDGFLDLDIGDTVKIQDTGFSPMLMLEARVSEQQISFTNPV 777

Query: 765 SNKTKFSNFVEKESQLASDLISDMLRLYDESIPYEIKLATSNGVAFKNGTGESVLTPSLQ 824
             NKT F+NF   +++++  L+S M +L +E+IPYE+KL+T NG  FKN TG+SVL  +L+
Sbjct: 778 ENKTVFANFQTLQNKVSDSLLSRMTKLAEEAIPYELKLSTDNGTTFKNSTGQSVLKATLE 837

Query: 825 KNGKDYEAVYFYKNGDSLIDIGPSLIVKASDFNHVLNITVEAYLNEELVASTQISFTDTE 884
             KNG+ Y+ ++F+KNGDS+I  G  L+VK +DF + L +TVEAYL++ELVAS +I+FTD
Sbjct: 838 KNGEVYQPIFFFKNGDSIIGTGNQLVVKPTDPENTLQVTVEAYLDDELVASAEITFTDVS 897

Query: 885 DGADGKDGAPGPQGPPGVNGLQGPKGDQGIQGPAGADGKATYTHIAYALDENGSTGFSVS 944
            DG                QGPKGD G+                 L    S G+
Sbjct: 898 DGK----------------QGPKGDDGVS-------------PINLIIESSNGYQFK 925

Query: 945 DNVGKTYI--GMYVDDNIIDSNDPK-KYKWNLIKGADG 979
             +N+  T     +Y D+  ID + +  Y W+ +  ADG
Sbjct: 926 NNIINTTFTAKLYQDNKEIDKDGTRYAYLWSKV-NADG 962
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1401> which encodes the amino acid sequence <SEQ ID 1402>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
```

```
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -3.56   Transmembrane  325 - 341 ( 323 - 343)

----- Final Results -----
               bacterial membrane --- Certainty=0.2423(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 = 23/55 (41%), Positives = 27/55 (48%)

Query: 886 GADGKDGAPGPQGPPGVNGLQGPKGDQGIQGPAGADGKATYTHIAYALDENGSTG 940
           G  GKDGAPG  G PG  G +G +G+ G QGP G  G+   T         G  G
Sbjct: 181 GEAGKDGAPGKDGAPGEKGEKGDRGETGAQGPVGPQGEKGETGAQGPAGPQGEAG 235
 Identities = 48/151 (31%), Positives = 58/151 (37%), Gaps = 19/151 (12%)

Query: 852 KASDFNHVLNITVEAYLNE--ELVASTQISFTDTEDGADGKDGAPGPQGPPGVNGLQGPK 909
           K  DF   L    E  L E  +L+  + I     + G  G  G  GPQG  G  G QGPK
Sbjct: 82  KEEDFQKELKDFTEKRLKEILDLIGKSGIK---GDRGETGPAGPAGPQGKTGERGAQGPK 138

Query: 910 GD---QGIQGPAGADGKATYTHIAYALDENGSTGFS----VSDNVGKTYIGMYVDDNIID 962
           GD   QGIQG AG  G+           E G  G +         GK       D
Sbjct: 139 GDRGEQGIQGKAGEKGERGEKGDKGETGERGEKGEAGIQGPQGEAGK-------DGAPGK 191

Query: 963 SNDPKKYKWNLIKGADGARGIQGPAGADGKT 993
              P +      +G  GA+G  GP G  G+T
Sbjct: 192 DGAPGEKGEKGDRGETGAQGPVGPQGEKGET 222
 Identities = 25/50 (50%), Positives = 29/50 (58%), Gaps = 9/50 (18%)

Query: 884 EDGADGKDGAPGPQGPPGVNGL---------QGPKGDQGIQGPAGADGKA 924
           +DGA GKDGAPG +G  G  G           QG KG+ G QGPAG  G+A
Sbjct: 185 KDGAPGKDGAPGEKGEKGDRGETGAQGPVGPQGEKGETGAQGPAGPQGEA 234
```

SEQ ID 1400 was expressed in four different forms. SDS-PAGE analysis of total cell extract is shown in Figure 122 (GBS105dN – lane 5 & 7; MW 102kDa), Figure 122 (GBS105dC – lane 8-10; MW 81kDa), Figure 179 (GBS105d – lane 8; MW 102kDa) and in Figure 181 (GBS105C – lane 2; MW 56kDa). GBS105dN-His was purified as shown in Figure 232 (lanes 9 & 10). GBS105dC-His was purified as shown in Figure 233 (lanes 3 & 4).

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

Example 438

A DNA sequence (GBSx0474) was identified in *S.agalactiae* <SEQ ID 1403> which encodes the amino acid sequence <SEQ ID 1404>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2502(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:AAC34412 GB:AF158600 putative minor structural protein
          [Streptococcus thermophilus bacteriophage Sfi11]
 Identities = 163/433 (37%), Positives = 244/433 (55%), Gaps = 21/433 (4%)
```

```
Query: 80  LSSKKPKMLMFSHIPGRYYLAVQVGDLNFKEIKMNGFGEIT--FIVADAYAHSTSYRRIK 137
           L +KK   L     P RYYLA+  G+++ K I  + + E T  F+V D  AHST+Y+R+
Sbjct: 93  LHTKKAVKLFLPTEPERYYLALVKGEVSLKGIS-DWYDEATIEPLVPDGVAHSTTYKRVT 151

Query: 138 DYTQDGNKMTFKIKNNGTAPAFPIFRIKHLGENGYIGITNETGAFAVGSPEEEDGTIVHR 197
           DY +   KM F I N G+  A+PI  +K   ENGY G+ ++  AF  G+ EE DG I+ +
Sbjct: 152 DYQEKDGKMIFSIDNEGSTDAYPIITLKANAENGYYGLVSDKFAFEAGNIEEADGKIISK 211

Query: 198 NETLFDY-SKAIAQAL-EGAPNVAKLNYMPPTFDSELKRMRLDNILGSGKGGEYVAIGAR 255
            E L+D+    I QA  +GA NV   N       +    + + N+ G         IG +
Sbjct: 212 AEVLYDFRDDRIPQAFAKGAKNVGITNVTGDLHGT----LEIQNVWGRPH------IGLK 261

Query: 256 GTTPGYGE-HVGTRTFIINPDSNGEY-TLNEHLWWKQIFIATAQDQKGFLKLCVTGENDE 313
                +   + T  I PDS+G    LNE++WW+QIF A +  Q GFLKL V+  +
Sbjct: 262 NPNANINQLQTASLTLDIPPDSSGNVGALNEYIWWRQIFWAGSISQYGFLKLTVSDADGN 321

Query: 314 FLYGIETYKRKNGFETEYNFFALDDDGVGWRFYKQFEFQA-DRNYHNPFSMNRSRAVEIF 372
           FLYG+ET+KR  G E+EYN  A D  G G+RF KQ+ F A +   HNPF+  R  + +I
Sbjct: 322 FLYGVETFKRSLGLESEYNALASDGYG-GFRFLKQWSFLATEYEDHNPFNEPRGWS-DIK 379

Query: 373 REEDKFRIYFNGAHHHVTVPSLKGKKSRKIHLAMGTCSDSSKYINYNLFEKVNFEKMGVS 432
           RE+DK   Y+ G ++  T+P +KGKKS KIHL +     S  ++ +  F+++ + K   +
Sbjct: 380 REDDKVTFYWWGTYNTFTIPEIKGKKSAKIHLTISNI-PSKSFVTHAYFDQLLYIKTNNA 438

Query: 433 HYNNIVNKYQPGDEVIINFENDTVSTKDIDSIQDVVLGSKMISIPPGESELVVHLSSWVA 492
            + +I N+Y  G  +IIN E+DT++  ++ ++ ++V GS    IPPGES++ V  S W
Sbjct: 439 FFEDIPNRYIQGSNLIINSEDDTLTLNNLLNLDEIVDGSLWPVIPPGESQIEVVQSPWAK 498

Query: 493 ALPDISIDFEERY 505
             P ++I+FEER+
Sbjct: 499 KKPSVTIEFEERW 511
```

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 439

A DNA sequence (GBSx0475) was identified in *S.agalactiae* <SEQ ID 1405> which encodes the amino acid sequence <SEQ ID 1406>. This protein is predicted to be PblA. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.11   Transmembrane  427 - 443 ( 424 - 445)
   INTEGRAL    Likelihood = -4.99   Transmembrane  449 - 465 ( 448 - 469)
   INTEGRAL    Likelihood = -2.71   Transmembrane   41 -  57 (  38 -  57)
   INTEGRAL    Likelihood = -0.37   Transmembrane  361 - 377 ( 361 - 377)
   INTEGRAL    Likelihood = -0.22   Transmembrane  324 - 340 ( 324 - 340)

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

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

```
>GP:AAG18638 GB:AY007505 PblA [Streptococcus mitis]
 Identities = 233/401 (58%), Positives = 296/401 (73%), Gaps = 17/401 (4%)

Query: 1   MATNLGQAYVQIMPSAKGISGSISKTLDPEASSAGSSAGSLLGGKLIGILGSVIAAAKIG 60
           MAT + QAYVQ++PSA+GI+G I   L+PEAS+AG SAG  LG  L+G++  VIAAA IG
```

```
Sbjct: 1   MATEIAQAYVQLIPSARGITGKIQSILNPEASAAGQSAGQSLGSSLVGVMTKVIAAAGIG 60

Query: 61  EMVTKAISSSISEGAALQQSLGGVETLFKSNANLVKKYADEAYKTTGLSANAYMESVTGF 120
            KA S++ISEGAALQQSLGG+ETLFK +A+ VK YA+EAYKTTGLSANAYME+VTGF
Sbjct: 61  ----KAFSAAISEGAALQQSLGGIETLFKGSADKVKGYANEAYKTTGLSANAYMENVTGF 116

Query: 121 SASLLQSLGGDTAKAAKVANMAMIDMADNSNKMGTSMESIQYAYQGFAKQNYTMLDNLKL 180
            SASLLQSLGGDT KAA+ ANMAMIDM+DN+NKMGTSMESIQ AYQGFAKQNYTMLDNLKL
Sbjct: 117 SASLLQSLGGDTNKAAETANMAMIDMSDNANKMGTSMESIQMAYQGFAKQNYTMLDNLKL 176

Query: 181 GYGGTQEEMKRLLSDAQKLTGKKYDISNLSDVYEAIHAIQGKIGITGTTAKEAATTFTGS 240
            GYGGT++EM+RLL+DA+KLTG KYDI+NLSDVY AIHAIQ + ITGTTAKEAA+TF+GS
Sbjct: 177 GYGGTKQEMQRLLADAEKLTGVKYDINNLSDVYSAIHAIQENLDITGTTAKEAASTFSGS 236

Query: 241 FEAMKAASKNLLGKMALGEDIKPSLKALFDTTSNFVLNNFIPMLTNVFKGFGSVISLTFS 300
            FE+MKAA++N+LGK+ALGE+I PSL AL  TTS F+ +NF+PM+ NVF G G V++   S
Sbjct: 237 FESMKAAAQNVLGKLALGENILPSLHALLKTTSTFLFDNFLPMIGNVFSGLGLVLTEGIS 296

Query: 301 ELIPKIVGFMQTSGPSLMQSGISFIISFVNGFLTAYPAFLTVAGKIFTDFVSFVMQSIPG 360
            ++  ++ G           S +   +S + G    +  F  + G +       ++ +I G
Sbjct: 297 QIASQLFG-------DAFGSAVFDQLSRITGIFETF--FDMIFGSLSKQDNIDILNTI-G 346

Query: 361 LLQAGATLVLNLIDGILANLPQIATS---AVSVISSFISML 398
              +  AT ++N+ D I      I ++    V ++  F+  L
Sbjct: 347 FSEEAATQIVNIADNIRVTFENIGSAIGDVVGIVGDFVGDL 387
 Identities = 112/386 (29%), Positives = 172/386 (44%), Gaps = 18/386 (4%)

Query: 235 TTFTGSFEAMKAASKNLLGKMA-LGEDIKPSLKA---LFDTTSNFVLNNFIPMLTNVFKG 290
            TT+    E++KA    ++   + L E IK +      L  T    V+  FI    N++
Sbjct: 580 TTWNAYVESLKAMWNAVVTFFSDLWESIKEAASTAWTLITTAVMMVVQPFIDGFMNIWNN 639

Query: 291 FGSVISLTFSELIPKIVGFMQTSGPSLMQSGISFIISFVNGFLTAYPAFLTVAGKIFTDF 350
               ++  +  +    G  +    S+    I  II  V G      A L++  +   + 
Sbjct: 640 ISEGLTQVWEGIKLIFEGAWEFI-KSIFLGAILIIIDLVTGNFGQLGADLSLIWEGIKNG 698

Query: 351 VSFVMQSIPGLLQAGATLVLNLIDGILANLPQIATSAVSVISSFISMLQANYPAILKKGF 410
            +S + + I          +++    G+  N   + ++    I +  SM    +  I
Sbjct: 699 ISLIWEGIKTYFSGVVDVIVGYATGVFENFSNVLSTIWEFIKTAASMA---WEWIKSTVS 755

Query: 411 EILSYLVQGIIARLPDIVITVGKL---IAILAGAIASNLPKVLALGVQLLITFVKGILSV 467
            +++ L+QG       + V  +  L    I  A A  S L K L  L  G  +  F+ G  S 
Sbjct: 756 NLITGLIQGAQNLWNNFVSFLSGLWENIKSTASAAWSGL-KSLVLG--FINGLVSGAQTA 812

Query: 468 IGKINETANNIGEK---LINAIKSIDLLSAGRAIMRGFLRGLEDVWGDIQNFVGDIAGWI 524
              + +  +++  K    + N IK+I+L  AG+AI+ GFL GL+  W  + NFVG IA WI
Sbjct: 813 WNNMKQAVSDLVTKVTNIFNGIKNINLWEAGKAILNGFLGGLKSAWEGVTNFVGGIANWI 872

Query: 525 KDHKGPISYDRRLLIPAGNAIMQGLHQGLVDKFKPVKNLVNGMAEEIQSSFGNPQLAFDM 584
            +DHKGPI YDR+LLIPAGNAIM  L  GL D FK VK  V GM+ EI    F     L  +
Sbjct: 873 RDHKGPIEYDRKLLIPAGNAIMGSLDNGLKDGFKDVKKTVGGMSGEISDVFSGDNLDLNS 932

Query: 585 DTNVNNGFE-RIGTLNKNLSSQVTST 609
              +V    E R+   +  L  Q + T
Sbjct: 933 TASVTKNLEARLAMPSAQLEVQESKT 958
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1407> which encodes the amino acid sequence <SEQ ID 1408>. Analysis of this protein sequence reveals the following:

```
     Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.76   Transmembrane  458 - 474 ( 458 - 474)
   INTEGRAL    Likelihood = -2.60   Transmembrane  483 - 499 ( 482 - 499)
   INTEGRAL    Likelihood = -2.02   Transmembrane  429 - 445 ( 429 - 445)
   INTEGRAL    Likelihood = -1.28   Transmembrane  397 - 413 ( 397 - 413)
   INTEGRAL    Likelihood = -0.53   Transmembrane  739 - 755 ( 738 - 755)
   INTEGRAL    Likelihood = -0.27   Transmembrane  356 - 372 ( 356 - 372)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAB18717 GB:U38906 ORF42 [Bacteriophage r1t]
 Identities = 261/579 (45%), Positives = 359/579 (61%), Gaps = 63/579 (10%)

Query: 184 MKRLLSDAEKLPAAMGKKFDLSNYADVVEAIHLVQDNMGIAGVAAEEAKTTFSGSLAAMK 243
           M+RLL+DA+KL    G+K+D+SN++D+ +AIH +Q  M I G  A+EA TTFSGS  +MK
Sbjct: 1   MQRLLTDAQKLT---GQKYDISNFSDITQAIHAIQTEMDITGTTAKEASTTFSGSFDSMK 57

Query: 244 SSFTNVMAGLSLGDDIRPALRGLAETTSNFLFGNFIPMVANIFKGLPSAIGTFIGAAAPI 303
           ++ +NV+  LSLG D++  L  L  TTS FLF NFIPMV NIFK LP AI TF+ AA
Sbjct: 58  AAMSNVLGNLSLGRDLQGPLNALVSTTSTFLFKNFIPMVGNIFKALPGAISTFVSAAGKE 117

Query: 304 ITSQ-----------------------------------FQGLMSSLG-ISIDLSPIT 325
           ++SQ                                   F  L+SS+G IS  +  +
Sbjct: 118 LSSQLGNGIGSGFSDFTAKFSSILSPLQGSFQTIVSGLKPVFDSLLSSIGPISTQIMGVF 177

Query: 326 AKFAQIGQNLQ----PVFNGLKTAFSQLPSFFTSIGSAVAPVIDTIISGLARLDFSGFEA 381
           +K  Q+  N+     PV + L  AF QLPS F +I  AV P+IDTI SG++RLDFSG +A
Sbjct: 178 SKLPQLFSNVISAVIPVISTLSVAFGQQLPSLFEAISVAVQPMIDTISSGISRLDFSGIQA 237

Query: 382 LISAILPALQAGFSNFAAIVGPAISGVVDSFVGMWNAAQPLISILSDALMPVFQILGSFL 441
           +ISA++PA+  G +    I+GP+I  +V+SFV MWN+ QPL ++++ ALMP FQ+LG+F+
Sbjct: 238 IISALVPAITTGITTMMGIIGPSIDTLVNSFVKMWNSIQPLATVIAGALMPAFQVLGAFI 297

Query: 442 GGVVKGALMGVSFAFDAVKVAIQLVTPIIDLLVQGLNFVQPVLSVIAEWIGVAIGMFGNL 501
           GGV+KGA++ +S  FD ++V +  +TPII  ++        PVL+ +A+W+G AIG F N
Sbjct: 298 GGVLKGAMLALSATFDTIRVVVGFLTPIIAAVLAKFQEFAPVLATVAQWVGTAIGFFANF 357

Query: 502 GTAGQGLSAFIKSAWTNIQTAISTAGTIISTVIDYIKLAFSGAGSAVGVLKNIFSLAWMA 561
           G AG  L   I SAW  I++ IS+  + I  +I+  K  F+G GSA G L+++ S AW
Sbjct: 358 GAAGTSLKGLITSAWNGIKSIISSVVSGIGGIINTAKAIFTGLGSAGGALRSMISGAWSG 417

Query: 562 MGDAINVAKGIISSVINGIKSAFSSFS-------SLVSSVGSAVNGVIDSISSTIRG--- 611
           +   I+   G IS  INGIKS FSS          S++S V S + G+I   SSTI G
Sbjct: 418 IRSIISSVGGSISGTINGIKSFFSSLGGSGNGLRSVMSGVWSGITGIISGASSTISGIID 477

Query: 612 --------LANIDISGAGAAIMNGFLNGLKSAWGAVKSFVSGIANWIAEHKGPISYDRVL 663
                   L NID++GAG A+++GF+ GLKS W A K FV GIA+WI +HKGPISYDR +
Sbjct: 478 GIKNIFNSLKNIDLAGAGRAVIDGFVGGLKSTWEAGKKFVGGIADWIKDHKGPISYDRKI 537

Query: 664 LKPAGKAIMGGLNTSLIDGFKEVKSNVSGMADDLASTMT 702
           L PAG+AIMGG N SL++ FK V+ NVSG+A  + S +T
Sbjct: 538 LIPAGQAIMGGFNDSLMENFKAVQKNVSGIAKQIQSAIT 576
```

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

```
 Identities = 272/701 (38%), Positives = 371/701 (52%), Gaps = 91/701 (12%)

Query: 1   MATNLGQAYVQIMPSAKGISGSISKTLDPEASSAGSSAGSLLGGKLIGILGSVIAAAKIG 60
           MAT LGQAYVQIMPSA+GISG+ISK LDPEA SAG SAGSL+GG L+ ++G  IAAA IG
Sbjct: 1   MATELGQAYVQIMPSARGISGAISKQLDPEARSAGLSAGSLIGGNLVKMIGGAIAAAGIG 60

Query: 61  EMVTKAISSSISEGAALQQSLGGVETLFKSNANLVKKYADEAYKTTGLSANAYMESVTGF 120
           +M    ISS++S GA LQQS GG++TL+K     VK +A EAYK  G+SAN Y E
Sbjct: 61  KM----ISSALSAGADLQQSFGGIDTLYKGAETAVKGFAKEAYKA-GISANTYAEQAVSM 115

Query: 121 SASLLQSLGGDTAKAAKVANMAMIDMADNSNKMGTSMESIQYAYQGFAKQNYTMLDNLKL 180
            ASL QSLGGD  AAK ANMA++DMADNS KMGT + SIQ AYQGFAKQNYTMLDNL+L
Sbjct: 116 GASLKQSLGGDAVAAAKAANMAIMDMADNSAKMGTDITSIQMAYQGFAKQNYTMLDNLRL 175

Query: 181 GYGGTQEEMKRLLSDAQKL---TGKKYDISNLSDVYEAIHAIQGKIGITGTTAKEAATTF 237
           GYGGT+EEMKRLLSDA+KL    GKK+D+SN +DV EAIH +Q  +GI G  A+EA TTF
```

```
Sbjct: 176 GYGGTKEEMKRLLSDAEKLPAAMGKKFDLSNYADVVEAIHLVQDNMGIAGVAAEEAKTTF 235

Query: 238 TGSFEAMKAASKNLLGKMALGEDIKPSLKALFDTTSNFVLNNFIPMLTNVFKGFGSVISL 297
           +GS  AMK++  N++  ++LG+DI+P+L+ L +TTSNF+  NFIPM+ N+FKG  S I
Sbjct: 236 SGSLAAMKSSFTNVMAGLSLGDDIRPALRGLAETTSNFLFGNFIPMVANIFKGLPSAIGT 295

Query: 298 TFSELIPKIV----GFMQTSGPSLMQSGISFIISFV--------NGFLTAY---PAFLTV 342
                P I     G M + G S+  S I+   + +         NG  TA+    P+F T
Sbjct: 296 FIGAAAPIITSQFQGLMSSLGISIDLSPITAKFAQIGQNLQPVFNGLKTAFSQLPSFFTS 355

Query: 343 AGKIFTDFVSFVMQSIPGL----LQAGATLVLNLIDGILANLPQIATSAVS-VISSFISM 397
            G      +  ++  +  L     +A  + +L  +    +N   I   A+S V+ SF+ M
Sbjct: 356 IGSAVAPVIDTIISGLARLDFSGFEALISAILPALQAGFSNFAAIVGPAISGVVDSFVGM 415

Query: 398 LQANYPAI------LKKGFEILSYLVQGI--------------------IARLPDIVIT 430
             A  P I       L   F+IL   + G+                      +  + D+++
Sbjct: 416 WNAAQPLISILSDALMPVFQILGSFLGGVVKGALMGVSFAFDAVKVAIQLVTPIIDLLVQ 475

Query: 431 ----VGKLIAILAGAIASNLPKVLALGV--QLLITFVKGILSVIGKINETANNIGEKLIN 484
               V  +++++A  I   +     LG   Q L  F+K   + I     TA  I   +I+
Sbjct: 476 GLNFVQPVLSVIAEWIGVAIGMFGNLGTAGQGLSAFIKSAWTNIQTAISTAGTIISTVID 535

Query: 485 AIKSI----------------------DLLSAGRAIMRGFLRGLEDVWGDIQNFVGDIA 521
            IK                        D ++  + I+   + G++  +    + V  +
Sbjct: 536 YIKLAFSGAGSAVGVLKNIFSLAWMAMGDAINVAKGIISSVINGIKSAFSSFSSLVSSVG 595

Query: 522 GWIKDHKGPISYDRRLLI-----PAGNAIMQGLHQGLVDKFKPVKNLVNGMAEEIQSSFG 576
             +      IS   R L       AG AIM G   GL   + VK+ V+G+A  I   G
Sbjct: 596 SAVNGVIDSISSTIRGLANIDISGAGAAIMNGFLNGLKSAWGAVKSFVSGIANWIAEHKG 655

Query: 577 NPQLAFDMDTNVNNGFERIGTLNKNLSSQVTSTDNYTSGNA 617
             +++D       G   +G LN +L         +  SG A
Sbjct: 656 --PISYDRVLLKPAGKAIMGGLNTSLIDGFKEVKSNVSGMA 694
```

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

Example 440

A DNA sequence (GBSx0477) was identified in *S.agalactiae* <SEQ ID 1409> which encodes the amino acid sequence <SEQ ID 1410>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2565(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:AAG18637 GB:AY007505 unknown [Streptococcus mitis]
 Identities = 64/119 (53%), Positives = 87/119 (72%), Gaps = 2/119 (1%)

Query: 1   MLKMDEDALVCDLAETYHIYDYKQLPPLKVAVFSLGLREESRINRVISGNRVSFERRILA 60
           M++ DEDAL+CDLAETY I+DY+QLP  +VAVF+ GLR++SRI    ++ ++V FE  +LA
Sbjct: 1   MIQTDEDALICDLAETYGIFDYRQLPADQVAVFAFGLRDDSRIKLAMTNSKVPFETFLLA 60

Query: 61  GMFDRLGMLIWMKTTDGQKGKNRPEMVSTMF--DNQQKDSEVVSFGSGKDFEETRNNIL 117
           G+ DRL  L+W KTTDGQKG N+P MV+       + K+S+ + F SG+DFEE R  IL
Sbjct: 61  GVLDRLSALVWFKTTDGQKGINKPLMVTEELTGKTKAKESKEMIFDSGEDFEEYRQKIL 119
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1411> which encodes the amino acid sequence <SEQ ID 1412>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2905(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 = 60/123 (48%), Positives = 82/123 (65%), Gaps = 2/123 (1%)

Query: 1   MLKMDEDALVCDLAETYHIYDYKQLPPLKVAVFSLGLREESRINRVISGNRVSFERRILA 60
           M+  D+DAL CDLAETY IYDY+QLP  +VAVF++GLR  SRI   +SG   + +  +LA
Sbjct: 1   MIAKDDDALTCDLAETYGIYDYRQLPAYQVAVFAVGLRSNSRIKMALSGETEALDTVLLA 60

Query: 61  GMFDRLGMLIWMKTTDGQKGKNRPEMV--STMFDNQQKDSEVVSFGSGKDFEETRNNILG 118
           G++D   +L W KT DGQ G+N+P+ V  +       QK ++V+SF SG+DFE  R  +LG
Sbjct: 61  GIYDNTNLLFWSKTKDGQSGQNKPKSVVEAISGSKSQKANDVISFVSGEDFENARKQLLG 120

Query: 119 FGG 121
             G
Sbjct: 121 GDG 123
```

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

Example 441

A DNA sequence (GBSx0478) was identified in *S.agalactiae* <SEQ ID 1413> which encodes the amino acid sequence <SEQ ID 1414>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2280(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:AAG18636 GB:AY007505 unknown [Streptococcus mitis]
 Identities = 40/80 (50%), Positives = 62/80 (77%), Gaps = 1/80 (1%)

Query: 3   TSSGFEYKIEESRLKNYELVEALADLESNPLSLPKVLRLLLGDQVESLKNHLRASDGTVS 62
           TS+GF ++I + RL+NYEL+EA++++++NP  LPKV++L+LG++ E LKNH+R +DG V
Sbjct: 24  TSTGFPPEITKERLENYELLEAISEVDTNPAVLPKVVKLMLGNKSEDLKNHVRTADGIVP 83

Query: 63  TEALMEEVKEIFES-GQLKK 81
            + +  E+ EIF S  QLKK
Sbjct: 84  LDKMGAEISEIFSSQNQLKK 103
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1415> which encodes the amino acid sequence <SEQ ID 1416>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4365(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 = 42/75 (56%), Positives = 60/75 (80%)

Query: 2  KTSSGFEYKIEESRLKNYELVEALADLESNPLSLPKVLRLLLGDQVESLKNHLRASDGTV 61
          KT+SGFEY+I + RLKN+ELVEA+A+ E++P ++ K++ LLLGD  +SLK H+R ++G V
Sbjct: 7  KTTSGFEYEIPKKRLKNFELVEAIAEEETDPTAVVKIVNLLLGDAAKSLKEHVRDAEGIV 66

Query: 62 STEALMEEVKEIFES 76
            EA+  E+KEIFES
Sbjct: 67 DVEAIGVEIKEIFES 81
```

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

Example 442

A DNA sequence (GBSx0479) was identified in *S.agalactiae* <SEQ ID 1417> which encodes the amino acid sequence <SEQ ID 1418>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3461(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:AAG18635 GB:AY007505 unknown [Streptococcus mitis]
 Identities = 114/183 (62%), Positives = 142/183 (77%)

Query: 2   VANSSNVTTAKPKIGGAIYTAPLGTELPKDTASELNEAFKSLGYISEDGLSNEDKRESEE 61
           +A  +NVTTAKPKIGGA+Y+APLGT LP D  ++L++AF++LGYIS+DG++N +  ESE
Sbjct: 1   MATEANVTTAKPKIGGAVYSAPLGTALPTDATTKLDQAFEALGYISDDGMTNSNSPESEN 60

Query: 62  IQAWGGDVVESAQKSKADKFTYTLIEALNIEVLKEIYGKDNVTGDLKTGITVKSNSKPLE 121
           I+AWGG VV S QK K D F Y LIEALN+ VLKE+YG DNV+GDL +GIT+K+NSK L
Sbjct: 61  IKAWGGVVVSSVQKEKTDTFKYMLIEALNLHVLKEVYGPDNVSGDLSSGITIKANSKELP 120

Query: 122 EHCLVIEMILKNNTVKRIVIPKGKVSEVGEIKYVDNEAAGYETTLQAFPDAEGNTHYEYI 181
            HCLVIE +LK   +KRIVIP GKV+ + EI Y D    GY TT+ AFP+A  +THYEYI
Sbjct: 121 HHCLVIETVLKGGVLKRIVIPSGKVTAIDEITYNDGSVLGYGTTVTAFPNAADDTHYEYI 180

Query: 182 KGA 184
           KGA
Sbjct: 181 KGA 183
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1419> which encodes the amino acid sequence <SEQ ID 1420>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2379(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 = 119/182 (65%), Positives = 142/182 (77%)
```

```
Query: 4   NSSNVTTAKPKIGGAIYTAPLGTELPKDTASELNEAFKSLGYISEDGLSNEDKRESEEIQ 63
           ++ NVT+AKPK GGAIY+APLGTELPKD  SELN  FK+LGY+SEDG+ NED R SE I+
Sbjct: 6   DTKNVTSAKPKTGGAIYSAPLGTELPKDAKSELNTKFKNLGYVSEDGVVNEDTRSSENIK 65

Query: 64  AWGGDVVESAQKSKADKFTYTLIEALNIEVLKEIYGKDNVTGDLKTGITVKSNSKPLEEH 123
           AWGGD+V + Q  K DKFTY LIE+LN+EVLKE+YG  NVTGDL  GI +KSNSK LE H
Sbjct: 66  AWGGDIVGAVQTEKEDKFTYKLIESLNVEVLKEVYGAVNVTGDLSGGIQIKSNSKELEAH 125

Query: 124 CLVIEMILKNNTVKRIVIPKGKVSEVGEIKYVDNEAAGYETTLQAFPDAEGNTHYEYIKG 183
            +V++MI+     +KRIV+P  KV EVGEIKYVD E  GYETTL+ FPD +G+TH EYI
Sbjct: 126 VIVVDMIMNGGILKRIVLPNAKVDEVGEIKYVDGEVVGYETTLKCFPDKDGDTHREYIVK 185

Query: 184 AG 185
           G
Sbjct: 186 PG 187
```

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

Example 443

A DNA sequence (GBSx0480) was identified in *S.agalactiae* <SEQ ID 1421> which encodes the amino acid sequence <SEQ ID 1422>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2214(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAB18710 GB:U38906 ORF35 [Bacteriophage r1t]
 Identities = 52/78 (66%), Positives = 66/78 (83%)

Query: 1  MSKFKFKLNKAGVAELMKSSEMQQVLTTKATAIRERCGDGYAQDIHVGKNRANAMVSAKT 60
          M+K  FKLN++GVA +MKS EMQ +L  KA+A+++RCG GY QD+HVGKNRANAMV A+T
Sbjct: 1  MAKNLFKLNRSGVASMMKSPEMQAILKEKASAVKQRCGPGYGQDMHVGKNRANAMVFAET 60

Query: 61 IKAKKDNSKNNTLLKAVR 78
           +AK+DN KNNT+LKAVR
Sbjct: 61 YQAKRDNMKNNTILKAVR 78
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1423> which encodes the amino acid sequence <SEQ ID 1424>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2446(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 75/78 (96%), Positives = 76/78 (97%)

Query: 1  MSKFKFKLNKAGVAELMKSSEMQQVLTTKATAIRERCGDGYAQDIHVGKNRANAMVSAKT 60
          MSKFKFKLN+AGVAELMKSSEMQQVLTTKATAIRERCGDGY QDIHVGKNRANAMVS KT
Sbjct: 1  MSKFKFKLNRAGVAELMKSSEMQQVLTTKATAIRERCGDGYVQDIHVGKNRANAMVSTKT 60
```

```
Query: 61 IKAKKDNSKNNTLLKAVR 78
          IKAKKDNSKNNTLLKAVR
Sbjct: 61 IKAKKDNSKNNTLLKAVR 78
```

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

Example 444

A DNA sequence (GBSx0481) was identified in *S.agalactiae* <SEQ ID 1425> which encodes the amino acid sequence <SEQ ID 1426>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2888(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:AAB18709 GB:U38906 ORF34 [Bacteriophage r1t]
 Identities = 41/59 (69%), Positives = 45/59 (75%)

Query: 1   MTGKKVEYILAIPKGDKHDWEDKEVCFFDKKWRTVGLALEGIEELIPLEWNKKVMVERY 59
           +TGKK  Y LAIPK D HDWE+K+V FF K WRT G  LEGIE LIPL+WNKKV VE Y
Sbjct: 56  LTGKKAIYTLAIPKKDTHDWENKKVRFFGKTWRTFGEPLEGIEGLIPLDWNKKVTVEHY 114
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1427> which encodes the amino acid sequence <SEQ ID 1428>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2779(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 = 51/60 (85%), Positives = 57/60 (95%)

Query: 1   MTGKKVEYILAIPKGDKHDWEDKEVCFFDKKWRTVGLALEGIEELIPLEWNKKVMVERYE 60
           +TGKKVEY+LAIPKGD+HDWE+KEV FF KKWRTVG+ LEGIEELIPL+WNKKVMVERYE
Sbjct: 50  LTGKKVEYVLAIPKGDEHDWENKEVRFFGKKWRTVGIPLEGIEELIPLDWNKKVMVERYE 109
```

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

Example 445

A DNA sequence (GBSx0482) was identified in *S.agalactiae* <SEQ ID 1429> which encodes the amino acid sequence <SEQ ID 1430>. 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 = 89/130 (68%), Positives = 106/130 (81%), Gaps = 1/130 (0%)

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  VLKSVTVDIVARTLMTATQGEPMSQESQSALGYTWSGTYLVPGGGLFIKDSELKRLGLKK 119
           V+KSVTVDIVARTLMT+T  EPM+Q ++SALGY+ SG+YLVPGGGLFIK+SEL RLGLKK
Sbjct: 61  VVKSVTVDIVARTLMTSTDQEPMTQTTESALGYSVSGSYLVPGGGLFIKNSELSRLGLKK 120

Query: 120 QRYGGIELYG 129
           QR+G I+ YG
Sbjct: 121 QRFGVIDFYG 130
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1431> which encodes the amino acid sequence <SEQ ID 1432>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2061(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 = 116/138 (84%), Positives = 129/138 (93%)

Query: 3   NFATTDDVILLWRQLSVDEIKRAEALLETVSDTLRLEASKVGKNLDEMILETPYFATVLK 62
           NFATTDDVILLWR LSVDE+KRA ALL+ VSDTLR+EA KVGK+LD+ +++ PYF  V+K
Sbjct: 3   NFATTDDVILLWRPLSVDELKRANALLKVVSDTLRMEADKVGKDLDKTMVDKPYFVNVIK 62

Query: 63  SVTVDIVARTLMTATQGEPMSQESQSALGYTWSGTYLVPGGGLFIKDSELKRLGLKKQRY 122
           SVTVDIVARTLMT+T+GEPM+QESQSALGYTWSGTYLVPGGGLFIKDSELKRLGLKKQRY
Sbjct: 63  SVTVDIVARTLMTSTRGEPMAQESQSALGYTWSGTYLVPGGGLFIKDSELKRLGLKKQRY 122

Query: 123 GGIELYGEIERNNSYFSR 140
           GGIELYGEIER+NS FSR
Sbjct: 123 GGIELYGEIERDNSCFSR 140
```

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

Example 446

A DNA sequence (GBSx0483) was identified in *S.agalactiae* <SEQ ID 1433> which encodes the amino acid sequence <SEQ ID 1434>. 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 = 132/296 (44%), Positives = 189/296 (63%), Gaps = 8/296 (2%)

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 LEPRTMTDASFKATNSFDGVVTGNVIKYEADK--IDDN--IDAAVTTIVANGNDVTGIAL 180
            + PR  T ++   TN FD  VT    K EA +   D N  I+ AV  +     DVTGIA+
Sbjct: 123 VNPRLGTASAVIGTNHFDSKVTQ---KVEAPRGIADPNGAIENAVELLTGVDADVTGIAI 179

Query: 181 SPQAGQDMSKRKDKFDNVMYPEFRFGQRPSNFFNMTLDINKTLTMKGGTAKDDHAIVGDF 240
            +P      ++K+KD  DN ++PE ++G  P      + +D+NKT++     T + D AI+GDF
Sbjct: 180 NPSFRSALAKQKDLQDNALFPELKWGATPDTINGLPVDVNKTVSDMSLTQR-DRAIIGDF 238

Query: 241 QNMFKWGYAENIPMEIIEYGDPDGSGRDLKAYNEILLRTEAFIGWGILDEKAFSRV 296
             N FKWGYA+ +P+E+I+YGDPD SG DLK YN++ +R E F+GWGILD   F+RV
Sbjct: 239 ANGFKWGYAKEVPLEVIQYGDPDNSGLDLKGYNQVYIRAELFLGWGILDATKFARV 294
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1435> which encodes the amino acid sequence <SEQ ID 1436>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2772(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 = 133/298 (44%), Positives = 187/298 (62%), Gaps = 2/298 (0%)

Query: 1   MAESIKAGTLFKPELVTEIMSKVKGHSTLAKLSGQTPIPFNGVEQFVFNLDGNAQIVGEG 60
            M        +LF   LV+++++KVKGHS+LAKLS Q PIPFNG ++F F LD +  +V E
Sbjct: 1   MGTETSKASLFDKHLVSDLINKVKGHSSLAKLSSQKPIPFNGSKEFTFTLDSDIDVVAEN 60

Query: 61  EQKLGNTAKVTSKIIKPLKFVYQARMTDEFKYASEEKRLNFLKHYADGFAKKMAEAFDIA 120
             +K       +    I P+K  Y AR++DEF YA+EE+++ LK + +GFAKK+A   D+
Sbjct: 61  GKKTHGGLSLEPVTIVPIKVEYGARLSDEFLYATEEEKIDILKAFNEGFAKKLARGIDLM 120

Query: 121 AIHGLEPRTMTDASFKATNSFDGVVTGNVIKYEADKIDDNIDAAVTTIVANGNDVTGIAL 180
            A+HG+ PRT   +    TN FD  VT  V   E++  D NI+AAV  I  +   VTG+A+
Sbjct: 121 AMHGINPRTKKASDVIGTNHFDSKVTQVVKFTESEDADANIEAAVNLIQGSEGVVTGLAM 180

Query: 181 SPQAGQDMSK-RKDKFDNVMYPEFRFGQRPSNFFNMTLDINKTLTMKGGTAKD-DHAIVG 238
             +   ++K   +    MYPE  +G  P +   +N T+      A+  D  I+G
Sbjct: 181 DTEFSTALAKVTNGEMGPKMYPELAWGANPDSINGLKSSVNTTVGAGADEAESKDLVIIG 240

Query: 239 DFQNMFKWGYAENIPMEIIEYGDPDGSGRDLKAYNEILLRTEAFIGWGILDEKAFSRV 296
            DF++MFKWGYA+ IPMEII+YGDPD SG+DLK YN+I LR EA+IGWGILD K+F+RV
Sbjct: 241 DFESMFKWGYAKQIPMEIIKYGDPDNSGKDLKGYNQIYLRAEAYIGWGILDAKSFARV 298
```

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

Example 447

A DNA sequence (GBSx0484) was identified in *S.agalactiae* <SEQ ID 1437> which encodes the amino acid sequence <SEQ ID 1438>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2224(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 9659> which encodes amino acid sequence <SEQ ID 9660> was also identified.

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

```
>GP:AAB18705 GB:U38906 ORF30 [Bacteriophage r1t]
 Identities = 64/158 (40%), Positives = 101/158 (63%), Gaps = 8/158 (5%)

Query: 43  MSEFKVIETQEELDTIVKARIARERE----KYQDYDQLKTRVEELETENSSLQTALNDAK 98
           MSE  + +TQEEL+ I++ R+AR++E     + DYD+LKT++  LE +N++ Q  + ++K
Sbjct: 1   MSENNLPKTQEELNQIIETRLARQKETIEANFADYDELKTKIAALEADNTAYQATIEESK 60

Query: 99  SNTDSYTEKITTLENQIAGYEAANLRTKVALQYGLPIDLANRLQGDDEDGLKVDAERLAS 158
           S    + ++    E QI+GY+   L+  +A++ GLP+DLA+RL GDDE+ LK DAER +
Sbjct: 61  S----WEQEKADYEKQISGYKTTQLKQSIAIKAGLPLDLADRLSGDDEESLKADAERFSG 116

Query: 159 FIKPSQPQPPTKSNEPIITDQKEAGWIEMARNLVNKGE 196
           FIKP  P  P K  EP + D K+  + ++   L  +GE
Sbjct: 117 FIKPKTPPAPLKDVEPNLGDGKDGAYRKLVDGLKTEGE 154
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1439> which encodes the amino acid sequence <SEQ ID 1440>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3476(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 = 128/149 (85%), Positives = 136/149 (90%)

Query: 43  MSEFKVIETQEELDTIVKARIAREREKYQDYDQLKTRVEELETENSSLQTALNDAKSNTD 102
           MSEFKVIETQEELDTIVKARIARE+EKYQDYDQLKTRVEELETENSSLQTALNDAKSNTD
Sbjct: 1   MSEFKVIETQEELDTIVKARIAREKEKYQDYDQLKTRVEELETENSSLQTALNDAKSNTD 60

Query: 103 SYTEKITTLENQIAGYEAANLRTKVALQYGLPIDLANRLQGDDEDGLKVDAERLASFIKP 162
           SYTE+I+TL+NQIA YE ANLRTKVALQYGLPIDLA+RLQGDDEDGLKVDAERLASFIKP
Sbjct: 61  SYTEEISTLKNQIADYETANLRTKVALQYGLPIDLADRLQGDDEDGLKVDAERLASFIKP 120

Query: 163 SQPQPPTKSNEPIITDQKEAGWIEMARNL 191
           SQPQPP KSNEP I    +A +  + + L
Sbjct: 121 SQPQPPAKSNEPNIDSNADANYRALVQGL 149
```

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

Example 448

A DNA sequence (GBSx0485) was identified in *S.agalactiae* <SEQ ID 1441> which encodes the amino acid sequence <SEQ ID 1442>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
```

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2888(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:AAB18704 GB:U38906 ORF29 [Bacteriophage r1t]
 Identities = 322/461 (69%), Positives = 383/461 (82%)

Query: 8   KLGNQRPTQSVNLHFAKTLAHEAINYYKKTGLSCYLWQENMLIPMMAINEDNLWVHQKYG 67
           + GNQ PTQSV L F +T   EAI  Y+K+   CY WQ+N+L  +MAI+ED LW HQK+G
Sbjct: 6   RFGNQYPTQSVILPFTETKYQEAIEIYEKSKHECYPWQKNLLKEVMAIDEDGLWTHQKFG 65

Query: 68  YAIPRRNGKTEVVYILELWALHKGLKILHTAHRISTSHSSFEKVKKYLEMSGYVDGEDFI 127
           Y+IPRRNGKTE+VYILELW+L +GL ILHTAHRISTSHSS+EK+KKYLE SGYV+GEDF
Sbjct: 66  YSIPRRNGKTEIVYILELWSLVQGLSILHTAHRISTSHSSYEKLKKYLEDSGYVEGEDFK 125

Query: 128 SNKAKGQERIEFKSSGSVIQFRTRTSNGGLGEGFDLLIIDEAQEYTAEQESALKYTVTDS 187
           S KAKGQER+E   SG VIQFRTRTS+GGLGEGFD+L+IDEAQEYT EQESALKYTVTDS
Sbjct: 126 SIKAKGQERLELIESGGVIQFRTRTSSGGLGEGFDILVIDEAQEYTTEQESALKYTVTDS 185

Query: 188 DNPMTIMCGTPPTMVSTGTVFESYRKECLKGDRRYSGWAEWSVDEMQPIHDVKSWYVANP 247
           DNPMTIMCGTPPT VS+GTVF +YR   + G  +YSGWAEWSV++++ IHDV++WY +NP
Sbjct: 186 DNPMTIMCGTPPTPVSSGTVFTNYRDNTIAGKAKYSGWAEWSVEDVKDIHDVEAWYNSNP 245

Query: 248 SMGYHLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVPELKSKLF 307
           SMGYHLNERKIEAELGED++DHN+QRLGYWP +NQKSVISE+EW  LKV ++P +K KLF
Sbjct: 246 SMGYHLNERKIEAELGEDKLDHNVQRLGYWPKYNQKSVISEQEWNALKVNRLPVIKGKLF 305

Query: 308 VGIKFGQDGNNVSLSIAARASENKVFVEAIDCLSVRNGTQWIINFLKSADIAKVVVDGAS 367
           VGIK+G DG NV++SIA +     KVFVE IDC S+RNG QWIINFLK AD+ KVV+DG S
Sbjct: 306 VGIKYGNDGANVAMSIAVKTLSGKVFVETIDCQSIRNGNQWIINFLKKADVEKVVIDGQS 365

Query: 368 GQELLAQEMREHGLKKPELPKVAEIITANTMWEQGIMQETICHNDQPSLTAVVTNCEKRQ 427
           GQ +L  EM++  LK+P LP V EII AN++WEQGI Q+  CH+ QPSL+ VVTNC+KR
Sbjct: 366 GQSILTSEMKDFKLKEPILPTVKEIINANSLWEQGIFQKNFCHSGQPSLSTVVTNCDKRN 425

Query: 428 IGSNGGFGYKSLYDDRDISLMDSALLAHWICYTTKPKRKQR 468
           IG++GGFGYKS +DD DISLMDSALLAHW C   KPK+KQ+
Sbjct: 426 IGTSGGFGYKSQFDDMDISLMDSALLAHWACSNNKPKKKQQ 466
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1443> which encodes the amino acid sequence <SEQ ID 1444>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3133(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 = 437/471 (92%), Positives = 459/471 (96%)

Query: 1   MVTKTKAKLGNQRPTQSVNLHFAKTLAHEAINYYKKTGLSCYLWQENMLIPMMAINEDNL 60
           MVTKTK KLGNQRPTQSVNLHFAK+LAHEAINYYKKTGLSCY WQ NMLIP+MAI+E+ L
Sbjct: 6   MVTKTKTKLGNQRPTQSVNLHFAKSLAHEAINYYKKTGLSCYPWQVNMLIPIMAIDENGL 65

Query: 61  WVHQKYGYAIPRRNGKTEVVYILELWALHKGLKILHTAHRISTSHSSFEKVKKYLEMSGY 120
           WVHQKYGYAIPRRNGKTEVVYI++LWALHKGLKILHTAHRISTSH+SFEKVKKYLEMSGY
Sbjct: 66  WVHQKYGYAIPRRNGKTEVVYIVQLWALHKGLKILHTAHRISTSHASFEKVKKYLEMSGY 125
```

```
Query: 121 VDGEDFISNKAKGQERIEFKSSGSVIQFRTRTSNGGLGEGFDLLIIDEAQEYTAEQESAL 180
           VDGEDFISNKAKGQERIEFK+SG+VIQFRTRTSNGGLGEGFDLLIIDEAQEYT+EQESAL
Sbjct: 126 VDGEDFISNKAKGQERIEFKASGAVIQFRTRTSNGGLGEGFDLLIIDEAQEYTSEQESAL 185

Query: 181 KYTVTDSDNPMTIMCGTPPTMVSTGTVFESYRKECLKGDRRYSGWAEWSVDEMQPIHDVK 240
           KYTVTDSDNPMTIMCGTPPTMVSTGTVFE+YRK+CLKG++RYSGWAEWSV EM  I+DV
Sbjct: 186 KYTVTDSDNPMTIMCGTPPTMVSTGTVFEAYRKDCLKGNKRYSGWAEWSVPEMVKINDVS 245

Query: 241 SWYVANPSMGYHLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVP 300
           SWY++NPSMG+HLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVP
Sbjct: 246 SWYISNPSMGFHLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVP 305

Query: 301 ELKSKLFVGIKFGQDGNNVSLSIAARASENKVFVEAIDCLSVRNGTQWIINFLKSADIAK 360
           ELKSKLFVGIKFGQDGNNVSLSIAAR SENKVFVE IDCLSVRNGTQWIINFLKSADIAK
Sbjct: 306 ELKSKLFVGIKFGQDGNNVSLSIAARTSENKVFVETIDCLSVRNGTQWIINFLKSADIAK 365

Query: 361 VVVDGASGQELLAQEMREHGLKKPELPKVAEIITANTMWEQGIMQETICHNDQPSLTAVV 420
           VV+DGASGQELLAQEM++ GLKKPELPKVAEIITAN MWEQGIMQETICH+DQPSLTAVV
Sbjct: 366 VVIDGASGQELLAQEMKDQGLKKPELPKVAEIITANMMWEQGIMQETICHSDQPSLTAVV 425

Query: 421 TNCEKRQIGSNGGFGYKSLYDDRDISLMDSALLAHWICYTTKPKRKQRTSC 471
           TNCEKRQIGSNGGFGYKSLYDDRDISLMDSALLAHWICYTTKPKRKQRTSC
Sbjct: 426 TNCEKRQIGSNGGFGYKSLYDDRDISLMDSALLAHWICYTTKPKRKQRTSC 476
```

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

Example 449

A DNA sequence (GBSx0486) was identified in *S.agalactiae* <SEQ ID 1445> which encodes the amino acid sequence <SEQ ID 1446>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2745(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 450

A DNA sequence (GBSx0487) was identified in *S.agalactiae* <SEQ ID 1447> which encodes the amino acid sequence <SEQ ID 1448>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2568(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAB18703 GB:U38906 ORF28 [Bacteriophage r1t]
```

```
Identities = 124/250 (49%), Positives = 164/250 (65%), Gaps = 3/250 (1%)

Query: 2   VDDVLPKLLKSVQQDFEKHFGKSEVVAKAFAELQAKKATYKTVNEFAVEVGRLLSLALAN 61
           ++D+LP LL+ + QDF++    S+ + ++   L+ KKATY   NEF VEVG++LS  L
Sbjct: 1   MEDILPPLLEKINQDFDERAANSKKLKQSMELLKTKKATYIQANEFGVEVGQILSDVLGT 60

Query: 62  SVISDELPDGKMYYNIANRLVNDTLRHNYKLISDYAGDVQQNLNKQAKISLKIQRPPLNQ 121
            V  D LPDGKMY+NIA+RL+N  L+ N+ LIS Y+ DVQ  LN+ A   LK Q P LNQ
Sbjct: 61  HVTVDVLPDGKMYFNIADRLLNSILKKNFDLISGYSTDVQSELNQLAGFKLKSQVPELNQ 120

Query: 122 DKIDGLVNRLASEPVFDDVKWLLDEPIVNFSQSIVDDCIRANADFHFKTGLKPTIERIST 181
           D+IDG+VNR++SE  F+ + WLL EPIV FSQS+VDD ++ N DF  K GLKP I R
Sbjct: 121 DRIDGIVNRISSEDDFEKILWLLKEPIVTFSQSVVDDTLKKNIDFQAKAGLKPKIVRKLV 180

Query: 182 GKCCDWCDRLAGRYVYHEEPKDFYKRHQHCQCVIDYHPK--NGKRQNSWSKKWTKETTDI 239
           GK CDWC  LAG Y Y   P D Y RH+ C+C ++Y P+  + KRQ+ WSK W     D
Sbjct: 181 GKACDWCRNLAGSYDYPNVPSDVYHRHERCRCTVEYDPRDIDKKRQDVWSKNWVDPDKDA 240

Query: 240 -LERRKQMNI 248
            +  RK +N+
Sbjct: 241 KIAERKNLNL 250
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1449> which encodes the amino acid sequence <SEQ ID 1450>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3099(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
Identities = 169/261 (64%), Positives = 207/261 (78%), Gaps = 2/261 (0%)

Query: 1   MVDDVLPKLLKSVQQDFEKHFGKSEVVAKAFAELQAKKATYKTVNEFAVEVGRLLSLALA 60
           MVDDVLPKLLKSV+QDFEK+FG+S+VV KAFAELQAKK TYKTVNEFA+EVGRLLSLAL
Sbjct: 1   MVDDVLPKLLKSVRQDFEKYFGESDVVTKAFAELQAKKVTYKTVNEFAIEVGRLLSLALT 60

Query: 61  NSVISDELPDGKMYYNIANRLVNDTLRHNYKLISDYAGDVQQNLNKQAKISLKIQRPPLN 120
            SV SD+LPDGKMYYNIA RL+++T+  NYKLIS YAGDVQ+ LN+ A+I LK+QRPPLN
Sbjct: 61  GSVSSDKLPDGKMYYNIAKRLLDETMGRNYKLISGYAGDVQRILNENAQIGLKVQRPPLN 120

Query: 121 QDKIDGLVNRLASEPVFDDVKWLLDEPIVNFSQSIVDDCIRANADFHFKTGLKPTIERIS 180
           +DKI+G+VNRL SE  FDDVKWL  EPIVNFSQSIVDD I+ANAD  +KTG+ P + R
Sbjct: 121 RDKINGMVNRLDSENTFDDVKWLFGEPIVNFSQSIVDDTIKANADLQYKTGMTPQVVRTE 180

Query: 181 TGKCCDWCDRLAGRYVYHEEPKDFYKRHQHCQCVIDYHPKNGKRQNSWSKKWTK--ETTD 238
           +G CC+WC  + G Y Y + PKD ++RHQ C+C +DY PKNGK Q++WSK W K  +T +
Sbjct: 181 SGNCCEWCREVVGTYSYPKVPKDVWRRHQRCRCTLDYDPKNGKVQSAWSKIWRKKEKTQE 240

Query: 239 ILERRKQMNIDIRDNNRKSDI 259
            +ER ++         + K+DI
Sbjct: 241 SIERVEKFKESALVESIKNDI 261
```

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

Example 451

A DNA sequence (GBSx0488) was identified in *S.agalactiae* <SEQ ID 1451> which encodes the amino acid sequence <SEQ ID 1452>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38   Transmembrane   93 - 109 (  93 - 110)

----- 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:AAC39307 GB:AF022773 ORF5 [Lactococcus bacteriophage phi31]
 Identities = 271/410 (66%), Positives = 326/410 (79%), Gaps = 2/410 (0%)

Query: 1   MNYMGMGYLQRKLALFKTGVDKRYRYYAMDDRDNTRSIVMPDNVREMYRSVIEWTAKGVD 60
            M   G+GYL+ KL++ K   + RY  YAM   D  + I +P  + + YRS++ W AKGVD
Sbjct: 1   MTEKGIGYLRFKLSVHKRRAEMRYEQYAMKHVDRFKGITIPQALSQQYRSILGWCAKGVD 60

Query: 61  SLADRIIFREFANDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPGKEDSLPKMQ 120
            SLADR+IFREF NDDF   EIF+ NNPDIFFD+A+ SALIASC F+YI  G+ D++ ++Q
Sbjct: 61  SLADRLIFREFENDDFTVNEIFEENNPDIFFDSAVLSALIASCSFIYISKGENDAV-RLQ 119

Query: 121 VIEASKATGILDPTTFLLTEGYAVLESDSNENPTLEAYFTGEKTWYYPKDEKP-YSIDNS 179
            VIEA  ATGI+DP T LLTEGYAVLE D N N  LEA+F  ++T YY +D +   SI N
Sbjct: 120 VIEAVNATGIIDPITGLLTEGYAVLERDENNNVVLEAHFLPDRTDYYYRDSRNNISIANP 179

Query: 180 TGHPLLVPVIHRPDAVRPFGRSRITKAGMYHQKAAKRTLERAEVTAEFYSFPQKYVLGMD 239
            TGHPLLVP+IHRPDAVRPFGRSRIT++GMY Q  AKRTLERA+VTAEFYSFPQKYV G+
Sbjct: 180 TGHPLLVPIIHRPDAVRPFGRSRITRSGMYWQSNAKRTLERADVTAEFYSFPQKYVTGLS 239

Query: 240 PDAEPMEKWRATVSTLLEISKDEDGDKPTVGQFTTASMAPFMDHLKMYASLFAGGSGLTL 299
             DAEPME W+ATVS++L+ +KDEDGDKPT+GQFT  SM+PF + L+  A+ FAG +GLTL
Sbjct: 240 DDAEPMETWKATVSSMLQFTKDEDGDKPTLGQFTQPSMSPFTEQLRTAAAGFAGETGLTL 299

Query: 300 DDLGFPSDNPSSVEAIKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRDDFPYLRNQFM 359
            DDLGF SDNPSSVEAIKA+HENLR AGRKAQRS  +G LNVAY+A CLRDD PYLR QF
Sbjct: 300 DDLGFVSDNPSSVEAIKASHENLRLAGRKAQRSLGAGLLNVAYLAACLRDDVPYLREQFS 359

Query: 360 DTEIKWEPLFEADANMLTLVGDGAIKLNQAIPGFMDADVIRDLTGVKGSD 409
             T+ KWEPLFEADA+ML+L+GDGAIKLNQAIP F++ D IRDLTG+KG++
Sbjct: 360 KTKPKWEPLFEADASMLSLIGDGAIKLNQAIPEFINKDTIRDLTGIKGAE 409
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1453> which encodes the amino acid sequence <SEQ ID 1454>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38   Transmembrane   93 - 109 (  93 - 110)

----- 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>
```

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

```
 Identities = 395/422 (93%), Positives = 407/422 (95%)

Query: 1   MNYMGMGYLQRKLALFKTGVDKRYRYYAMDDRDNTRSIVMPDNVREMYRSVIEWTAKGVD 60
            MNYMGMGYL+RKLALFKTGVDKRYRYYAMDDRD+TRSIVMP+NVREMYRSV+EWTAKGVD
```

```
Sbjct: 1   MNYMGMGYLRRKLALFKTGVDKRYRYYAMDDRDDTRSIVMPNNVREMYRSVLEWTAKGVD 60

Query: 61  SLADRIIFREFANDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPGKEDSLPKMQ 120
           SLADRIIFREF NDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPG ED LPKMQ
Sbjct: 61  SLADRIIFREFTNDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPGAEDGLPKMQ 120

Query: 121 VIEASKATGILDPTTFLLTEGYAVLESDSNENPTLEAYFTGEKTWYYPKDEKPYSIDNST 180
           VIEASKATGILDPTTFLLTEGYA+LESDSN NPTLEAYFT +  WYYPK  KPY+I N T
Sbjct: 121 VIEASKATGILDPTTFLLTEGYAILESDSNGNPTLEAYFTDKDIWYYPKKGKPYNIKNPT 180

Query: 181 GHPLLVPVIHRPDAVRPFGRSRITKAGMYHQKAAKRTLERAEVTAEFYSFPQKYVLGMDP 240
           GHPLLVP+IHRPDAVRPFGRSRITKAGMYHQKAAKRTLERAEVTAEFYSFPQKYVLGMDP
Sbjct: 181 GHPLLVPIIHRPDAVRPFGRSRITKAGMYHQKAAKRTLERAEVTAEFYSFPQKYVLGMDP 240

Query: 241 DAEPMEKWRATVSTLLEISKDEDGDKPTVGQFTTASMAPFMDHLKMYASLFAGGSGLTLD 300
           DAEPMEKWRATVSTLLEISKDEDGDKPTVGQFTTASMAPFM+HLKMYASLFAGGSGLTLD
Sbjct: 241 DAEPMEKWRATVSTLLEISKDEDGDKPTVGQFTTASMAPFMEHLKMYASLFAGGSGLTLD 300

Query: 301 DLGFPSDNPSSVEAIKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRDDFPYLRNQFMD 360
           DLGFPSDNPSSVE+IKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRD+FPYLRNQFMD
Sbjct: 301 DLGFPSDNPSSVESIKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRDEFPYLRNQFMD 360

Query: 361 TEIKWEPLFEADANMLTLVGDGAIKLNQAIPGFMDADVIRDLTGVKGSDNPIPKATEVTT 420
           T IKWEPLFEADANMLTLVGDGAIKLNQAIPGFMDADVIRDLTGVKG+D PIP  TEVTT
Sbjct: 361 TVIKWEPLFEADANMLTLVGDGAIKLNQAIPGFMDADVIRDLTGVKGADKPIPAITEVTT 420

Query: 421 DG 422
           DG
Sbjct: 421 DG 422
```

SEQ ID 1452 (GBS364) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 6; MW 50kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 11; MW 75kDa).

GBS364-GST was purified as shown in Figure 216, lane 10.

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

Example 452

A DNA sequence (GBSx0489) was identified in *S.agalactiae* <SEQ ID 1455> which encodes the amino acid sequence <SEQ ID 1456>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4063(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 1457> which encodes the amino acid sequence <SEQ ID 1458>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4120(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 = 101/118 (85%), Positives = 110/118 (92%)

Query: 1   MKKKCLICKKTFQAKTNRSLYCSEECRKKGIREKQRKLMKQKRADKKKEKIKVLNTNADV 60
           +KKKCLICKK FQAKTNR+LYCSEECRKKG REKQRKLMKQKRA+++KEK KVLN N DV
Sbjct: 1   LKKKCLICKKNFQAKTNRTLYCSEECRKKGNREKQRKLMKQKRAEQRKEKKKVLNPNTDV 60

Query: 61  TEKPKKIRNLVQHYKKLKREILDNESEFGFTGIALVEGIDIHEENFVDLVMQKIKEQQ 118
           TEKPKKIRNL QHYKKLK+EIL NESEFGFTGI L+EGID+HEENFVDLVMQKIKEQ+
Sbjct: 61  TEKPKKIRNLAQHYKKLKKEILANESEFGFTGITLIEGIDVHEENFVDLVMQKIKEQK 118
```

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

Example 453

A DNA sequence (GBSx0490) was identified in *S.agalactiae* <SEQ ID 1459> which encodes the amino acid sequence <SEQ ID 1460>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0633(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:AAC39305 GB:AF022773 ORF3 [Lactococcus bacteriophage phi31]
 Identities = 75/109 (68%), Positives = 87/109 (79%), Gaps = 1/109 (0%)

Query: 29  LRADKKGTHRVAFEKNKRRLLKTAHLCGICGRPVDKSLKYPHPLSAAIDHIVPIAKGGHP 88
           LRAD+ G HRVAF+KN++ LLKT + CGICG+P+DK LK P PLS  +DHI+PI KGGHP
Sbjct: 3   LRADRTGAHRVAFDKNRKILLKTQNTCGICGKPIDKRLKAPDPLSPVVDHIIPINKGGHP 62

Query: 89  SSIDNLQLTHWQCNRQKSDKLFINQTAVRATVVGNRNLPQSRDWSSYAS 137
           S++DNLQL HW CNRQKSDKLF N        V+GNRNLPQSRDWSSY S
Sbjct: 63  SAMDNLQLAHWTCNRQKSDKLF-NVKQEEPKVLGNRNLPQSRDWSSYVS 110
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1461> which encodes the amino acid sequence <SEQ ID 1462>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4185(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 = 88/112 (78%), Positives = 102/112 (90%)

Query: 28  KLRADKKGTHRVAFEKNKRRLLKTAHLCGICGRPVDKSLKYPHPLSAAIDHIVPIAKGGH 87
           +LRADKKGTHRVAF++NK++LLK A +CGICG+PVDKSLKYPHPLSAAIDHIVPIAKGGH
Sbjct: 3   QLRADKKGTHRVAFDRNKKKLLKAATVCGICGKPVDKSLKYPHPLSAAIDHIVPIAKGGH 62

Query: 88  PSSIDNLQLTHWQCNRQKSDKLFINQTAVRATVVGNRNLPQSRDWSSYASKE 139
           PS+++NLQLTHWQCNRQKSDKLF NQ +      +GNRNLPQSRDWSS+A K+
Sbjct: 63  PSALENLQLTHWQCNRQKSDKLFANQASNEPKTIGNRNLPQSRDWSSFAFKK 114
```

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

Example 454

A DNA sequence (GBSx0491) was identified in *S.agalactiae* <SEQ ID 1463> which encodes the amino acid sequence <SEQ ID 1464>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4481(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 455

A DNA sequence (GBSx0492) was identified in *S.agalactiae* <SEQ ID 1465> which encodes the amino acid sequence <SEQ ID 1466>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2907(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:AAF43508 GB:AF145054 ORF15 [Streptococcus thermophilus
          bacteriophage 7201]
 Identities = 61/187 (32%), Positives = 90/187 (47%), Gaps = 31/187 (16%)

Query: 1   MNIEEAKKLIDKQSIGKGGVGDIPVVKTHIVKVLLDQIDQPQPEVPRFVADWYEKHKDSL 60
            MN +EA K I K+            +     + L D I   +P VP++VADWYE+HKD
Sbjct: 1   MNRDEAVKKIAKEGY----------ISIEHAEDLYDSIIT-KPVVPQYVADWYEEHKDEF 49

Query: 61  ECDL------YLYHMSIY--DEEVEKDDFYYWMQTSKNPVYTLINMHQFGYTIQKEKLYT 112
              +L      +  H++ Y  +E     DF  W   +KN +  L+NMHQFGY ++KEK YT
Sbjct: 50  YLNLHRVVRDPFEHLNAYYFNENPIDYDFACWYYNTKNAIQILVNMHQFGYEVKKEKRYT 109

Query: 113 VEIPN--PNERQLSFVLMRQLSGNVSIKVMHRDNLDLLKTDNDLQLTESEIRKDFDWAWQ 170
            V I N    E  L++   R+     +     RDN D  +T +    T  E+ ++  + W
Sbjct: 110 VRIRNLDDEETYLNYDKFRE-----TWVFYSRDNTDRFRTIH----THKEL-EEGGFGWV 159

Query: 171 FREEVVE 177
           F  E +E
Sbjct: 160 FDCEGIE 166
```

A related GBS nucleic acid sequence <SEQ ID 10927> which encodes amino acid sequence <SEQ ID 10928> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1467> which encodes the amino acid sequence <SEQ ID 1468>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
```

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3815(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 = 70/180 (38%), Positives = 98/180 (53%), Gaps = 30/180 (16%)

Query: 1   MNIEEAKKLIDKQSI-GKGGVGDIPVVKTHIVKVLLDQIDQPQPEVPRFVADWYEKHKDS 59
           MNIEEAK+L+D     GK       V+K   V+ ++DQ++QP+PEVP+ VADW E+ K+
Sbjct: 1   MNIEEAKELVDNSKFYGKTS----SVIKAE-VRDIIDQLNQPKPEVPQCVADWIEECKEE 55

Query: 60  LECDLYLYHMSIYDEEVEKDDFYYWMQTSKNPVYTLINMHQFGYTIQKEKLYTVEIPN-- 117
              DL L    ++         + W+  S      +      GYT++KEKLYTV++PN
Sbjct: 56  ---DLTL--KGLFSNSDMPAKIFDWIFGSDENCRLMAEAWINGYTVEKEKLYTVDLPNGQ 110

Query: 118 PNERQLSFVLMRQLSGNVSIKVMHRDNLDLLKTDNDLQLTESEIRKDFDWAWQFREEVVE 177
           P  R ++ +   Q                 L T+N ++LTESEIRKDF+WAWQF EEV E
Sbjct: 111 PLVRGINTLYFSQN----------------LATEN-VKLTESEIRKDFEWAWQFAEEVTE 153
```

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

Example 456

A DNA sequence (GBSx0493) was identified in *S.agalactiae* <SEQ ID 1469> which encodes the amino acid sequence <SEQ ID 1470>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5365(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 457

A DNA sequence (GBSx0494) was identified in *S.agalactiae* <SEQ ID 1471> which encodes the amino acid sequence <SEQ ID 1472>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.55   Transmembrane   34 -  50 (  31 -  54)

----- Final Results -----
               bacterial membrane --- Certainty=0.4418(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 9657> which encodes amino acid sequence <SEQ ID 9658> 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 1473> which encodes the amino acid sequence <SEQ ID 1474>. Analysis of this protein sequence reveals the following:

```
      Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.25   Transmembrane   26 -  42 (  20 -  49)

----- 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 no significant homology with any sequences in the GENPEPT database.

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

```
 Identities = 56/89 (62%), Positives = 71/89 (78%)

Query: 8  MTEQQMIDCLLYELAKKDKLNIRRNNIITFLSIVLMAISILNVALQDHYKSQITELRTQL 67
          MTE+QMIDCLLYEL KKDK   +++ II  L+++L+ +S L V+L+ +Y+ QI  LRTQL
Sbjct: 1  MTEEQMIDCLLYELVKKDKAIKKKSIIIAALTVMLIVVSGLCVSLKSYYEPQIYGLRTQL 60

Query: 68 SRTQKQLKRASDDRARQTKRIAELTGNGG 96
          SRTQKQLKRAS+   RQTKRIA+LT NGG
Sbjct: 61 SRTQKQLKRASEQNQRQTKRIADLTNNGG 89
```

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

Example 458

A DNA sequence (GBSx0495) was identified in *S.agalactiae* <SEQ ID 1475> which encodes the amino acid sequence <SEQ ID 1476>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2040(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 459

A DNA sequence (GBSx0496) was identified in *S.agalactiae* <SEQ ID 1477> which encodes the amino acid sequence <SEQ ID 1478>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3044(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:AAD37108 GB:AF109874 unknown [Bacteriophage Tuc2009]
 Identities = 50/143 (34%), Positives = 67/143 (45%), Gaps = 29/143 (20%)

Query: 1   MIPNFRAFNKETKKM-YG-VDGFELSVRKIYRCSLADDEFRCGRLETFHFVEDNFDDYIL 58
           MIP  RA++K+ ++M YG V+ F+ S+   YR                HF     +D
Sbjct: 1   MIPKLRAWDKQDERMSYGEVEYFDDSIN--YRFD--------------HFCTGADEDVEF 44

Query: 59  MQSTGMFDKNGVEIFDGDIVLTTRL-------IDY-TYKNFKGVVKMLEGRWLIDTGKDA 110
           MQSTG+ DKNGVEI++GDI+    +       I Y  Y    G   + EG  L    +
Sbjct: 45  MQSTGIKDKNGVEIYEGDILKLHAIFLAPDDKIGYLEYSPKYGYSIICEGNRLY---RQE 101

Query: 111 VGLWTEVDENEAIGNIYQNSELL 133
               T     E IGNIY+N ELL
Sbjct: 102 YWASTNKLNYEVIGNIYENPELL 124
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1479> which encodes the amino acid sequence <SEQ ID 1480>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4779(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 44/52 (84%), Positives = 47/52 (89%)

Query: 1  MIPNFRAFNKETKKMYGVDGFELSVRKIYRCSLADDEFRCGRLETFHFVEDN 52
          MIPNFR FNK+TKKMY +DGF+ S RKIYRCSLADDEFR GRLETFHFVEDN
Sbjct: 1  MIPNFRGFNKKTKKMYSIDGFKSSERKIYRCSLADDEFRSGRLETFHFVEDN 52
```

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

Example 460

A DNA sequence (GBSx0497) was identified in *S.agalactiae* <SEQ ID 1481> which encodes the amino acid sequence <SEQ ID 1482>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3843(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 9655> which encodes amino acid sequence <SEQ ID 9656> was also identified.

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

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

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

Example 461

A DNA sequence (GBSx0498) was identified in *S.agalactiae* <SEQ ID 1483> which encodes the amino acid sequence <SEQ ID 1484>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5189(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 9653> which encodes amino acid sequence <SEQ ID 9654> was also identified.

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

```
>GP:AAF43503 GB:AF145054 ORF10 [Streptococcus thermophilus
          bacteriophage 7201]
 Identities = 92/147 (62%), Positives = 121/147 (81%)

Query: 15  IEPKPQTRPKFSKFGTYEDPKMKRWRKEVSGWIEKNYDGPFFDDCIKVEVTFYMKAPKTL 74
           IEPKPQTRP+FSKFGTYEDPKMK WR+E S  IE+ YDG FF   I V+VTFYMKAP ++
Sbjct: 7   IEPKPQTRPRFSKFGTYEDPKMKAWRRECSRLIEQEYDGQFFYGPISVDVTFYMKAPLSV 66

Query: 75  SKEPTQRSKGKTIQIYQNFVRELIWHAKKPDIDNLIKAVFDSISDAGYDRIQKSGIVWSD 134
           SK+PT +++ KT   ++ F+ E +WH++KPDIDNLIKA+FDSIS AGY+++ K GIVW+D
Sbjct: 67  SKKPTPKARAKTWDAFKKFMAERLWHSRKPDIDNLIKALFDSISTAGYNKVDKKGIVWTD 126

Query: 135 DNIVCDLRAKKKYSQNPRIKVRIEEID 161
           D+IVC L A+K+YS+NPRI+  I+E++
Sbjct: 127 DSIVCKLSAQKRYSENPRIEFEIKELE 153
```

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

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

Example 462

A DNA sequence (GBSx0499) was identified in *S.agalactiae* <SEQ ID 1485> which encodes the amino acid sequence <SEQ ID 1486>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4007(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 463

A DNA sequence (GBSx0500) was identified in *S.agalactiae* <SEQ ID 1487> which encodes the amino acid sequence <SEQ ID 1488>. This protein is predicted to be pXO1-07. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3664(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:AAC38715 GB:AF030367 maturase-related protein [Streptococcus pneumoniae]
 Identities = 146/373 (39%), Positives = 216/373 (57%), Gaps = 18/373 (4%)

Query: 35  LYDKVYRKDILKVAWFYVKRNKGSAGIDDFTIEEIEAYGVQKFLDEIEDQLRNKKYQPKA 94
           L DK+  ++ +  A+  VK NKGSAGID  TIEE++ Y  Q +    ++ ++ +KY+P+
Sbjct: 4   LLDKILSRENMLEAYNQVKSNKGSAGIDGMTIEEMDNYLRQNWR-LTKELIKQRKYKPQP 62

Query: 95  VKRVYIPKANGKKRPLGIPTVRDRVVQTAVKIVIEPIFEADFQEFSYGFRPKRSANQAIR 154
           V +V IPK +G  R LGIPTV DR++Q A+  V+ PI E  F + SYGFRP RS  +AI
Sbjct: 63  VLKVEIPKPDGGIRQLGIPTVMDRMIQQAIVQVMSPICEPHFSDTSYGFRPNRSCEKAIM 122

Query: 155 EIYKYLNYGCEWVIDADLKGYFDTIPHDKLLLLVKERVTDKSIIKLLSLWLEAGIMEDNQ 214
           ++ +YLN G EW++D DL+ +FDT+P D+L+ LV   + D      L+  +L +G++ + Q
Sbjct: 123 KLLEYLNDGYEWIVDIDLEKFFDTVPQDRLMSLVHNIIEDGDTESLIRKYLHSGVIINGQ 182

Query: 215 VRSNILGTPQGGVISPLLANIYLNALDRYWKNNRLEGRGHDAHLIRYADDFVI-LCSNNP 273
               ++GTPQGG +SPLL+NI LN LD+      LE RG      +RYADD VI + S
Sbjct: 183 RYKTLVGTPQGGNLSPLLSNIMLNELDK-----ELEKRG--LRFVRYADDCVITVGSEAA 235

Query: 274 KKYYQYAKQRI--DKLGLTLNEEKTRIVHATEGFDFLGYTLRKSKSHKSGKYKTYYYPSR 331
             K   Y+  R     +LGL +N  KT+I    E   +LG+   KS       +      P +
Sbjct: 236 AKRVMYSVSRFIEKRLGLKVNMTKTKITRPRE-LKYLGFGFWKSSDGWKSR------PHQ 288

Query: 332 KSMKSIKGKVKDVIQTGQHLNLPDVMERLNPMLRGWANYFKAGNSKQHFKSIDNYVIYNL 391
            S++  K K+K + Q    ++L   +E+LN  +RGW NYF  GN K    SID  +   L
Sbjct: 289 DSVRRFKLKLKKLTQRKWSIDLTRRIEQLNLSIRGWINYFSLGNMKSIVASIDERLRTRL 348

Query: 392 TIMLRKKHKKSGK 404
           +++ K+ KK  +
Sbjct: 349 RMIIWKQWKKKSR 361
```

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 464

A DNA sequence (GBSx0501) was identified in *S.agalactiae* <SEQ ID 1489> which encodes the amino acid sequence <SEQ ID 1490>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3833(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 9651> which encodes amino acid sequence <SEQ ID 9652> was also identified.

A further related DNA sequence (GBSx2517) was identified in *S.agalactiae* <SEQ ID 7217> which encodes the amino acid sequence <SEQ ID 7218>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3833(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 1491> which encodes the amino acid sequence <SEQ ID 1492>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2299(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 = 113/163 (69%), Positives = 128/163 (78%), Gaps = 25/163 (15%)

Query: 1   MINNIVLVGRMTKDAELRYTPSNQAVATFSLAVNRNFKNQSGEREADFINCVIWRQQAEN 60
           MINN+VLVGRMTKDAELRYTPS  AVATF+LAVNR FK+Q+GEREADFINCVIWRQ AEN
Sbjct: 1   MINNVVLVGRMTKDAELRYTPSQVAVATFTLAVNRTFKSQNGEREADFINCVIWRQPAEN 60

Query: 61  LANWAKKGALVGITGRIQTRNYENQQGQRIYVTEVVAENFQLLESRNSQQ---------Q 111
           LANWAKKGAL+G+TGRIQTRNYENQQGQR+YVTEVVA+NFQ+LESR +++
Sbjct: 61  LANWAKKGALIGVTGRIQTRNYENQQGQRVYVTEVVADNFQMLESRATREGGSTGSFNGG 120

Query: 112 TNQSGNSSNSY----------------FGNANKMDISDDDLPF 138
            N + +SSNSY                FGN+N MDISDDDLPF
Sbjct: 121 FNNNTSSSNSYSAPAQQTPNFGRDDSPFGNSNPMDISDDDLPF 163
```

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

Example 465

A DNA sequence (GBSx0502) was identified in *S.agalactiae* <SEQ ID 1493> which encodes the amino acid sequence <SEQ ID 1494>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.33   Transmembrane   17 -  33 (  17 -  33)

----- Final Results -----
             bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has 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 466

A DNA sequence (GBSx0503) was identified in *S.agalactiae* <SEQ ID 1495> which encodes the amino acid sequence <SEQ ID 1496>. This protein is predicted to be p22 erf-like protein. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2469(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:BAA97824 GB:AB044554 orf 17 [Staphylococcus aureus prophage
          phiPV83]
 Identities = 93/183 (50%), Positives = 120/183 (64%), Gaps = 5/183 (2%)

Query: 1   MRKSESITEYAKAFCKAQLEVKQPLKDKDNPFFKSKYVPLENVTEAITTAFANNGISFSQ 60
           M KSE++ E  KA    + EVKQPLKDK+NPFFKSKYVPLENV EAI  A   +G+S++Q
Sbjct: 1   MNKSETVVEINKAMVAFRKEVKQPLKDKNNPFFKSKYVPLENVVEAIDEAATPHGLSYTQ 60

Query: 61  DPTTNTENGYIDVATLVMHTSGEWVEYGPLSVKPTKNDVQGAGSAITYAKRYALSAIFGI 120
               N  +G + VAT++MH SGE++EY P+ +   KN  QGAGS I+Y KRY+LSAIFGI
Sbjct: 61  W-ALNDVDGRVGVATMLMHESGEYIEYDPVFMNAEKNTPQGAGSLISYLKRYSLSAIFGI 119

Query: 121 TSDQDDDGNEDSKPNNSRQSPKATTKKTQKTGYQTPKISNIQIETYKSDLNDIAKATNQN 180
           TSDQDDDGNE S  NN   +PK  T +TQ    +T  I   ++ ++   +    K   QN
Sbjct: 120 TSDQDDDGNEASGKNN---NPKQQT-RTQWASSETIGILRKEVISFTKLIKGTDKEAPQN 175

Query: 181 VEE 183
           + E
Sbjct: 176 IVE 178
```

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 467

A DNA sequence (GBSx0504) was identified in *S.agalactiae* <SEQ ID 1497> which encodes the amino acid sequence <SEQ ID 1498>. This protein is predicted to be gp157. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3148(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:AAD44102 GB:AF115103 orf157 gp [Streptococcus thermophilus
          bacteriophage Sfi21]
```

```
 Identities = 59/160 (36%), Positives = 100/160 (61%), Gaps = 3/160 (1%)

Query: 1   MAYLYELEGIYAQLQSMDLDEETFQDTLDSIDFQSDLENNIEYFVKMLKNVQADAEKYKA 60
           MA LYEL G + ++ +M++D+ET  DTL++ID+ SD EN +E +VK++K+++AD E  K
Sbjct: 1   MATLYELTGQFLEIYNMEIDDETKLDTLEAIDWTSDYENKVEGYVKVIKSLEADIEARKN 60

Query: 61  EKEAFYKKQKQAEAKAEKYKETIRLAMELSQKKKVDAGMFKVSLRRSKKVEILDETKIPL 120
           EK+       K  ++K +K K  + ++M  + + +VD  +FK+   +SK V +++E K+P
Sbjct: 61  EKKRLDGLNKSDQSKIDKLKAALAISMTETGQTRVDTTLFKIGFHKSKAV-VVNEEKLPK 119

Query: 121 DYMQEKIEYKPMKAEISKALKSGIDISGVELIETESLQVK 160
           +Y  +   YKP K  + + LKSG  I G  L E  +L ++
Sbjct: 120 EY--QIATYKPDKKTLKELLKSGKHIEGATLEERRNLNIR 157
```

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

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

Example 468

A DNA sequence (GBSx0505) was identified in *S.agalactiae* <SEQ ID 1499> which encodes the amino acid sequence <SEQ ID 1500>. This protein is predicted to be tropomyosin 2. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4474(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 469

A DNA sequence (GBSx0506) was identified in *S.agalactiae* <SEQ ID 1501> which encodes the amino acid sequence <SEQ ID 1502>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4114(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 9649> which encodes amino acid sequence <SEQ ID 9650> was also identified.

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

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

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

Example 470

A DNA sequence (GBSx0507) was identified in *S.agalactiae* <SEQ ID 1503> which encodes the amino acid sequence <SEQ ID 1504>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3799(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 1505> which encodes the amino acid sequence <SEQ ID 1506>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3775(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 = 43/46 (93%), Positives = 46/46 (99%)

Query: 1  MTKQHRETLIWYRASHQEREKLLDFGLVDKSQYVTLLRQLRKKYAI 46
          MTKQHRETLIWYRASHQERE+LLDFGLVDK++YVTLLRQLRKKYAI
Sbjct: 1  MTKQHRETLIWYRASHQERERLLDFGLVDKARYVTLLRQLRKKYAI 46
```

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

Example 471

A DNA sequence (GBSx0508) was identified in *S.agalactiae* <SEQ ID 1507> which encodes the amino acid sequence <SEQ ID 1508>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4308(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 1509> which encodes the amino acid sequence <SEQ ID 1510>. Analysis of this protein sequence reveals the following:

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

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

```
                bacterial cytoplasm --- Certainty=0.4308(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 = 76/77 (98%), Positives = 76/77 (98%)

Query: 1  MDQEIFNFFNKQIKKDFGKTASKETFAKFASYCAEGIEKNGVKPIFNWINLYAFGTGMTT 60
          MDQEIFNFFNKQIKKDFGKTASKETFAKFASYCAEGIEKNGVKPIFNWINLYAFGTGMTT
Sbjct: 1  MDQEIFNFFNKQIKKDFGKTASKETFAKFASYCAEGIEKNGVKPIFNWINLYAFGTGMTT 60

Query: 61 AEADRLRIERYKQENTL 77
          AEADRLRIERYKQEN L
Sbjct: 61 AEADRLRIERYKQENAL 77
```

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

Example 472

A DNA sequence (GBSx0509) was identified in *S.agalactiae* <SEQ ID 1511> which encodes the amino acid sequence <SEQ ID 1512>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.2706(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 1513> which encodes the amino acid sequence <SEQ ID 1514>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.3316(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 = 52/127 (40%), Positives = 75/127 (58%), Gaps = 1/127 (0%)

Query: 160 EDRFVDVVEANLGRGLVKFEFDMINDYLIGQNVSKDLFLEAVKVAVANNVRKFNYIARIL 219
           E +  +  +   GR +  FE + I  ++   N+  ++   A++ AV NN   + YI +IL
Sbjct: 3   EKKLFENFQLTFGRMISPFEIEDIQKWIHEDNMPIEVVNLALREAVENNKISWKYINKIL 62

Query: 220 DNWINDGIKTPEQAYQAQRDFKAKKANKTMQSQSNVPSWSNPDYKGPDLKEFALGSIDDI 279
            +W   G  T E+     + F   K  +++ + SNVPSWSNPDYK PDL+EFALGS+D I
Sbjct: 63  VDWYKSGDITVEKVRDRLQRFDDSKKQRSVTT-SNVPSWSNPDYKEPDLEEFALGSMDGI 121

Query: 280 EDGSGDF 286
           EDGSGDF
Sbjct: 122 EDGSGDF 128
```

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

Example 473

A DNA sequence (GBSx0510) was identified in *S.agalactiae* <SEQ ID 1515> which encodes the amino acid sequence <SEQ ID 1516>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.63   Transmembrane   13 -  29 (  11 -  31)

----- Final Results -----
               bacterial membrane --- Certainty=0.3251(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 9647> which encodes amino acid sequence <SEQ ID 9648> was also identified.

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

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

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

Example 474

A DNA sequence (GBSx0511) was identified in *S.agalactiae* <SEQ ID 1517> which encodes the amino acid sequence <SEQ ID 1518>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5822(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 475

A DNA sequence (GBSx0512) was identified in *S.agalactiae* <SEQ ID 1519> which encodes the amino acid sequence <SEQ ID 1520>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4175(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 476

A DNA sequence (GBSx0513) was identified in *S.agalactiae* <SEQ ID 1521> which encodes the amino acid sequence <SEQ ID 1522>. This protein is predicted to be P1-antirepressor homolog. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3411(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 9645> which encodes amino acid sequence <SEQ ID 9646> was also identified.

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

```
>GP:AAG31333 GB:AF182207 ORF 169a [Bacteriophage mv4]
 Identities = 88/167 (52%), Positives = 122/167 (72%)

Query: 100 MLQRNEKSKQVRKYFIQVEKDFNSPEKIMARALLMADKKITNLTMENNQLQLDLKEAQKQ 159
           M+ +  K K++R+YFIQVEK++NSPE I+ RAL +++ +I  L  +N  L L L+E+ K+
Sbjct: 1   MMSKTAKGKEIRQYFIQVEKNWNSPEMIIQRALEISNARIQELQAQNKSLTLQLEESNKK 60

Query: 160 ARYLDLIIESKGALRVTQIAADYGMSVNKFNKTLLEFGVQHKVNGQWILYKRHMGKGYTD 219
           A YLD+I+ +   L  TQIAADYG S   FN+ L E G+QHKVNGQWILYK +MGKGY
Sbjct: 61  ASYLDIILGTPDLLATTQIAADYGYSARTFNQLLKEVGIQHKVNGQWILYKAYMGKGYVQ 120

Query: 220 SHTFDYQDKNGHTRANVTTTWTQKGRLFLYELLKDNNILPLIEQEDI 266
           S +F ++D+ GH R+  +T WTQKGR  +Y++LK+N  LPLIE++DI
Sbjct: 121 SKSFAFKDRKGHDRSKPSTYWTQKGRKLIYDVLKENGTLPLIERDDI 167
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1523> which encodes the amino acid sequence <SEQ ID 1524>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4214(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 = 130/249 (52%), Positives = 163/249 (65%), Gaps = 14/249 (5%)

Query: 19  MNQLINITLNENQEPVVSGRDLHNVLNIKTQYTKWLERMSEYGFEENVDYIAISQKRLTA 78
           MNQLIN+TLNENQEPVVSGRDLH VL IKTQYTKWLERMSEYGF EN D++AISQKRLTA
Sbjct: 1   MNQLINVTLNENQEPVVSGRDLHKVLEIKTQYTKWLERMSEYGFVENEDFMAISQKRLTA 60

Query: 79  QGNRTEYIDHVLKLDMAKEIAMLQRNEKSKQVRKYFIQVEKDFNSPEKIMARALLMADKK 138
           QGN+TEY DHVLKLDMAKEIAMLQRNEKSK+VRKYFIQVEKDFNSPEKIMARALLMADKK
Sbjct: 61  QGNQTEYTDHVLKLDMAKEIAMLQRNEKSKEVRKYFIQVEKDFNSPEKIMARALLMADKK 120

Query: 139 ITNLTMENNQLQLDLKEAQKQARYLDLIIESKGALRVTQIAA-----DYGMSVNKFNKTL 193
           +       ++L+  ++  + +  + D +  S  ++ V ++A          +   +    L
Sbjct: 121 V-------HKLEAQIEADRPKVLFADAVSASHTSILVGELAKLLKQNGVNIGATRLFTWL 173
```

```
Query: 194 LEFGVQHKVNGQ-WIL-YKRHMGKGYTDSHTFDYQDKNGHTRANVTTTWTQKGRLFLYEL 251
           + G   K NG+ W +  ++ +  G            +GH   + T   T KG+ +
Sbjct: 174 RKHGYLIKRNGRDWNMPTQKSVELGLIRVKETSITHSDGHITVSKTPLVTGKGQQYFINK 233

Query: 252 LKDNNILPL 260
             +   LP+
Sbjct: 234 FLNQEYLPV 242
```

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

Example 477

A DNA sequence (GBSx0514) was identified in *S.agalactiae* <SEQ ID 1525> which encodes the amino acid sequence <SEQ ID 1526>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4205(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 1527> which encodes the amino acid sequence <SEQ ID 1528>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> 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>
```

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

```
 Identities = 21/63 (33%), Positives = 31/63 (48%), Gaps = 1/63 (1%)

Query: 1  MQQFNLKQLREKKGFTQNELADKANVSRSLVVGLETGSYSETSTASLKKLAKALDVKIKD 60
          M+   LK  R  K  +Q  LAD   VSR  +  +E G Y+ T    +  + + LD  + D
Sbjct: 1  MKNLKLKAARAGKDLSQQALADLVGVSRQTIAAVEKGDYNPTINLCI-AICRVLDKTLDD 59

Query: 61 LFF 63
          LF+
Sbjct: 60 LFW 62
```

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

Example 478

A DNA sequence (GBSx0515) was identified in *S.agalactiae* <SEQ ID 1529> which encodes the amino acid sequence <SEQ ID 1530>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0396(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:BAA17582 GB:D90907 hypothetical protein [Synechocystis sp.]
 Identities = 45/164 (27%), Positives = 79/164 (47%), Gaps = 33/164 (20%)

Query: 102 EEEELRNLFTKLIASSMDKSKNEFNHPSFIEIIKQFDKIDAQNFKIISDLYFKKGFVATG 161
           ++E L+ L+  L+AS++ +S    +  SF+E++KQ D +DA+   ++  L+ +
Sbjct: 97  DDENLQTLWANLLASALTESDRTNSTKSFVEVLKQVDIVDAELLNVLYLLHLRV------ 150

Query: 162 TYYTTIIGQDKPLEHIASHVFVDNLEQNDIAIQSSSLTNLERLGLIQINY--KAHVDEKE 219
                  KP E   ++   D+ + N + I S +L NLERLGL+ I+      VDE+
Sbjct: 151 --------MAKPDEFTYAN---DSRKYNIVQI-SVALNNLERLGLLIIHKYDDTPVDEEA 198

Query: 220 YYNILNNSFITKKNSELKEQNKRVLTNLGMITLTLFGVRFSKTC 263
             +I    ++   N   K            ++LTLFG+ F + C
Sbjct: 199 RISIW---YMQDGNRSFKAH----------VSLTLFGIHFMRVC 229
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1531> which encodes the amino acid sequence <SEQ ID 1532>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0151(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 64/215 (29%), Positives = 105/215 (48%), Gaps = 23/215 (10%)

Query: 65  QKLAKEIQDVVSKNIE-NLQEPSLSIAGPALEASKFYLEEEELRNLFTKLIASSMDKSKN 123
           +K   EI    SK  + +L+EP   I  PA+  S+ YL  E LRN+F + IAS+ ++ K
Sbjct: 72  EKFKNEIDCEFSKIPQTSLKEPVEYILYPAINESEQYLSNETLRNMFARTIASTFNQDKE 131

Query: 124 EFNHPSFIEIIKQFDKIDAQNFKIISDLYFKKGFVATGTYYTTIIGQDKPLEHI------ 177
           +  H +F++IIKQ   +DAQN  +I+                  IG     E++
Sbjct: 132 KDLHSAFVQIIKQMTPLDAQNLLLINQ-------EGNNLIANLQIGVHYSKENLSGTVNK 184

Query: 178 ASHVFVDNLEQNDIAIQSSSLTNLERLGLIQINYKAHVDEKEYYNILNNSFITKKNSELK 237
           A+++++  L+ +   I +SS+ NL RLGLI+++Y  +  +  Y +I   +      SE+
Sbjct: 185 ANNIYLSKLDYSPDII-ASSIDNLTRLGLIKVDYLHYPLDSNYESIKQTTIYKSLESEIN 243

Query: 238 EQNKRVLTNL--------GMITLTLFGVRFSKTCL 264
             N    +N          G ++LT FG +F   CL
Sbjct: 244 TLNLFKTSNTKYDIKIEKGKVSLTDFGKKFISVCL 278
```

SEQ ID 1530 (GBS261) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 8; MW 31kDa).

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

Example 479

A DNA sequence (GBSx0516) was identified in *S.agalactiae* <SEQ ID 1533> which encodes the amino acid sequence <SEQ ID 1534>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.55   Transmembrane    3 -  19 (   1 -  26)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.4418(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 480

A DNA sequence (GBSx0517) was identified in *S.agalactiae* <SEQ ID 1535> which encodes the amino acid sequence <SEQ ID 1536>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.99   Transmembrane   35 -  51 (  30 -  51)

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

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

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1537> which encodes the amino acid sequence <SEQ ID 1538>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.94   Transmembrane   31 -  47 (  30 -  51)

----- Final Results -----
               bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
Identities = 45/52 (86%), Positives = 48/52 (91%)

Query: 1  MNWKKLMLGDLEHTFTSRDGKEKTSVEFEGGVLPALLVLGGITWLIAWLITK 52
          MNWKKLM GDLEHTFT+ DGKEKTS+EFEGGVLPALLVLGGI W+IAW ITK
Sbjct: 1  MNWKKLMFGDLEHTFTNHDGKEKTSIEFEGGVLPALLVLGGIAWMIAWFITK 52
```

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

Example 481

A DNA sequence (GBSx0518) was identified in *S.agalactiae* <SEQ ID 1539> which encodes the amino acid sequence <SEQ ID 1540>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3445(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 482

A DNA sequence (GBSx0519) was identified in *S.agalactiae* <SEQ ID 1541> which encodes the amino acid sequence <SEQ ID 1542>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3934(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 483

A DNA sequence (GBSx0520) was identified in *S.agalactiae* <SEQ ID 1543> which encodes the amino acid sequence <SEQ ID 1544>. This protein is predicted to be repressor 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=0.0905(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 9643> which encodes amino acid sequence <SEQ ID 9644> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1545> which encodes the amino acid sequence <SEQ ID 1546>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3117(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 = 175/264 (66%), Positives = 207/264 (78%), Gaps = 19/264 (7%)
```

```
Query: 34  LGKYIKKYRDTNNLSMAEFAKESGISKAY--VSILEKNRDPRNGKEIIPSIPIIKKVSDT 91
           LG  I+K R+  N++  E ++  G+ K Y  VS  EKN +   GK++        KK+++
Sbjct: 24  LGDRIRKLREGRNMTQTELSEILGM-KTYTTVSKWEKNENFPKGKDL-------KKLAEI 75

Query: 92  IGISFDDLLNSLDENQIVALNETKTEKNLTSSTLQKITSTSSQLEQPRQEKVLSFANEQL 151
             ++ D LL          L ++K  K    +   +I S  +QLEQPRQEKVL+FANEQL
Sbjct: 76  FNVTSDYLLG---------LTDSKLGKITIQNEQPEIVSIYNQLEQPRQEKVLNFANEQL 126

Query: 152 EEQNKVVSMFDRKVEETENYITDYVEGLVAAGLGAYQEDNLHMEVKLRADDVPDKYDTIA 211
           EEQNK VS+FD+K EETE+YITDYVEGLVAAGLGAYQEDNLHM+VKLR+DDVPD+YDTIA
Sbjct: 127 EEQNKTVSIFDKKSEETEDYITDYVEGLVAAGLGAYQEDNLHMKVKLRSDDVPDEYDTIA 186

Query: 212 KVAGNSMEPLIQDNDLLFVKVSSQVDMNDIGIFQVNGKNFVKKLKRDYDGAWYLQSLNKS 271
           KVAG+SMEPLIQDNDLLF+KVSSQVDMNDIGIFQVNGKNFVKKLKRDYDGAWYLQSLNKS
Sbjct: 187 KVAGDSMEPLIQDNDLLFIKVSSQVDMNDIGIFQVNGKNFVKKLKRDYDGAWYLQSLNKS 246

Query: 272 YEEIYLSENDNIRTIGEVVDIYRE 295
           YEEIYLS++D+IRTIGEVVDIYRE
Sbjct: 247 YEEIYLSKDDDIRTIGEVVDIYRE 270
```

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

Example 484

A DNA sequence (GBSx0521) was identified in *S.agalactiae* <SEQ ID 1547> which encodes the amino acid sequence <SEQ ID 1548>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3760(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 485

A DNA sequence (GBSx0522) was identified in *S.agalactiae* <SEQ ID 1549> which encodes the amino acid sequence <SEQ ID 1550>. This protein is predicted to be integrase (ripX). Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2719(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:CAB96616 GB:AJ400629 integrase [Streptococcus pneumoniae
          bacteriophage MM1]
 Identities = 36/59 (61%), Positives = 48/59 (81%), Gaps = 1/59 (1%)
```

```
Query: 2   KIYGDYHTHLFRHSHISFLAEKGIPLNAIMDRVGHSDPKTTLSIYSHTTVNMKE-IINK 59
           KI  +  +H+FRHSHISFLAE G+P+ +IMDRVGHS+ K TL IYSHTT +M++ ++NK
Sbjct: 312 KIEKNLSSHIFRHSHISFLAESGLPIKSIMDRVGHSNAKMTLEIYSHTTEDMEDKLVNK 370
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1551> which encodes the amino acid sequence <SEQ ID 1552>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2719(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 = 63/71 (88%), Positives = 66/71 (92%)

Query: 1  MKIYGDYHTHLFRHSHISFLAEKGIPLNAIMDRVGHSDPKTTLSIYSHTTVNMKEIINKQ 60
          +KIYGDYHTHLFRHSHISFLAEKGIPLNAIMDRVGHSDPKTTLSIYSHTTVNMKEIINKQ
Sbjct: 1  LKIYGDYHTHLFRHSHISFLAEKGIPLNAIMDRVGHSDPKTTLSIYSHTTVNMKEIINKQ 60

Query: 61 TAPFVPLLKSE 71
          T PF   +K +
Sbjct: 61 TDPFKTGIKQK 71
```

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

Example 486

A DNA sequence (GBSx0523) was identified in *S.agalactiae* <SEQ ID 1553> which encodes the amino acid sequence <SEQ ID 1554>. This protein is predicted to be 50S ribosomal protein L19 (rplS). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3331(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 9641> which encodes amino acid sequence <SEQ ID 9642> was also identified.

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

```
>GP:AAC01534 GB:U88973 ribosomal protein L19 [Streptococcus thermophilus]
 Identities = 110/115 (95%), Positives = 112/115 (96%)

Query: 25  MNPLIQSLTEGQLRSDIPEFRAGDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISEMYT 84
           MNPLIQSLTEGQLR+DIP FR GDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISEMYT
Sbjct: 1   MNPLIQSLTEGQLRTDIPSFRPGDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISEMYT 60

Query: 85  VRKISGGIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYYLRALQGKAARIKEIRR 139
           VRKIS GIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYYLRALQGKAARIKEIR+
Sbjct: 61  VRKISSGIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYYLRALQGKAARIKEIRK 115
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1555> which encodes the amino acid sequence <SEQ ID 1556>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4849(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 = 111/115 (96%), Positives = 113/115 (97%)

Query: 25  MNPLIQSLTEGQLRSDIPEFRAGDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISEMYT 84
           MNPLIQSLTEGQLRSDIP FR GDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISEMYT
Sbjct: 1   MNPLIQSLTEGQLRSDIPNFRPGDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISEMYT 60

Query: 85  VRKISGGIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYYLRALQGKAARIKEIRR 139
           VRKISGGIGVERTFPIHTPRVDKIEV+R+GKVRRAKLYYLRALQGKAARIKEIRR
Sbjct: 61  VRKISGGIGVERTFPIHTPRVDKIEVIRHGKVRRAKLYYLRALQGKAARIKEIRR 115
```

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

Example 487

A DNA sequence (GBSx0524) was identified in *S.agalactiae* <SEQ ID 1557> which encodes the amino acid sequence <SEQ ID 1558>. This protein is predicted to be ISL2 protein. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> 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:CAC18596 GB:AJ278419 IS1381 transposase [Streptococcus pneumoniae]
 Identities = 111/129 (86%), Positives = 117/129 (90%)

Query: 1   MKAQAIVTSQGRIVSLDIAVNYCHDMKLFKMSRRNIGQAAKILADSGYQGIMKMYSQAQT 60
           MK QAIVTSQGRIVSLDI VNYCHDMKLFKMSRRNIGQA KILADSGYQG+MK+Y QAQT
Sbjct: 1   MKTQAIVTSQGRIVSLDITVNYCHDMKLFKMSRRNIGQAGKILADSGYQGLMKIYPQAQT 60

Query: 61  PRKSSKLKPLTLEDKTYNHTLSKERIKVENIFAKVKTFKIFSTTYRNRRKRFGLRMNLIA 120
            RKSSKLKPLT+EDK  NH LSKER KVENIFAKVKTFK+FSTTYR+ RKRFGLRMNL A
Sbjct: 61  SRKSSKLKPLTVEDKACNHALSKERSKVENIFAKVKTFKMFSTTYRSHRKRFGLRMNLSA 120

Query: 121 GMINRELGF 129
           G+IN ELGF
Sbjct: 121 GIINHELGF 129
```

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 488

A DNA sequence (GBSx0526) was identified in *S.agalactiae* <SEQ ID 1559> which encodes the amino acid sequence <SEQ ID 1560>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.99   Transmembrane   81 -  97 (  67 - 107)
   INTEGRAL    Likelihood = -6.32   Transmembrane    8 -  24 (   6 -  25)
   INTEGRAL    Likelihood = -2.76   Transmembrane  120 - 136 ( 120 - 136)

----- Final Results -----
               bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:BAB04382 GB:AP001509 unknown conserved protein in others
          [Bacillus halodurans]
 Identities = 53/150 (35%), Positives = 82/150 (54%), Gaps = 1/150 (0%)

Query: 1   MLNPYKRIFTLGLLATFLLFIFHFGRYSGLGTNLIEASFTNKNLYDYDWLLKLCLTVITL 60
           M N   R F  GL+   L +I     Y+G G +++E SFT +++  Y +L KL  T +T+
Sbjct: 251 MKNHTVRAFVGGLIIVALTYIIGSYDYNGRGLDMLEDSFT-QDVPPYAFLAKLVFTAVTM 309

Query: 61  AAGYQGGEVTPLFAIGASLGVIIAPILGLPVILVAALGYTSVFGSATNTLLGPILIGGEV 120
             G+ GGE  PLF +GA+LG  +   + LP+  +AALG    FG   NT +   L+G E+
Sbjct: 310 GMGFVGGEAIPLFFVGATLGNTLHAFIDLPLSFLAALGMIVTFGGGANTPIAAFLLGVEM 369

Query: 121 FGFANTPYFVIVCLVAYSISHAHTIYGAQS 150
           F      +F + CL +Y  S  H ++ +Q+
Sbjct: 370 FNGKGIEFFFVACLTSYLFSGHHGLWPSQT 399
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1561> which encodes the amino acid sequence <SEQ ID 1562>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.99   Transmembrane   56 -  72 (  53 -  78)
   INTEGRAL    Likelihood = -7.17   Transmembrane  337 - 353 ( 327 - 355)
   INTEGRAL    Likelihood = -6.74   Transmembrane  264 - 280 ( 260 - 282)
   INTEGRAL    Likelihood = -6.16   Transmembrane  167 - 183 ( 161 - 187)
   INTEGRAL    Likelihood = -5.26   Transmembrane  223 - 239 ( 217 - 242)
   INTEGRAL    Likelihood = -5.10   Transmembrane   20 -  36 (  19 -  42)
   INTEGRAL    Likelihood = -0.37   Transmembrane  102 - 118 ( 102 - 119)
   INTEGRAL    Likelihood = -0.16   Transmembrane  300 - 316 ( 300 - 316)

----- 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 databases:

```
>GP:BAB04382 GB:AP001509 unknown conserved protein in others
          [Bacillus halodurans]
 Identities = 129/397 (32%), Positives = 210/397 (52%), Gaps = 14/397 (3%)

Query: 20  VLGLVGLALPIGGAVGVVDVIFGKGLLFLSEYRDHHLFLLLPFLALAGLVIVFLYDKLG- 78
           +L  +   + IG  VG    +        L E R++  + +L FL LAGL + +LY K G
Sbjct: 9   LLTWIFFGIMIGAIVGSATALLLTVNDHLGETRENRPWFVL-FLPLAGLALGYLYMKAGT 67

Query: 79  ---KEVRQGMGLVFQVGHGQKNQIPPMLIPLILFSTWVTHLFGASAGREGVAVQIGATIS 135
              E+ +G  LV +   G K ++   L PL+   T++T LFG S GREG A+Q+G  +++
Sbjct: 68  SAGNELYKGNNLVIESVQG-KGKMLLRLGPLVYLGTFMTILFGGSTGREGAAIQMGGSVA 126
```

```
Query: 136 HYCRR-FVTSQEAARHLLIMGMAAGFAGLFQTPIAAVVFALEVLLVGTLRYSALLPSLVA 194
               + F        R LL+ G++AGF    F TPI A +F +E+   +G L++ AL+P LVA
Sbjct: 127 EAVNKLFKVKLIDTRILLMGGISAGFGAAFGTPITAAIFGMEMASLGRLKFEALVPCLVA 186

Query: 195 AYVASWTSHALG-LEKFTIVLEETLTITPLTLVKLIGLGLIFGLVGNSFAYLL-GWFKPY 252
           ++V  +T+      +E     ++      ++ LT  K+I L ++F LV    +  L  G  K
Sbjct: 187 SFVGHYTTEKFWHVEHEKFIIATVPEVSALTFSKVILLAIVFSLVSVLYCQLRHGIHKLS 246

Query: 253 LSQKLPNPYFRIAFIGALLSICL--MIGHVGRYSGLGTNLIAAAFSGQTILTYDWLLKMI 310
               + N    R AF+G L+ + L  +IG    Y+G G +++   +F+ Q +  Y +L K++
Sbjct: 247 EKYTMKNHTVR-AFVGGLIIVALTYIIGSYD-YNGRGLDMLEDSFT-QDVPPYAFLAKLV 303

Query: 311 VTVISLSAGFQGGEVTPLFAIGASLGIVLAPYLGLPVLLVAALGYTTVFGSATNTFWAPI 370
            T +++   GF GGE   PLF +GA+LG  L  ++ LP+   +AALG    FG   NT  A
Sbjct: 304 FTAVTMGMGFVGGEAIPLFFVGATLGNTLHAFIDLPLSFLAALGMIVTFGGGANTPIAAF 363

Query: 371 FIGIEVFGPENALAYFVTSAAAYMVSHRHSIYSYQKV 407
           +G+E+F  +    +FV    +Y+ S  H ++  Q +
Sbjct: 364 LLGVEMFNGKGIEFFFVACLTSYLFSGHHGLWPSQTI 400
```

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

```
 Identities = 91/147 (61%), Positives = 111/147 (74%)

Query: 3   NPYKRIFTLGLLATFLLFIFHFGRYSGLGTNLIEASFTNKNLYDYDWLLKLCLTVITLAA 62
           NPY RI  +G L +  L I H GRYSGLGTNLI A+F+ + +  YDWLLK+ +TVI+L+A
Sbjct: 259 NPYFRIAFIGALLSICLMIGHVGRYSGLGTNLIAAAFSGQTILTYDWLLKMIVTVISLSA 318

Query: 63  GYQGGEVTPLFAIGASLGVIIAPILGLPVILVAALGYTSVFGSATNTLLGPILIGGEVFG 122
           G+QGGEVTPLFAIGASLG+++AP LGLPV+LVAALGYT+VFGSATNT   PI IG EVFG
Sbjct: 319 GFQGGEVTPLFAIGASLGIVLAPYLGLPVLLVAALGYTTVFGSATNTFWAPIFIGIEVFG 378

Query: 123 FANTPYFVIVCLVAYSISHAHTIYGAQ 149
             N   + +    AY +SH H+IY  Q
Sbjct: 379 PENALAYFVTSAAAYMVSHRHSIYSYQ 405
```

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

Example 489

A DNA sequence (GBSx0527) was identified in *S.agalactiae* <SEQ ID 1563> which encodes the amino acid sequence <SEQ ID 1564>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.65   Transmembrane   47 -  63 (  45 -  70)
   INTEGRAL    Likelihood = -5.04   Transmembrane  219 - 235 ( 208 - 237)
   INTEGRAL    Likelihood = -3.35   Transmembrane  168 - 184 ( 168 - 187)
   INTEGRAL    Likelihood = -0.48   Transmembrane  141 - 157 ( 141 - 157)

----- Final Results -----
              bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9317> which encodes amino acid sequence <SEQ ID 9318> was also identified.

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

```
>GP:BAB04382 GB:AP001509 unknown conserved protein in others
          [Bacillus halodurans]
 Identities = 75/223 (33%), Positives = 119/223 (52%), Gaps = 18/223 (8%)
```

```
Query: 17  FSLLIGGVVGAITAVFGRVLLFLTAFRSDYIAYLLPFLSIVGLFIVFVYQKFGGKS---- 72
           F ++IG +VG+ TA+   V   L   R +   ++L FL + GL + ++Y K G  +
Sbjct: 15  FGIMIGAIVGSATALLLTVNDHLGETRENRPWFVL-FLPLAGLALGYLYMKAGTSAGNEL 73

Query: 73  VKGMGLVFEVGHGNEETIPKRLVPLVILTTWLTHLFGGSAGREGVAVQIGATVSHYFQKY 132
            KG  LV E   G  + +  RL PLV L T++T LFGGS GREG A+Q+G +V+    K
Sbjct: 74  YKGNNLVIESVQGKGKML-LRLGPLVYLGTFMTILFGGSTGREGAAIQMGGSVAEAVNKL 132

Query: 133 CRLQNASQLFLVM-GMAAGFAGLFQTPLAATFFAIEVLVVGRLMVSYVLPSLIAALTANF 191
            +++      L+M G++AGF   F TP+ A  F +E+  +GRL    ++P L+A+   ++
Sbjct: 133 FKVKLIDTRILLMGGISAGFGAAFGTPITAAIFGMEMASLGRLKFEALVPCLVASFVGHY 192

Query: 192 VSHSLGLEKFSH------SIATSMALTPDIILKLLVLGLCFGL 228
           +    EKF H       IAT   ++     K+++L + F L
Sbjct: 193 TT-----EKFWHVEHEKFIIATVPEVSALTFSKVILLAIVFSL 230
```

There is also homology to SEQ ID 1562.

A related GBS gene <SEQ ID 8577> and protein <SEQ ID 8578> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:       9.66
GvH: Signal Score (-7.5): -1.12
     Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 7 value: -10.99 threshold:  0.0
   INTEGRAL    Likelihood =-10.99   Transmembrane  328 - 344 ( 314 - 354)
   INTEGRAL    Likelihood = -8.65   Transmembrane   47 -  63 (  45 -  70)
   INTEGRAL    Likelihood = -6.32   Transmembrane  255 - 271 ( 253 - 272)
   INTEGRAL    Likelihood = -4.41   Transmembrane  214 - 230 ( 208 - 238)
   INTEGRAL    Likelihood = -3.35   Transmembrane  168 - 184 ( 168 - 187)
   INTEGRAL    Likelihood = -2.76   Transmembrane  367 - 383 ( 367 - 383)
   INTEGRAL    Likelihood = -0.48   Transmembrane  141 - 157 ( 141 - 157)
   PERIPHERAL  Likelihood =  0.42       94
modified ALOM score:   2.70

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
ORF01989(349 - 1491 of 1794)
GP|4512350|dbj|BAA75315.1||AB011836(15 - 399 of 424) similar to Bordetella paraperlussis
transposase for insertion sequence element(27%-identity) {Bacillus halodurans}
PIR|T44296|T44296 hypothetical protein [imported] - Bacillus halodurans
%Match = 15.4
%Identity = 33.4  %Similarity = 54.7
Matches = 129  Mismatches = 167  Conservative Sub.s = 82

222       252       282       312       342       372       402       432
MY*RKSKTINLTM*YEQLSKTL*QNLVFIKRRIL*TVIKRFDNYAQYVLVLGFSLLIGGVVGAITAVFGRVLLFLTAFRS
                                                | ::|| :||: ||:: |   |   |
                                              MNKTFWLTLLTWIFFGIMIGAIVGSATALLLTVNDHLGETRE
                                                      10        20        30        40

462       492       513       540       570       600       630       660
DYIAYLLPFLSIVGLFIVFVYQKFG---GKSV-KGMGLVFEVGHGNEETIPKRLVPLVILTTWLTHLFGGSAGREGVAVQ
:   ::| || : || : ::| | |   | : ||  || | :| :  :  || ||| | |::| ||||| |||| |:|
NRPWFVL-FLPLAGLALGYLYMKAGTSAGNELYKGNNLVIESVQG-KGKMLLRLGPLVYLGTFMTILFGGSTGREGAAIQ
                  60        70        80        90       100       110       120
```

```
690       720       747       777       807       837       867       894
IGATVSHYFQKYCRLQNASQLFLVMG-MAAGFAGLFQTPLAATFFAIEVLVVGRLMVSYVLPSLIAALTANFVSHSL-GL
:| :|:    |  :::      |:|| ::|||   | ||: |  | :|:  :|||    ::| |:|::  :: :  :  :
MGGSVAEAVNKLFKVKLIDTRILLMGGISAGFGAAFGTPITAAIFGMEMASLGRLKFEALVPCLVASFVGHYTTEKFWHV
      130       140       150       160       170       180       190       200

924       954       984       1014      1041      1071      1101      1131
EKFSHSIATSMALTPDIILKLLVLGLCFGLCGNLFAYLLAKA-KLIASSRLLNPYKRIFTLGLLATFLLFIFHFGRYSGL
|     |||   ::     |:::| : | |   |:  |      ||      : |   | |  ||:   | :|      |:|
EHEKFIIATVPEVSALTFSKVILLAIVFSLVSVLYCQLRHGIHKLSEKYTMKNHTVRAFVGGLIIVALTYIIGSYDYNGR
      210       220       230       240       250       260       270       280

1161      1191      1221      1251      1281      1311      1341      1371
GTNLIEASFTNKNLYDYDWLLKLCLTVITLAAGYQGGEVTPLFAIGASLGVIIAPILGLPVILVAALGYTSVFGSATNTL
| :::| |||  :::  | :| || :| :|:  |: |||  ||| :||:|| :   : ||: ::|||     ||   ||
GLDMLEDSFT-QDVPPYAFLAKLVFTAVIMGMGFVGGEAIPLFFVGATLGNTLHAFIDLPLSFLAALGMIVTFGGGANTP
      290       300       310       320       330       340       350

1401      1431      1461      1491      1521      1551      1581      1611
LGPILIGGEVFGFANTPYFVIVCLVAYSISHAHTIYGAQSR*LVMSFKRVYQFVERNIPFSFLFS*SL*KWSLSIL*MQK
:   |:| |:|      :| : || :|  |  | :: :|:
IAAFLLGVEMFNGKGIEFFFVACLTSYLFSGHHGLWPSQTIYEPKSRLYGVRKGETIKRTEEMKE
      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 490

A DNA sequence (GBSx0528) was identified in *S.agalactiae* <SEQ ID 1565> which encodes the amino acid sequence <SEQ ID 1566>. 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
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1567> which encodes the amino acid sequence <SEQ ID 1568>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2356(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 = 45/91 (49%), Positives = 62/91 (67%)

Query: 1  MELEEIRQEIDEIDQQLVSLLETRMGLILEVIAFKKKHRLPVLDNNRENEVLNNVLKKVQ 60
          M LE+IRQEI+ ID  LV+LLE RM L+ +V A+K  + LPVLD  REN++L+ V   V+
Sbjct: 1  MRLEKIRQEINGIDHHLVALLEKRMALVEQVTAYKLANHLPVLDQARENQILDRVSYLVK 60

Query: 61 NHQFDDVIRATFKDIMTESRVYQKENIVDGD 91
          +  F+  I  TFK IM+ SR YQ +++  GD
Sbjct: 61 DQAFEPAIHETFKTIMSLSRQYQTQHLTGGD 91
```

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

Example 491

A DNA sequence (GBSx0529) was identified in *S.agalactiae* <SEQ ID 1569> which encodes the amino acid sequence <SEQ ID 1570>. This protein is predicted to be neuraminidase. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.35   Transmembrane   28 -  44 (  28 -  47)

----- Final Results -----
               bacterial membrane --- Certainty=0.2338(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 10191> which encodes amino acid sequence <SEQ ID 10192> was also identified.

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

```
>GP:CAA51473 GB:X72967 neuraminidase [Streptococcus pneumoniae]
 Identities = 294/504 (58%), Positives = 380/504 (75%), Gaps = 10/504 (1%)

Query: 303 EDIKSYFQYYCHLNHQLKLPKGAILSAKTEVYRGGDFGRKNKDNVFGYRIPSLLKTEKGT 362
           E+++   Q +   + + KLP+GA L+ KT+++  G  G+ NKD +  YRIP+LLKT+KGT
Sbjct: 299 EEVQKRSQLFKRSDLEKKLPEGAALTEKTDIFESGRNGKPNKDGIKSYRIPALLKTDKGT 358

Query: 363 LLVGADERIEQACDWGNIGMVIRRSEDDGVTWGKRETIVNLRNNPRVPLVTSGDYSGSPI 422
           L+ GADER   + DWG+IGMVIRRSED+G TWG R TI NLR+NP+     S   GSP+
Sbjct: 359 LIAGADERRLHSSDWGDIGMVIRRSEDNGKTWGDRVTITNLRDNPKA----SDPSIGSPV 414

Query: 423 NMDMALVQDTSSKTKRIFSIYDMFPEGRGVISIANTPEKEYTQIGGQSYLNLYNNGKKSK 482
           N+DM LVQD   +TKRIFSIYDMFPEG+G+  +++  E+ Y +I G++Y  LY  G+K
Sbjct: 415 NIDMVLVQDP--ETKRIFSIYDMFPEGKGIFGMSSQKEEAYKKIDGKTYQILYREGEKG- 471

Query: 483 VFTIRDKGIVYNFKGKKTDYHVITETTKSDHSNLGDIYKGKQLLGNIYFTKHKTSPFRLA 542
            +TIR+ G VY   GK TDY V+ +  K  +S+ GD+YKG QLLGNIYFT +KTSPFR+A
Sbjct: 472 AYTIRENGTVYTPDGKATDYRVVVDPVKPAYSDKGDLYKGNQLLGNIYFTTNKTSPFRIA 531

Query: 543 KSSYVWMSYSDDDGRTWSSPRDITASLRQKGMKFLGIGPGKGIVLKWGPHAGRIIIPAYS 602
           K SY+WMSYSDDDG+TWS+P+DIT  ++   MKFLG+GPG GIVL+ GPH GRI+IP Y+
Sbjct: 532 KDSYLWMSYSDDDGKTWSAPQDITPMVKADWMKFLGVGPGTGIVLRNGPHKGRILIPVYT 591

Query: 603 TNWKSHLRGSQSSRLIYSDDHGKTWHTGKAVNDNRILSNGEKIHSLTMDNKKEQNTESVP 662
           TN  SHL GSQSSR+IYSDDHGKTWH G+AVNDNR + +G+KIHS TM+N++ QNTES
Sbjct: 592 TNNVSHLNGSQSSRIIYSDDHGKTWHAGEAVNDNRQV-DGQKIHSSTMNNRRAQNTESTV 650

Query: 663 VQLKNGDIKLFMRNLTGNLEVATSKDGGETWQNHVKRYKEVHDAYVQLSAIRFEHDKKEY 722
           VQL NGD+KLFMR LTG+L+VATSKDGG TW+  +KRY +V D YVQ+SAI   H+ KEY
Sbjct: 651 VQLNNGDVKLFMRGLTGDLQVATSKDGGVTWEKDIKRYPQVKDVYVQMSAIHTMHEGKEY 710
```

```
Query: 723 ILLVNANGPGKKRQDGYARLAQVNRNGSFKWLYHHHIQDGSFAYNSVQQLNNDKFGVLYE 782
          I+L NA GP  KR++G   LA+V  NG   WL H+ IQ G FAYNS+Q+L N ++G+LYE
Sbjct: 711 IILSNAGGP--KRENGMVHLARVEENGELTWLKHNPIQKGEFAYNSLQELGNGEYGILYE 768

Query: 783 HREKHQNSFTLNYKVFNWSFLSQN 806
           H EK QN++TL+++ FNW FLS++
Sbjct: 769 HTEKGQNAYTLSFRKFNWDFLSKD 792
```

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

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

Example 492

A DNA sequence (GBSx0530) was identified in *S.agalactiae* <SEQ ID 1571> which encodes the amino acid sequence <SEQ ID 1572>. This protein is predicted to be unnamed protein product (gatC). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.63   Transmembrane  154 - 170 ( 149 - 178)
   INTEGRAL    Likelihood =-11.99   Transmembrane  103 - 119 (  98 - 123)
   INTEGRAL    Likelihood = -7.91   Transmembrane   21 -  37 (  14 -  40)
   INTEGRAL    Likelihood = -6.53   Transmembrane  448 - 464 ( 444 - 467)
   INTEGRAL    Likelihood = -5.89   Transmembrane   47 -  63 (  45 -  68)
   INTEGRAL    Likelihood = -5.10   Transmembrane  356 - 372 ( 352 - 373)
   INTEGRAL    Likelihood = -4.78   Transmembrane  330 - 346 ( 328 - 350)
   INTEGRAL    Likelihood = -4.41   Transmembrane  376 - 392 ( 375 - 393)
   INTEGRAL    Likelihood = -3.72   Transmembrane  243 - 259 ( 235 - 266)
   INTEGRAL    Likelihood = -2.55   Transmembrane  277 - 293 ( 275 - 293)

----- Final Results -----
               bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1573> which encodes the amino acid sequence <SEQ ID 1574>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.31   Transmembrane  154 - 170 ( 150 - 179)
   INTEGRAL    Likelihood =-11.68   Transmembrane  104 - 120 (  99 - 124)
   INTEGRAL    Likelihood = -9.82   Transmembrane  447 - 463 ( 442 - 469)
   INTEGRAL    Likelihood = -7.91   Transmembrane   22 -  38 (  11 -  41)
   INTEGRAL    Likelihood = -7.11   Transmembrane  377 - 393 ( 375 - 403)
   INTEGRAL    Likelihood = -5.89   Transmembrane   48 -  64 (  46 -  69)
   INTEGRAL    Likelihood = -4.78   Transmembrane  331 - 347 ( 329 - 351)
   INTEGRAL    Likelihood = -3.88   Transmembrane  357 - 373 ( 353 - 373)
   INTEGRAL    Likelihood = -2.55   Transmembrane  278 - 294 ( 276 - 294)
   INTEGRAL    Likelihood = -1.22   Transmembrane  240 - 256 ( 240 - 257)

----- Final Results -----
               bacterial membrane --- Certainty=0.5925(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 = 419/482 (86%), Positives = 447/482 (91%)

Query: 1   MQVFLNIVNKFFDPIIHMGSGVVMLIVMTGLAMIFGVKFSKALEGGIKLAIALTGIGAII 60
           MQ FL+I+NK      I +GSGVVMLIVMTGLAMIFGVKF+KALEGGIKLAIALTGIGAII
```

```
Sbjct: 2   MQPFLDIINKILGFPIQLGSGVVMLIVMTGLAMIFGVKFTKALEGGIKLAIALTGIGAII 61

Query: 61  GILTGAFSESLQAFVKNTGINLSIIDVGWAPLATITWGSPYTLYFLLIMLIVNIVMIVMK 120
           GILTGAFSESLQAFVKNTGI+L+IIDVGWAPLATITWGSPYTLYFLL+ML+VNIVMIVMK
Sbjct: 62  GILTGAFSESLQAFVKNTGISLNIIDVGWAPLATITWGSPYTLYFLLVMLVVNIVMIVMK 121

Query: 121 KTDTLDVDIFDIWHLSITGLLIMWYAKKNNLPTLLSVIIATVAIIFVGVLKIINSDLMKP 180
           KTDTLDVDIFDIWHLSITGLLIMWYA +N+LP  +S++IATVA+I VGVLKIINSDLMKP
Sbjct: 122 KTDTLDVDIFDIWHLSITGLLIMWYAARNHLPVFVSLLIATVAVILVGVLKIINSDLMKP 181

Query: 181 TFDDLLGTGPTSPMTSTHMNYMMNPIIMVLDKLFDKVFPGLDKYDFDAAKLNKAIGFWGS 240
           TFDDLLGTGP SPMTSTHMNYMMNPIIMVLDK+FDKVFPGLDKYDFDAAKLNK IGFWGS
Sbjct: 182 TFDDLLGTGPQSPMTSTHMNYMMNPIIMVLDKIFDKVFPGLDKYDFDAAKLNKKIGFWGS 241

Query: 241 KFFIGMILGLVIGIMGNPVFSFAALGGWFSLGFTAGACLELFSLIGSWFIAAVEPLSQGI 300
           KFFIGM LG VIGIMG+P F+  ++  WF LGFTAGACLELFSLIGSWFIAAVEPLSQGI
Sbjct: 242 KFFIGMALGFVIGIMGDPHFTVESIKNWFGLGFTAGACLELFSLIGSWFIAAVEPLSQGI 301

Query: 301 TNFANGKMHGRRFNIGLDWPFIAGRAEIWACANILAPIMLVEAILLSKVGNGILPLAGII 360
           TNFAN +MHGRRFNIGLDWPFIAGRAEIWACANILAPIML+EA+LLSKVGNGILPLAGII
Sbjct: 302 TNFANARMHGRRFNIGLDWPFIAGRAEIWACANILAPIMLIEAVLLSKVGNGILPLAGII 361

Query: 361 AMGVTPALLVVTRGRLIRMITFGTLLLPLFLLSGTMIAPFATELAKKVGAFPAGARAGSL 420
           AMG+TPALLVVTRGRLIRMI FG+LLLPLFLLSGTMIAPFATELAKKVGAFPAG  AGSL
Sbjct: 362 AMGMTPALLVVTRGRLIRMIIFGSLLLPLFLLSGTMIAPFATELAKKVGAFPAGTSAGSL 421

Query: 421 ITHSTLEGPMEKIFGYVIGKATTGQLSAIITLIIFATAYLGLFMWYAKQMKRRNAEYAAN 480
           ITHSTLEGPMEKIFGYVIG+ATTGQ+++IITLIIF   YL LF WYA QMK RNAEYA
Sbjct: 422 ITHSTLEGPMEKIFGYVIGQATTGQIASIITLIIFVAIYLSLFAWYANQMKARNAEYAKT 481

Query: 481 QK 482
            K
Sbjct: 482 MK 483
```

A related GBS gene <SEQ ID 8579> and protein <SEQ ID 8580> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:      4.31
GvH: Signal Score (-7.5): -2.64
     Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 6 value: -12.63 threshold:  0.0
   INTEGRAL    Likelihood =-12.63   Transmembrane  154 - 170 ( 149 - 178)
   INTEGRAL    Likelihood =-11.99   Transmembrane  103 - 119 (  98 - 123)
   INTEGRAL    Likelihood = -7.91   Transmembrane   21 -  37 (  14 -  40)
   INTEGRAL    Likelihood = -5.89   Transmembrane   47 -  63 (  45 -  68)
   INTEGRAL    Likelihood = -4.88   Transmembrane  243 - 259 ( 235 - 265)
   INTEGRAL    Likelihood = -1.22   Transmembrane  268 - 284 ( 268 - 284)
   PERIPHERAL  Likelihood =  0.85     127
modified ALOM score:   3.03

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
ORF00838(343 - 1122 of 1455)
EGAD|91348|EC2092(9  -  344  of  451)  PTS  system,  galactitol  specific  IIC  component
(Escherichia coli) OMNI|NT01EC2494 PTS system galactitol-specific enzyme IIC component
SP|P37189|PTKC_ECOLI PTS SYSTEM, GALACTITOL-SPECIFIC IIC COMPONENT (EIIC-GAT) (GALACTICOL-
PERMEASE  IIC  COMPONENT)  (PHOSPHOTRANSFERASE  ENZYME  II,  C  COMPONENT).
GP|1736809|dbj|BAA15955.1||D90847 PTS system, Galactitol-specific IIC component (EIIC-GAT)
```

```
(Galactitol- permease IIC component) (Phosphotransferase enzyme II, C component).
{Escherichia coli} GP|17884
%Match = 10.9
%Identity = 29.8  %Similarity = 59.2
Matches = 68  Mismatches = 88  Conservative Sub.s = 67

282       312       342       372       402       432       462       492
LS*HI*NWN*S*RRRNMQVFLNIVNKFFDPIIHMGSGVVMLIVMTGLAMIFGVKFSKALEGGIKLAIALTGIGAIIGILT
                                |: :|  |:: ||:  :: |:|:|    :: |: : | : ||| :||::
                          MFSEVMRYILDLGPTVMLPIVIIIFSKILGMKAGDCFKAGLHIGIGFVGIGLVIGLML
                                10        20        30        40        50

522       552       582       612       642       672       702
GAFSESLQAFVKNTGINLSIIDVGWAPLATITWGSPYTLYFLLIMLIVNIVMIVMKKTDTLDVDIFDIWHLSITGLLIM-
 :   : :|  :|  :|| ::|||||   : :|| |   |  : | ::||: |:: : |  ::|||::|||:: || |:
DSIGPAAKAMAENFDLNLHVVDVGWPGSSPMTWASQIALVAIPIAILVNVAMLLTRMTRVVNVDIWNIWHMTFTGALLHL
          70        80        90       100       110       120       130

747       774       804       834       864       894
----------------------WYAKKN-NLPTLLSVIIATVAIIFVGVLKIINSDLMKPTFDDLLGTGPTSPMTSTH
                       |:|:   |:  |  : |      ::|
ATGSWMIGMAGVVIHAAFVYKLGDWFARDTRNFFELEGIAIPHGTSAYMG----------------------------
         150       160       170       180

924       954       984       1014      1044
MNYMMNPIIMVLDKLFDKVFPGLDKYDFDAAKLNKAIGFWGSKFFIGMILGLVIXIM----------------------~~~
     || :::| : :|: ||:::  | |  : :  | :|    :| ::||:| |:
------PIAVLVDAIIEKI-PGVNRIKFSADDIQRKFGPFGEPVTVGFVMGLIIGILAGYDVKGVLQLAVKTAAVML~~~
               200       210       220       230       240       250

1092      1122      1152      1182      1212      1242
~--------------------GNPVFSFASIRWLVFFFVLQQGACLEVGLF*LVSWVQLLQ*NHFLRKLLILLMVNAXX*
                     || |: |  :  : ||  :
-VVSASLIFIPLTILIAVCVPGNQVLPFGDLATIGFFVAMAVAVHRGNLFRTLISGVIIMSITLWIATQTIGLHTQLAAN
             320       330       340       350       360       370       380
```

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

Example 493

A DNA sequence (GBSx0531) was identified in *S.agalactiae* <SEQ ID 1575> which encodes the amino acid sequence <SEQ ID 1576>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0302(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 1577> which encodes the amino acid sequence <SEQ ID 1578>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0302(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 = 85/100 (85%), Positives = 96/100 (96%)

Query: 1   MIKILAACGAGVNSSHQIKDAIETQLGDRGYNVHCDAVMVKDITEEMVNKYDIFTPIAKT 60
           MIKILAACGAGVNSSHQIKDAIETQ+ DRGY+VHCDAVMVKDITEE+V++YDIFTPIAKT
Sbjct: 1   MIKILAACGAGVNSSHQIKDAIETQMSDRGYDVHCDAVMVKDITEELVSRYDIFTPIAKT 60

Query: 61  DLGFNVPIPVVEAGPILYRIPVMSEPVFTALEQVIKEHNL 100
           DLGF +PIP+VEAGPILYRIP+MSEPVF  LE+VIKE++L
Sbjct: 61  DLGFEMPIPIVEAGPILYRIPIMSEPVFAELERVIKENHL 100
```

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

Example 494

A DNA sequence (GBSx0532) was identified in *S.agalactiae* <SEQ ID 1579> which encodes the amino acid sequence <SEQ ID 1580>. This protein is predicted to be GatA. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2078(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 10193> which encodes amino acid sequence <SEQ ID 10194> was also identified.

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

```
>GP:AAG09977 GB:AF248038 GatA [Streptococcus agalactiae]
 Identities = 39/135 (28%), Positives = 76/135 (55%), Gaps = 9/135 (6%)

Query: 16  QEELFDLVSKALIKQHYVSPNYRQAVKERERERFPTGLKIDLKDGTPIQYVAIPHTETQYC 75
           Q  L +++S+ L+++ YV   + +A+ +RE+++PTGL+++          VAIPHT ++Y
Sbjct: 20  QTNLLEVLSQYLLQKGYVKTEFSKAILQREKDYPTGLQLE------NMAVAIPHTYSEYV 73

Query: 76  LVDRIFYVKNSQPITFKHMINPEEECRVQDFFFIINSRN-SNQSDILSNLITFFITKGNL 134
           L   I+  K  +PI+F  M   E+E  +  +  ++   N  +Q+ +L+ L+T F     +
Sbjct: 74  LKPFIYINKLKEPISFIQM-GTEDEIVMARYVIVLGISNPKDQAGLLAELMTLFSNPKIV 132

Query: 135 DRLHELGDNKEKINH 149
            +L E+   KE + +
Sbjct: 133 QQL-EMAQTKEALKN 146
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1581> which encodes the amino acid sequence <SEQ ID 1582>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3130(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 = 102/154 (66%), Positives = 122/154 (78%)

Query: 4   VTQDILFIDAHSQEELFDLVSKALIKQHYVSPNYRQAVKERERERFPTGLKIDLKDGTPIQ 63
```

```
           V  +ILF +A +Q ELFDLV+  L K  YV+  Y QA+ ERE  FPTGLK+DLKDG+ I
Sbjct: 1   VFPNILFTEARTQPELFDLVASHLEKVGYVTQEYHQALVEREAVFPTGLKVDLKDGSDIL 60

Query: 64  YVAIPHTETQYCLVDRIFYVKNSQPITFKHMINPEEECRVQDFFFIINSRNSNQSDILSN 123
           Y AIPHTET+YCLVD++ YV+NSQ +TFKHMINPEE+C V DFFFIINS+N  Q+ ILSN
Sbjct: 61  YAAIPHTETKYCLVDQVVYVRNSQALTFKHMINPEEDCLVTDFFFIINSQNEGQTTILSN 120

Query: 124 LITFFITKGNLDRLHELGDNKEKINHYLIEKGVF 157
           LITFFITKGNL  L  L D+K+ I++YLIEKGVF
Sbjct: 121 LITFFITKGNLSYLASLKDDKQAISNYLIEKGVF 154
```

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

Example 495

A DNA sequence (GBSx0533) was identified in *S.agalactiae* <SEQ ID 1583> which encodes the amino acid sequence <SEQ ID 1584>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1429(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:AAA25176 GB:M60447 repressor protein [Lactococcus lactis]
 Identities = 139/255 (54%), Positives = 189/255 (73%), Gaps = 6/255 (2%)

Query: 1   MLKRERLQKIIEKVNINGIVTVNEIMEELDVSDMTVRRDLDELDKAGLLIRIHGGAQKVN 60
           M K+ RL+KI++ + I+G +T+ EI++ELD+SDMT RRDLD L+  GLL R HGGAQ ++
Sbjct: 7   MNKKRRLEKILDMLKIDGTITIKEIIDELDISDMTARRDLDALEADGLLTRTHGGAQLLS 66

Query: 61  ASPTPQNYEKSNTEKYDIQTNEKLEIAQFAKQFINDGETIFIGPGTTLEKLATQLLD--- 117
           +    + EK++ EK  + T EK++IA+ A   I DG+TIFIGPGTTL +LA +L
Sbjct: 67  SK---KPLEKTHIEKKSLNTKEKIDIAKKACSLIKDGDTIFIGPGTTLVQLALELKGRKG 123

Query: 118 FKIRVVTNSLPVFNILNQSSTLDLILVGGEYREITGAFVGSVTINSIKSLNFSKAFVSSN 177
           +KIRV+TNSLPVF ILN S T+DL+L+GGEYREITGAFVGS+   ++K++ F+KAFV +N
Sbjct: 124 YKIRVITNSLPVFLILNDSETIDLLLLGGEYREITGAFVGSMASTNLKAMRFAKAFVRAN 183

Query: 178 GVFEKSIATYDEGEGEIQRIALNNSFEKFLLVDSQKFGKYDFYTFYQLDDIDFVLTDHNI 237
            V   SIATY + EG IQ++ALNN+ EKFLLVDS KF +YDF+ FY LD +D ++TD+ I
Sbjct: 184 AVTHNSIATYSDKEGVIQQLALNNAVEKFLLVDSTKFDRYDFFNFYNLDQLDTIITDNQI 243

Query: 238 DNVVKEQYSSFTKIL 252
                E++S +T IL
Sbjct: 244 SPQHLEEFSQYTTIL 258
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1585> which encodes the amino acid sequence <SEQ ID 1586>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0740(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 = 161/252 (63%), Positives = 195/252 (76%), Gaps = 3/252 (1%)

Query: 1   MLKRERLQKIIEKVNINGIVTVNEIMEELDVSDMTVRRDLDELDKAGLLIRIHGGAQKVN 60
           MLKRERL KI E VN  GIVTVN+I++ L+VSDMTVRRDLDEL+KAG LIRIHGGAQ +
Sbjct: 1   MLKRERLLKITEIVNEQGIVTVNDIIQTLNVSDMTVRRDLDELEKAGKLIRIHGGAQSIT 60

Query: 61  ASPTPQNYEKSNTEKYDIQTNEKLEIAQFAKQFINDGETIFIGPGTTLEKLATQLLDFKI 120
               P   E+SN EK  +QT EK E+A +A Q +NDGETIFIGPGTTLE  A QL + +I
Sbjct: 61  M---PNKKERSNIEKQTVQTKEKWELASYATQLVNDGETIFIGPGTTLECFAEQLKNRQI 117

Query: 121 RVVTNSLPVFNILNQSSTLDLILVGGEYREITGAFVGSVTINSIKSLNFSKAFVSSNGVF 180
           R+VTNSLPVFNIL  S T+DLIL+GGEYR ITGAFVGS+   +I SL F+KAF+S NG++
Sbjct: 118 RIVTNSLPVFNILQDSETIDLILIGGEYRSITGAFVGSLASQNISSLKFAKAFISCNGIY 177

Query: 181 EKSIATYDEGEGEIQRIALNNSFEKFLLVDSQKFGKYDFYTFYQLDDIDFVLTDHNIDNV 240
           +  IATY E EGEIQ++A NNS EK+LLVD+QKF  YDF+ FY L++ID V+TD  I
Sbjct: 178 KNDIATYSETEGEIQKLAFNNSIEKYLLVDNQKFNAYDFFIFYHLNNIDAVVTDSQITED 237

Query: 241 VKEQYSSFTKIL 252
           V E+YS FT++L
Sbjct: 238 VIERYSQFTQLL 249
```

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

Example 496

A DNA sequence (GBSx0534) was identified in *S.agalactiae* <SEQ ID 1587> which encodes the amino acid sequence <SEQ ID 1588>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3436(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:AAD13797 GB:AF062533 unknown [Streptococcus agalactiae]
 Identities = 86/371 (23%), Positives = 136/371 (36%), Gaps = 79/371 (21%)

Query: 11  DLSESELKAAQEFLSGKSEANQDKPKTGKTAQEIYEAIEPKAIVKPEDLLFGIAQATDYK 70
           DL++      +  L  K     D   TG         IEP+  V    L      AT
Sbjct: 526 DLTQIAFAEQELMLKDKKHYRYDIVDTG---------IEPRLAVDVSSLPMHAGNATYDT 576

Query: 71  NGTFVIPHKDHYHYVELKWFDEEKDLLADSDKTYSLEDYLATAKYYMMHPEKRPKVEGWG 130
             +FVIPH DH H V   W    +               +AT KY M HPE RP V  W
Sbjct: 577 GSSFVIPHIDHIHVVPYSWLTRNQ---------------IATIKYVMQHPEVRPDV--WS 619

Query: 131 KDAEIYKEKDSNKADKPSPAPTDNKSTSNSSDKNLSAAEVFKQAKPEKIVPLDKIAAHMA 190
           K        + + +  P+  P D ++   +     SA EV      +K +   + AA
Sbjct: 620 KPGH-----EESGSVIPNVTPLDKRAGMPNWQIIHSAEEV------QKALAEGRFAA--- 665

Query: 191 YAVGFEDDQLIVPHHDHYHNVPMAWFDKGGLWKAPEGYTLQQLFST--IKYYMEHPNELP 248
                D  I    D       W D       +G +L+ +  +   +    +   EL
Sbjct: 666 ------PDGYIFDPRDVLAKETFVWKDGSFSIPRADGSSLRTINKSDLSQAEWQQAQELL 719

Query: 249 KEKGWGHDSDHNKGSNKDNKAKNYAPDEEPEDSGKVTHNYGFYDVNKGSDEEEP-EKQED 307
            +K  G  +D +K             P+E+ +                +K ++ ++P E  ++
Sbjct: 720 AKKNAGDATDTDK-----------PEEKQQ-------------ADKSNENQQPSEASKE 754

Query: 308 ESELDEYELGMAQNAKKYGMDRQSFEKQLIQLSNKYSVSFESFNYINGSQVQVTKKDGSK 367
           E E D++      +   YG+DR + E  + QL+ K ++    +       VQ   K+G
Sbjct: 755 EKESDDF----IDSLPDYGLDRATLEDHINQLAQKANID-PKYLIFQPEGVQFYNKNGEL 809
```

```
Query: 368 VLVDIKTLTEV 378
          V  DIKTL ++
Sbjct: 810 VTYDIKTLQQI 820
```

A related DNA sequence was identified in *S.agalactiae* <SEQ ID 6983> which encodes the amino acid sequence <SEQ ID 6984>. 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>
```

A related GBS gene <SEQ ID 8581> and protein <SEQ ID 8582> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 2
McG: Discrim Score:      6.06
GvH: Signal Score (-7.5): -5.61
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 0 value:   2.23 threshold:  0.0
   PERIPHERAL  Likelihood =  2.23       6
 modified ALOM score:  -0.95

*** Reasoning Step: 3

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

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1589> which encodes the amino acid sequence <SEQ ID 1590>. 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>
```

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

```
 Identities = 808/825 (97%), Positives = 816/825 (97%), Gaps = 3/825 (0%)

Query: 2    KKTYGYIGSVAAILLATHIGSYQLGKHHMGLATKDNQIAYIDDSKGKVKAPKTNKTMDQ 60
            KKTYGYIGSVAAILLATHIGSYQLGKHHMG ATKDNQIAYIDDSKGK KAPKTNKTMDQ
Sbjct: 2    KKTYGYIGSVAAILLATHIGSYQLGKHHMGSATKDNQIAYIDDSKGKAKAPKTNKTMDQ 60

Query: 61   ISAEEGISAEQIVVKITDQGYVTSHGDHYHFYNGKVPYDAIISEELLMTDPNYHFKQSDV 120
            ISAEEGISAEQIVVKITDQGYVTSHGDHYHFYNGKVPYDAIISEELLMTDPNY FKQSDV
Sbjct: 61   ISAEEGISAEQIVVKITDQGYVTSHGDHYHFYNGKVPYDAIISEELLMTDPNYRFKQSDV 120

Query: 121  INEILDGYVIKVNGNYYVYLKPGSKRKNIRTKQQIAEQVAKGTKEAKEKGLAQVAHLSKE 180
            INEILDGYVIKVNGNYYVYLKPGSKRKNIRTKQQIAEQVAKGTKEAKEKGLAQVAHLSKE
Sbjct: 121  INEILDGYVIKVNGNYYVYLKPGSKRKNIRTKQQIAEQVAKGTKEAKEKGLAQVAHLSKE 180

Query: 181  EVAAVNEAKRQGRYTTDDGYIFSPTDIIDDLGDAYLVPHGNHYHYIPKKDLSPSELAAAQ 240
            EVAAVNEAKRQGRYTTDDGYIFSPTDIIDDLGDAYLVPHGNHYHYIPKKDLSPSELAAAQ
Sbjct: 181  EVAAVNEAKRQGRYTTDDGYIFSPTDIIDDLGDAYLVPHGNHYHYIPKKDLSPSELAAAQ 240
```

```
Query: 241 AYWSQKQGRGARPSDYRPTPAP--GRRKAPIPDVTPNPGQGHQPDNGGYHPAPPRPNDAS 298
           AYWSQKQGRGARPSDYRPTPAP  GRRKAPIPDVTPNPGQGHQPDNGGYHPAPPRPNDAS
Sbjct: 241 AYWSQKQGRGARPSDYRPTPAPAPGRRKAPIPDVTPNPGQGHQPDNGGYHPAPPRPNDAS 300

Query: 299 QNKHQRDEFKGKTFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKSNAFGYVVPHGDHYH 358
           QNKHQRDEFKGKTFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKSNAFGYVVPHGDHYH
Sbjct: 301 QNKHQRDEFKGKTFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKSNAFGYVVPHGDHYH 360

Query: 359 IIPRSQLSPLEMELADRYLAGQTDDNDSGSDHSKPSDKEVTHTFLGHRIKAYGKGLDGKP 418
           IIPRSQLSPLEMELADRYLAGQT+D+DSGSDHSKPSDKEVTHTFLGHRIKAYGKGLDGKP
Sbjct: 361 IIPRSQLSPLEMELADRYLAGQTEDDDSGSDHSKPSDKEVTHTFLGHRIKAYGKGLDGKP 420

Query: 419 YDTSDAYVFSKESIHSVDKSGVTAKHGDHFHYIGFGELEQYELDEVANWVKAKGQADELV 478
           YDTSDAYVFSKESIHSVDKSGVTAKHGDHFHYIGFGELEQYELDEVANWVKAKGQADEL
Sbjct: 421 YDTSDAYVFSKESIHSVDKSGVTAKHGDHFHYIGFGELEQYELDEVANWVKAKGQADELA 480

Query: 479 AALDQEQGKEKPLFDTKKVSRKVTKDGKVGYIMPKDGKDYFYARYQLDLTQIAFAEQELM 538
           AALDQEQGKEKPLFDTKKVSRKVTKDGKVGY+MPKDGKDYFYAR QLDLTQIAFAEQELM
Sbjct: 481 AALDQEQGKEKPLFDTKKVSRKVTKDGKVGYMMPKDGKDYFYARDQLDLTQIAFAEQELM 540

Query: 539 LKDKKHYRYDIVDTGIEPRLAVDLSSLPMHAGNATYDTGSSFVIPHIDHIHVVPYSWLTR 598
           LKDKKHYRYDIVDTGIEPRLAVD+SSLPMHAGNATYDTGSSFVIPHIDHIHVVPYSWLTR
Sbjct: 541 LKDKKHYRYDIVDTGIEPRLAVDVSSLPMHAGNATYDTGSSFVIPHIDHIHVVPYSWLTR 600

Query: 599 NQIATIKYVMQHPEVRPDVWSKPGHEESGSVIPNVTPLDKRAGMPNWQIIHSAEEVQKAL 658
           +QIATIKYVMQHPEVRPD+WSKPGHEESGSVIPNVTPLDKRAGMPNWQIIHSAEEVQKAL
Sbjct: 601 DQIATIKYVMQHPEVRPDIWSKPGHEESGSVIPNVTPLDKRAGMPNWQIIHSAEEVQKAL 660

Query: 659 AEGRFAAPDGYIFDPRDVLAKETFVWKDGSFSIPRADGSSLRTINKSDLSQAEWQQAQEL 718
           AEGRFA PDGYIFDPRDVLAKETFVWKDGSFSIPRADGSSLRTINKSDLSQAEWQQAQEL
Sbjct: 661 AEGRFATPDGYIFDPRDVLAKETFVWKDGSFSIPRADGSSLRTINKSDLSQAEWQQAQEL 720

Query: 719 LAKKNAGDATDTDKPEEKQQADKSNENQQPSEASK-EEKESDDFIDSLPDYGLDRATLED 777
           LAKKNAGDATDTDKP+EKQQADKSNENQQPSEASK EEKESDDFIDSLPDYGLDRATLED
Sbjct: 721 LAKKNAGDATDTDKPKEKQQADKSNENQQPSEASKEEEKESDDFIDSLPDYGLDRATLED 780

Query: 778 HINQLAQKANIDPKYLIFQPEGVQFYNKNGELVTYDIKTLQQINP 822
           HINQLAQKANIDPKYLIFQPEGVQFYNKNGELVTYDIKTLQQINP
Sbjct: 781 HINQLAQKANIDPKYLIFQPEGVQFYNKNGELVTYDIKTLQQINP 825
```

SEQ ID 8582 was expressed in *E.coli* in two different forms. GBS293dNterm was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 14; MW 74kDa + lanes 17 & 18; MW 48.8kDa). GBS293C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figures 148 (lane 2-4; MW 71kDa + lanes 5 & 7; MW 46kDa) and 182 (lane 7; MW 46kDa). Purified GBS293C-His is shown in Figure 241, lanes 8& 9.

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

Example 497

A DNA sequence (GBSx0535) was identified in *S.agalactiae* <SEQ ID 1591> which encodes the amino acid sequence <SEQ ID 1592>. 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>
```

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

```
>GP:AAD13797 GB:AF062533 unknown [Streptococcus agalactiae]
 Identities = 213/463 (46%), Positives = 277/463 (59%), Gaps = 41/463 (8%)

Query: 4   KKTV-IISALSVALFGTGVGAYQLGSYNA--QKSDNSVSYVKTDKSDSKAQATAVNKTPD 60
           KKT   I +++  L  T +G+YQLG ++       DN ++Y+  D S  K +A   NKT D
Sbjct: 2   KKTYGYIGSVAAILLATHIGSYQLGKHHMGLATKDNQIAYI--DDSKGKVKAPKTNKTMD 59

Query: 61  QISKEEGISAEQIVVKITDDGYVTSHGDHYHYYNGKVPYDAIISEELIMKDPSYVFNKAD 120
           QIS EEGISAEQIVVKITD GYVTSHGDHYH+YNGKVPYDAIISEEL+M DP+Y F ++D
Sbjct: 60  QISAEEGISAEQIVVKITDQGYVTSHGDHYHFYNGKVPYDAIISEELLMTDPNYHFKQSD 119

Query: 121 VINEVKDGYIIKVNGKYYLYLKEGSKRTNVRTKEQIQKQREEWSKGGSKGESGKHSSAKT 180
           VINE+ DGY+IKVNG YY+YLK GSKR N+RTK+QI +Q  + +K     E+ +   A+
Sbjct: 120 VINEILDGYVIKVNGNYYVYLKPGSKRKNIRTKQQIAEQVAKGTK-----EAKEKGLAQV 174

Query: 181 QALS----ASVREAKASGRYTTDDGYVFSPTDVIDDMGDAFLVPHGDHFHYIPKADLSPS 236
             LS    A+V EAK  GRYTTDDGY+FSPTD+IDD+GDA+LVPHG+H+HYIPK DLSPS
Sbjct: 175 AHLSKEEVAAVNEAKRQGRYTTDDGYIFSPTDIIDDLGDAYLVPHGNHYHYIPKKDLSPS 234

Query: 237 ELSAAQAYWNRKTGRSGNSS--KPSNSSSYIHASAPSGNVSTGRHANAPISIPRVTHANH 294
           EL+AAQAYW++K GR     S  +P+ +     A  P    + G+              H
Sbjct: 235 ELAAAQAYWSQKQGRGARPSDYRPTPAPGRRKAPIPDVTPNPGQGHQPD------NGGYH 288

Query: 295 WSKPAGNHATAPKHHAPTTKPINKDSALDKMLKRLYAQPLYARHVESDGLVYDPAQVNAF 354
            + P  N A+  KH      +    K      ++L +L++   L  RHVE DGL+++P QV
Sbjct: 289 PAPPRPNDASQNKHQ----RDEFKGKTFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKS 344

Query: 355 TAIGVSIPHGNHFHFIHYKDMSPLELE-ATRMVAEHRGHHIDALGKKDSTEKPKHISHEP 413
            A G  +PHG+H+H I     +SPLE+E A R +A          G+ D  +     S
Sbjct: 345 NAFGYVVPHGDHYHIIPRSQLSPLEMELADRYLA----------GQTDDNDSGSDHSKPS 394

Query: 414 NKE-PHTEEEHHAVTPKDQRKGKP---NSQIVYSAQEIEEAKK 452
           +KE  HT   H          GKP   +   V+S + I    K
Sbjct: 395 DKEVTHTFLGHRIKAYGKGLDGKPYDTSDAYVFSKESIHSVDK 437
```

There is also homology to SEQ ID 1590.

SEQ ID 1592 (GBS94) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 3; MW 52.5kDa).

GBS94-His was purified as shown in Figure 194, lane 8.

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

Example 498

A DNA sequence (GBSx0536) was identified in *S.agalactiae* <SEQ ID 1593> which encodes the amino acid sequence <SEQ ID 1594>. This protein is predicted to be Lmb. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> May be a lipoprotein

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

There is also homology to SEQ IDs 1596 and 5548.

A related GBS gene <SEQ ID 8583> and protein <SEQ ID 8584> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: 22   Crend: 5
McG: Discrim Score:     13.64
GvH: Signal Score (-7.5): -5.75
     Possible site: 24
>>> May be a lipoprotein
ALOM program   count: 0 value:   4.83 threshold:  0.0
   PERIPHERAL  Likelihood =  4.83      33
 modified ALOM score:  -1.47

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

SEQ ID 8584 (GBS22) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 6; MW 35kDa).

The GBS22-His fusion product was purified (Figure 94A; see also Figure 193, lane 4) and used to immunise mice (lane 2 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 94B), FACS (Figure 94C), 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.

SEQ ID 8584 (GBS22) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 183 (lane 7 & 8; MW 35kDa).

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

Example 499

A DNA sequence (GBSx0537) was identified in *S.agalactiae* <SEQ ID 1597> which encodes the amino acid sequence <SEQ ID 1598>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.59   Transmembrane   19 -  35 (  19 -  35)

----- 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 GENPEPT database:

```
>GP:CAA51352 GB:X72832 ORF1 [Streptococcus equisimilis]
 Identities = 104/145 (71%), Positives = 126/145 (86%)

Query: 1   MKIIIQRVNQASVSIEDDVVGSIEKGLVLLVGIAPEDTTEDIAYAVRKITSMRIFSDDEG 60
           MK+++QRV +ASVSI+  + G+I +GL+LLVG+ P+D  ED+AYAVRKI +MRIFSD +G
Sbjct: 1   MKLVLQRVKEASVSIDGKIAGAINQGLLLLVGVGPDDAAEDLAYAVRKIVNMRIFSDADG 60

Query: 61  KMNLSIQDIKGSVLSISQFTLFADTKKGNRPAFTGAADPVKANQFYDIFNQELANHVSVE 120
           KMN SIQDIKGS+LS+SQFTL+ADTKKGNRPAFTGAA P  A+QFYD FN++LA+ V VE
Sbjct: 61  KMNQSIQDIKGSILSVSQFTLYADTKKGNRPAFTGAAKPDMASQFYDRFNEQLADFVPVE 120

Query: 121 TGQFGADMQVSLINDGPVTIVLDTK 145
            G FGADMQVSLINDGPVTI+LDTK
Sbjct: 121 RGVFGADMQVSLINDGPVTIILDTK 145
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1599> which encodes the amino acid sequence <SEQ ID 1600>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1430(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 = 103/145 (71%), Positives = 124/145 (85%)

Query: 1   MKIIIQRVNQASVSIEDDVVGSIEKGLVLLVGIAPEDTTEDIAYAVRKITSMRIFSDDEG 60
           MK+++QRV +ASVSI+  + G+I +GL+LLVG+ P+D  ED+AYAVRKI +MRIFSD +G
Sbjct: 1   MKLVLQRVKEASVSIDGKIAGAINQGLLLLVGVGPDDNAEDLAYAVRKIVNMRIFSDADG 60

Query: 61  KMNLSIQDIKGSVLSISQFTLFADTKKGNRPAFTGAADPVKANQFYDIFNQELANHVSVE 120
           KMN SIQDIKGS+LS+SQFTL+ADTKKGNRPAFTGAA P  A+Q YD FN++LA  V VE
Sbjct: 61  KMNQSIQDIKGSILSVSQFTLYADTKKGNRPAFTGAAKPDLASQLYDSFNEQLAEFVPVE 120

Query: 121 TGQFGADMQVSLINDGPVTIVLDTK 145
            G FGADMQVSLINDGPVTI+LDTK
Sbjct: 121 RGVFGADMQVSLINDGPVTIILDTK 145
```

SEQ ID 1598 (GBS368) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 4; MW 20kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 70 (lane 4; MW 45kDa).

GBS368-GST was purified as shown in Figure 215, lane 6.

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

Example 500

A DNA sequence (GBSx0538) was identified in *S.agalactiae* <SEQ ID 1601> which encodes the amino acid sequence <SEQ ID 1602>. This protein is predicted to be stringent response-like protein (rel) (relA). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane   60 -  76 (  60 -  76)

----- 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:CAA51353 GB:X72832 stringent response-like protein
          [Streptococcus equisimilis]
 Identities = 647/739 (87%), Positives = 696/739 (93%), Gaps = 1/739 (0%)

Query: 1   MVKEINLTGEEVVAITSQYMSETDVAFVKFALNYATAAHYYQARKSGEPYIIHPIQVAGI 60
           M KEINLTGEEVVA+ ++YM+ETD AFVK AL+YATAAH+YQ RKSGEPYI+HPIQVAGI
Sbjct: 1   MAKEINLTGEEVVALAAKYMNETDAAFVKKALDYATAAHFYQVRKSGEPYIVHPIQVAGI 60
```

```
Query: 61  LADLHLDAVTVACGFLHDVVEDTEITLDEIETDFGKDVRDIIDGVTKLGKVEYKSHEEQL 120
           LADLHLDAVTVACGFLHDVVEDT+ITLD IE DFGKDVRDI+DGVTKLGKVEYKSHEEQL
Sbjct: 61  LADLHLDAVTVACGFLHDVVEDTDITLDNIEFDFGKDVRDIVDGVTKLGKVEYKSHEEQL 120

Query: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180
           AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI
Sbjct: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180

Query: 181 SRIKWELEDLSFRYLNETEFYKISHMMSEKRRREEELVDIIVDKIRSYTEEQGLYGDIYG 240
           SRIKWELEDL+FRYLNETEFYKISHMM+EKRRERE LVD IV KI+SYT EQGL+GD+YG
Sbjct: 181 SRIKWELEDLAFRYLNETEFYKISHMMNEKRRERERALVDDIVTKIKSYTTEQGLFGDVYG 240

Query: 241 RPKHIYSIYRKMRDKKKRFDQIYDLIAIRCIMETASDVYAMVGYIHELWRPMPGRFKDYI 300
           RPKHIYSIYRKMRDKKKRFDQI+DLIAIRC+MET SDVYAMVGYIHELWRPMPGRFKDYI
Sbjct: 241 RPKHIYSIYRKMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 300

Query: 301 AAPKANGYQSIHTTVYGPKGPIEIQIRTKEMHQVAEFGVAAHWAYKKGITSKVNQAEQSV 360
           AAPKANGYQSIHTTVYGPKGPIEIQIRTKEMHQVAE+GVAAHWAYKKG+  KVNQAEQ V
Sbjct: 301 AAPKANGYQSIHTTVYGPKGPIEIQIRTKEMHQVAEYGVAAHWAYKKGVRGKVNQAEQKV 360

Query: 361 GMGWIQELVELQDESK-DAKDFVDSVKEDIFTERIYVFTPNGAVQELPRESGPIDFAYAI 419
           GM WI+ELVELQD S  DA DFVDSVKEDIF+ERIYVFTP GAVQELP++SGPIDFAYAI
Sbjct: 361 GMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKDSGPIDFAYAI 420

Query: 420 HTQVGEKATGAKVNGRMVPLTAKLKTGDVVEIITNPNSFGPSRDWIKIVKTNKARNKIRQ 479
           HTQVGEKA GAKVNGRMVPLTAKLKTGDVVEI+TNPNSFGPSRDWIK+VKTNKARNKIRQ
Sbjct: 421 HTQVGEKAIGAKVNGRMVPLTAKLKTGDVVEIVTNPNSFGPSRDWIKLVKTNKARNKIRQ 480

Query: 480 FFKNQDKETSINKGRELLVDYFQEQGYVPNKYLDKKHIEEILPRVSVKSEEALYAAVGFG 539
           FFKNQDKE S+NKGR++LV YFQEQGYV NKYLDKK IE ILP+VSVKSEE+LYAAVGFG
Sbjct: 481 FFKNQDKELSVNKGRDMLVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG 540

Query: 540 DLSPISIFNKLTEKERREEERAKAKAEADELINGGEIKTDKRDVLKVKSENGVIIQGASG 599
           D+SP+S+FNKLTEKERREEERAKAKAEA+EL+NGGEIK + +DVLKV+SENGVIIQGASG
Sbjct: 541 DISPVSVFNKLTEKERREEERAKAKAEAEELVNGGEIKHENKDVLKVRSENGVIIQGASG 600

Query: 600 LLMRIAKCCNPVPGDLIEGYITKGRGVAIHRSDCQNLKSQENYEQRLIDVEWDDDGSKKE 659
           LLMRIAKCCNPVPGD IEGYITKGRG+AIHR+DC N+KSQ+ Y++RLI+VEWD D S K+
Sbjct: 601 LLMRIAKCCNPVPGDPIEGYITKGRGIAIHRADCNNIKSQDGYQERLIEVEWDLDNSSKD 660

Query: 660 YMAEIDIYGLNRSGLLNDVLQTLSNATKLVSTVNAQPTKDMKFANIHVSFGISNLAQLTT 719
           Y AEIDIYGLNR GLLNDVLQ LSN+TK +STVNAQPTKDMKFANIHVSFGI NL  LTT
Sbjct: 661 YQAEIDIYGLNRRGLLNDVLQILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT 720

Query: 720 VVDKIKIIPDVYSVKRTNG 738
           VV+KIK +PDVYSVKRTNG
Sbjct: 721 VVEKIKAVPDVYSVKRTNG 739
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1603> which encodes the amino acid sequence <SEQ ID 1604>. Analysis of this protein sequence reveals the following:

```
     Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane   64 -  80 (  64 -  80)

----- 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 databases:

```
>GP:CAA51353 GB:X72832 stringent response-like protein
          [Streptococcus equisimilis]
 Identities = 700/739 (94%), Positives = 721/739 (96%)

Query: 5   MAKIMNVTGEEVIALAATYMTKADVAFVAKALAYATAAHFYQVRKSGEPYIVHPIQVAGI 64
```

```
            MAK +N+TGEEV+ALAA YM + D AFV KAL YATAAHFYQVRKSGEPYIVHPIQVAGI
Sbjct: 1    MAKEINLTGEEVVALAAKYMNETDAAFVKKALDYATAAHFYQVRKSGEPYIVHPIQVAGI 60

Query: 65   LADLHLDAVTVACGFLHDVVEDTDITLDEIEADFGHDARDIVDGVTKLGEVEYKSHEEQL 124
            LADLHLDAVTVACGFLHDVVEDTDITLD IE DFG D RDIVDGVTKLG+VEYKSHEEQL
Sbjct: 61   LADLHLDAVTVACGFLHDVVEDTDITLDNIEFDFGKDVRDIVDGVTKLGKVEYKSHEEQL 120

Query: 125  AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 184
            AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI
Sbjct: 121  AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180

Query: 185  SRIKWELEDLAFRYLNETEFYKISHMMKEKRREREALVEAIVSKVKTYTTQQGLFGDVYG 244
            SRIKWELEDLAFRYLNETEFYKISHMM EKRREREALV+ IV+K+K+YTT+QGLFGDVYG
Sbjct: 181  SRIKWELEDLAFRYLNETEFYKISHMMNEKRREREALVDDIVTKIKSYTTEQGLFGDVYG 240

Query: 245  RPKHIYSIYRKMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 304
            RPKHIYSIYRKMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI
Sbjct: 241  RPKHIYSIYRKMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 300

Query: 305  AAPKANGYQSIHTTVYGPKGPIEIQIRTKDMHQVAEYGVAAHWAYKKGVRGKVNQAEQAV 364
            AAPKANGYQSIHTTVYGPKGPIEIQIRTK+MHQVAEYGVAAHWAYKKGVRGKVNQAEQ V
Sbjct: 301  AAPKANGYQSIHTTVYGPKGPIEIQIRTKEMHQVAEYGVAAHWAYKKGVRGKVNQAEQKV 360

Query: 365  GMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKESGPIDFAYAI 424
            GMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPK+SGPIDFAYAI
Sbjct: 361  GMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKDSGPIDFAYAI 420

Query: 425  HTQIGEKATGAKVNGRMVPLTAKLKTGDVVEIITNANSFGPSRDWVKLVKTNKARNKIRQ 484
            HTQ+GEKA GAKVNGRMVPLTAKLKTGDVVEI+TN NSFGPSRDW+KLVKTNKARNKIRQ
Sbjct: 421  HTQVGEKAIGAKVNGRMVPLTAKLKTGDVVEIVTNPNSFGPSRDWIKLVKTNKARNKIRQ 480

Query: 485  FFKNQDKELSVNKGRDLLVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG 544
            FFKNQDKELSVNKGRD+LVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG
Sbjct: 481  FFKNQDKELSVNKGRDMLVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG 540

Query: 545  DISPISVFNKLTEKERREEERAKAKAEAEELVKGGEVKHENKDVLKVRSENGVIIQGASG 604
            DISP+SVFNKLTEKERREEERAKAKAEAEELV GGE+KHENKDVLKVRSENGVIIQGASG
Sbjct: 541  DISPVSVFNKLTEKERREEERAKAKAEAEELVNGGEIKHENKDVLKVRSENGVIIQGASG 600

Query: 605  LLMRIAKCCNPVPGDPIDGYITKGRGIAIHRSDCHNIKSQDGYQERLIEVEWDLDNSSKD 664
            LLMRIAKCCNPVPGDPI+GYITKGRGIAIHR+DC+NIKSQDGYQERLIEVEWDLDNSSKD
Sbjct: 601  LLMRIAKCCNPVPGDPIEGYITKGRGIAIHRADCNNIKSQDGYQERLIEVEWDLDNSSKD 660

Query: 665  YQAEIDIYGLNRSGLLNDVLQILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT 724
            YQAEIDIYGLNR GLLNDVLQILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT
Sbjct: 661  YQAEIDIYGLNRRGLLNDVLQILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT 720

Query: 725  VVEKIKAVPDVYSVKRTNG 743
            VVEKIKAVPDVYSVKRTNG
Sbjct: 721  VVEKIKAVPDVYSVKRTNG 739
```

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

```
Identities = 635/739 (85%), Positives = 691/739 (92%), Gaps = 1/739 (0%)

Query: 1    MVKEINLTGEEVVAITSQYMSETDVAFVKFALNYATAAHYYQARKSGEPYIIHPIQVAGI 60
            M K +N+TGEEV+A+ + YM++ DVAFV  AL YATAAH+YQ RKSGEPYI+HPIQVAGI
Sbjct: 5    MAKIMNVTGEEVIALAATYMTKADVAFVAKALAYATAAHFYQVRKSGEPYIVHPIQVAGI 64

Query: 61   LADLHLDAVTVACGFLHDVVEDTEITLDEIETDFGKDVRDIIDGVTKLGKVEYKSHEEQL 120
            LADLHLDAVTVACGFLHDVVEDT+ITLDEIE DFG D RDI+DGVTKLG+VEYKSHEEQL
Sbjct: 65   LADLHLDAVTVACGFLHDVVEDTDITLDEIEADFGHDARDIVDGVTKLGEVEYKSHEEQL 124

Query: 121  AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180
            AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI
Sbjct: 125  AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 184

Query: 181  SRIKWELEDLSFRYLNETEFYKISHMMSEKRREREELVDIIVDKIRSYTEEQGLYGDIYG 240
```

```
            SRIKWELEDL+FRYLNETEFYKISHMM EKRRERE LV+ IV K+++YT +QGL+GD+YG
Sbjct: 185 SRIKWELEDLAFRYLNETEFYKISHMMKEKRREREALVEAIVSKVKTYTTQQGLFGDVYG 244

Query: 241 RPKHIYSIYRKMRDKKKRFDQIYDLIAIRCIMETASDVYAMVGYIHELWRPMPGRFKDYI 300
            RPKHIYSIYRKMRDKKKRFDQI+DLIAIRC+MET SDVYAMVGYIHELWRPMPGRFKDYI
Sbjct: 245 RPKHIYSIYRKMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 304

Query: 301 AAPKANGYQSIHTTVYGPKGPIEIQIRTKEMHQVAEFGVAAHWAYKKGITSKVNQAEQSV 360
            AAPKANGYQSIHTTVYGPKGPIEIQIRTK+MHQVAE+GVAAHWAYKKG+  KVNQAEQ+V
Sbjct: 305 AAPKANGYQSIHTTVYGPKGPIEIQIRTKDMHQVAEYGVAAHWAYKKGVRGKVNQAEQAV 364

Query: 361 GMGWIQELVELQDESK-DAKDFVDSVKEDIFTERIYVFTPNGAVQELPRESGPIDFAYAI 419
            GM WI+ELVELQD S  DA DFVDSVKEDIF+ERIYVFTP GAVQELP+ESGPIDFAYAI
Sbjct: 365 GMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKESGPIDFAYAI 424

Query: 420 HTQVGEKATGAKVNGRMVPLTAKLKTGDVVEIITNPNSFGPSRDWIKIVKTNKARNKIRQ 479
            HTQ+GEKATGAKVNGRMVPLTAKLKTGDVVEIITN NSFGPSRDW+K+VKTNKARNKIRQ
Sbjct: 425 HTQIGEKATGAKVNGRMVPLTAKLKTGDVVEIITNANSFGPSRDWVKLVKTNKARNKIRQ 484

Query: 480 FFKNQDKETSINKGRELLVDYFQEQGYVPNKYLDKKHIEEILPRVSVKSEEALYAAVGFG 539
            FFKNQDKE S+NKGR+LLV YFQEQGYV NKYLDKK IE ILP+VSVKSEE+LYAAVGFG
Sbjct: 485 FFKNQDKELSVNKGRDLLVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG 544

Query: 540 DLSPISIFNKLTEKERREEERAKAKAEADELINGGEIKTDKRDVLKVKSENGVIIQGASG 599
            D+SPIS+FNKLTEKERREEERAKAKAEA+EL+ GGE+K + +DVLKV+SENGVIIQGASG
Sbjct: 545 DISPISVFNKLTEKERREEERAKAKAEAEELVKGGEVKHENKDVLKVRSENGVIIQGASG 604

Query: 600 LLMRIAKCCNPVPGDLIEGYITKGRGVAIHRSDCQNLKSQENYEQRLIDVEWDDDGSKKE 659
            LLMRIAKCCNPVPGD I+GYITKGRG+AIHRSDC N+KSQ+ Y++RLI+VEWD D S K+
Sbjct: 605 LLMRIAKCCNPVPGDPIDGYITKGRGIAIHRSDCHNIKSQDGYQERLIEVEWDLDNSSKD 664

Query: 660 YMAEIDIYGLNRSGLLNDVLQTLSNATKLVSTVNAQPTKDMKFANIHVSFGISNLAQLTT 719
            Y AEIDIYGLNRSGLLNDVLQ LSN+TK +STVNAQPTKDMKFANIHVSFGI NL  LTT
Sbjct: 665 YQAEIDIYGLNRSGLLNDVLQILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT 724

Query: 720 VVDKIKIIPDVYSVKRTNG 738
            VV+KIK +PDVYSVKRTNG
Sbjct: 725 VVEKIKAVPDVYSVKRTNG 743
```

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

Example 501

A DNA sequence (GBSx0539) was identified in *S.agalactiae* <SEQ ID 1605> which encodes the amino acid sequence <SEQ ID 1606>. This protein is predicted to be 2',3'-cyclic-nucleotide 2'-phosphodiesterase precursor (cpdB). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -5.79   Transmembrane  779 - 795 ( 778 - 797)

----- Final Results -----
               bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB12613 GB:Z99108 similar to 2',3'-cyclic-nucleotide
          2'-phosphodiesterase [Bacillus subtilis]
 Identities = 297/630 (47%), Positives = 419/630 (66%), Gaps = 21/630 (3%)

Query: 102 KVDLRIMSTTDLHTNLVNYDYYQDKESQKIGLAKTAVLIEEAKKENPNTVLVDNGDVIQG 161
           +V L I++TTD+H N+++YDYY DKE+   GLA+TA LI++ +++NPNT+LVDNGD+IQG
```

```
Sbjct: 42  QVHLSILATTDIHANMMDYDYYSDKETADFGLARTAQLIQKHREQNPNTLLVDNGDLIQG 101

Query: 162 TPLGTYKAIVKP---VAENEEHPMYQAMNALGYDASTLGNHEFNYGLDYLKKIIATANLP 218
            PLG Y    +    ++  + HP+   MNAL YDA TLGNHEFNYGLD+L   I  A+ P
Sbjct: 102 NPLGEYAVKYQKDDIISGTKTHPIISVMNALKYDAGTLGNHEFNYGLDFLDGTIKGADFP 161

Query: 219 ILNANVLDFKTHQPVFKTYDIITKTFKDSTGRAVALNIGITGIVPPQILNWDKANLEGKV 278
            I+NANV    + +  +  Y I  KT  D  G    + +G  G VPPQI+ WDK NLEG+V
Sbjct: 162 IVNANVKT-TSGENRYTPYVINEKTLIDENGNEQKVKVGYIGFVPPQIMTWDKKNLEGQV 220

Query: 279 IVKDSVKAIEEIVPTMRAKGADVILVLSHSGIGDDRYEEGEENVGYQIAS-IKGVDAVVT 337
            V+D V++  E +P M+A+GADVI+ L+H+GI       G EN  + +A+  KG+DA+++
Sbjct: 221 QVQDIVESANETIPKMKAEGADVIIALAHTGIEKQAQSSGAENAVFDLATKTKGIDAIIS 280

Query: 338 GHSHAEFPSGNGTGFYEKYTGVDGIN---GKINGTPVTMAGKYGDHLGIIDLGLSYTNGK 394
            GH H  FPS        +Y GV   N  G ING PV M   +G +LG+IDL L   +G
Sbjct: 281 GHQHGLFPSA-------EYAGVAQFNVEKGTINGIPVVMPSSWGKYLGVIDLKLEKADGS 333

Query: 395 WQVSESSAKIRKIDMNSTTADERIIALAKEAHDGTINYVRQQVGTTTAPITSYFALVKDD 454
            W+V++S   I  I  N T+ +E +    ++ H  T+ YVR+ VG T A I S+FA VKDD
Sbjct: 334 WKVADSKGSIESIAGNVTSRNETVTNTIQQTHQNTLEYVRKPVGKTEADINSFFAQVKDD 393

Query: 455 PSVQIVNNAQRWYVANELKGTPEANLPLLSAAAPFKAGTRGDATAYTDIPAGPVAIKNVA 514
            PS+QIV +AQ+WY   E+K T   NLP+LSA APFKAG R  A  YT+IPAG +AIKNV
Sbjct: 394 PSIQIVTDAQKWYAEKEMKDTEYKNLPILSAGAPFKAGGRNGANYYTNIPAGDLAIKNVG 453

Query: 515 DLYLYDNVTALLKVTGADLREWLEMSAGQFNQIDPNNKAPQNIINTEYRTYNFDVIDGLT 574
            DLYLYDN   ++K+TG+++++WLEMSAGQFNQIDP     Q ++N  +R+YNFDVIDG+T
Sbjct: 454 DLYLYDNTVQIVKLTGSEVKDWLEMSAGQFNQIDPAKGGDQALLNENFRSYNFDVIDGVT 513

Query: 575 YKFDITQPNKYNKDGKVVNSQASRVRDLMYNGKPVADKQEFMIVTNNYRASGTFPGAKNA 634
            Y+ D+T+P KYN++GKV+N+ +SR+ +L Y GKP++  QEF++VTNNYRASG   G  +
Sbjct: 514 YQVDVTKPAKYNENGKVINADSSRIINLSYEGKPISPSQEFLVVTNNYRASGG-GGFPHL 572

Query: 635 TMNRLLN---LENRQTIINYIISEKTINPTADNNWGFTESIKDLDLRFQTADKAKNLVTN 691
            T +++++   +ENRQ +++YII +KT+NP ADNNW     +   +L F+++  AK
Sbjct: 573 TSDKIVHGSAVENRQVLMDYIIEQKTVNPKADNNWSIA-PVSGTNLTFESSLLAKPFADK 631

Query: 692 SKDIQYIASSTKDEGFGDYRFVYTEQEKVD 721
            + D+ Y+  S  +EG+G Y+  + +    D
Sbjct: 632 ADDVAYVGKSA-NEGYGVYKLQFDDDSNPD 660
 Identities = 133/567 (23%), Positives = 214/567 (37%), Gaps = 147/567 (25%)

Query: 104 DLRIMSTTDLHTNLVNYDYYQDKESQKIGLAKTAVLIEEAKKENPNTVLVDNGDVIQGTP 163
            DL +M T D H +L +                A+    I E + E  + +L+D GDV  G
Sbjct: 668 DLTVMHTNDTHAHLDD-------------AARRMTKINEVRSETNHNILLDAGDVFSGD- 713

Query: 164 LGTYKAIVKPVAENEEHPMYQAMNALGYDASTLGNHEFNYG----LDYLKKIIATAN--- 216
              Y      +A+       + MN +GYDA T GNHEF+ G     D+L    AT +
Sbjct: 714 --LYFTKWNGLAD------LKMMNMMGYDAMTFGNHEFDKGPTVLSDFLSGNSATVDPAN 765

Query: 217 --------LPILNANVLDFKTHQPVFKTYDIITKTF----KDSTGRAVALNIGITG--IV 262
                   PI++ANV    +++P  K++    +TF    K   G    + + + G  +
Sbjct: 766 RYHFEAPEFPIVSANV--DVSNEPKLKSFVKKPQTFTAGEKKEAGIHPYILLDVDGEKVA 823

Query: 263 PPQILNWDKANLE--GKVIV--------KDSVKAIEEIVPTMRAKGADVILVLSHSGIGD 312
              +  D A     GK IV        +++VKAI+E       +  + I+ L+H G
Sbjct: 824 VFGLTTEDTATTSSPGKSIVFNDAFETAQNTVKAIQE------EEKVNKIIALTHIG--- 874

Query: 313 DRYEEGEENVGYQIA-SIKGVDAVVTGHSHAEFPSGNGTGFYEKYTGVDGINGKINGTP- 370
            N   ++A  +KG+D ++ GH+H               T VD +    N  P
Sbjct: 875 -------HNRDLELAKKVKGIDLIIGGHTH---------------TLVDKMEVVNNEEPT 912

Query: 371 -VTMAGKYGDHLGIIDLGLSYTNGKWQVSESSAKIRKIDMNSTTADERIIALAKEAHDGT 429
            V  A +YG  LG +D+       G  Q  +S+  +  ID ++    E     AK+  D
Sbjct: 913 IVAQAKEYGQFLGRVDVAFD-EKGVVQTDKSNLSVLPIDEHTEENPE-----AKQELDQF 966

Query: 430 INYV----RQQVGTTTAPITSYFALVKDDPSVQIVNNAQRWYVANELKGTPEANLPLLSA 485
```

```
              N +     ++VG T                  +  + QR +V  +           + A
Sbjct: 967   KNELEDVKNEKVGYT----------------DVALDGQREHVRTKETNLGNFIADGMLA 1009

Query: 486   AAPFKAGTRGDAT----AYTDIPAGPVAIKNVADLYLYDNVTALLKVTGADLREWLEMSA 541
               A   AG R   T        I  G + +  V ++  + N   +  +TG  ++E LE
Sbjct: 1010  KAKEAAGARIAITNGGGIRAGIDKGDITLGEVLNVMPFGNTLYVADLTGKQIKEALE--- 1066

Query: 542   GQFNQIDPNNKAPQNIINTEYRTYNFDVIDGLTYKFDITQPNKYNKDGKVVNSQASRVRD 601
                           Q + N E     F  + G+ Y F +      NK G     +   V+
Sbjct: 1067  ------------QGLSNVENGGGAFPQVAGIEYTFTLN-----NKPG----HRVLEVKI 1104

Query: 602   LMYNGKPVADKQE--FMIVTNNYRASG 626
                 NG  VA   +  + + TNN+  +G
Sbjct: 1105  ESPNGDKVAINTDDTYRVATNNFVGAG 1131
```

There is also homology to SEQ ID 1608. A related sequence was also identified in GAS <SEQ ID 9129> which encodes the amino acid sequence <SEQ ID 9130>. Analysis of this protein sequence reveals the following:

```
Possible cleavage site: 27

>>> Seems to have an uncleavable N-term signal seq
     INTEGRAL     Likelihood = -4.67   Transmembrane  649 - 665 ( 648 - 666)
     INTEGRAL     Likelihood = -2.02   Transmembrane    6 -  22 (   5 -  22)
     PERIPHERAL   Likelihood =  1.85
```

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

A related GBS gene <SEQ ID 8585> and protein <SEQ ID 8586> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:      6.68
GvH: Signal Score (-7.5): 0.87
     Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 1 value:  -5.79 threshold:  0.0
   INTEGRAL    Likelihood = -5.79   Transmembrane  779 - 795 ( 778 - 797)
   PERIPHERAL  Likelihood =  0.53     251
 modified ALOM score:   1.66

*** Reasoning Step: 3

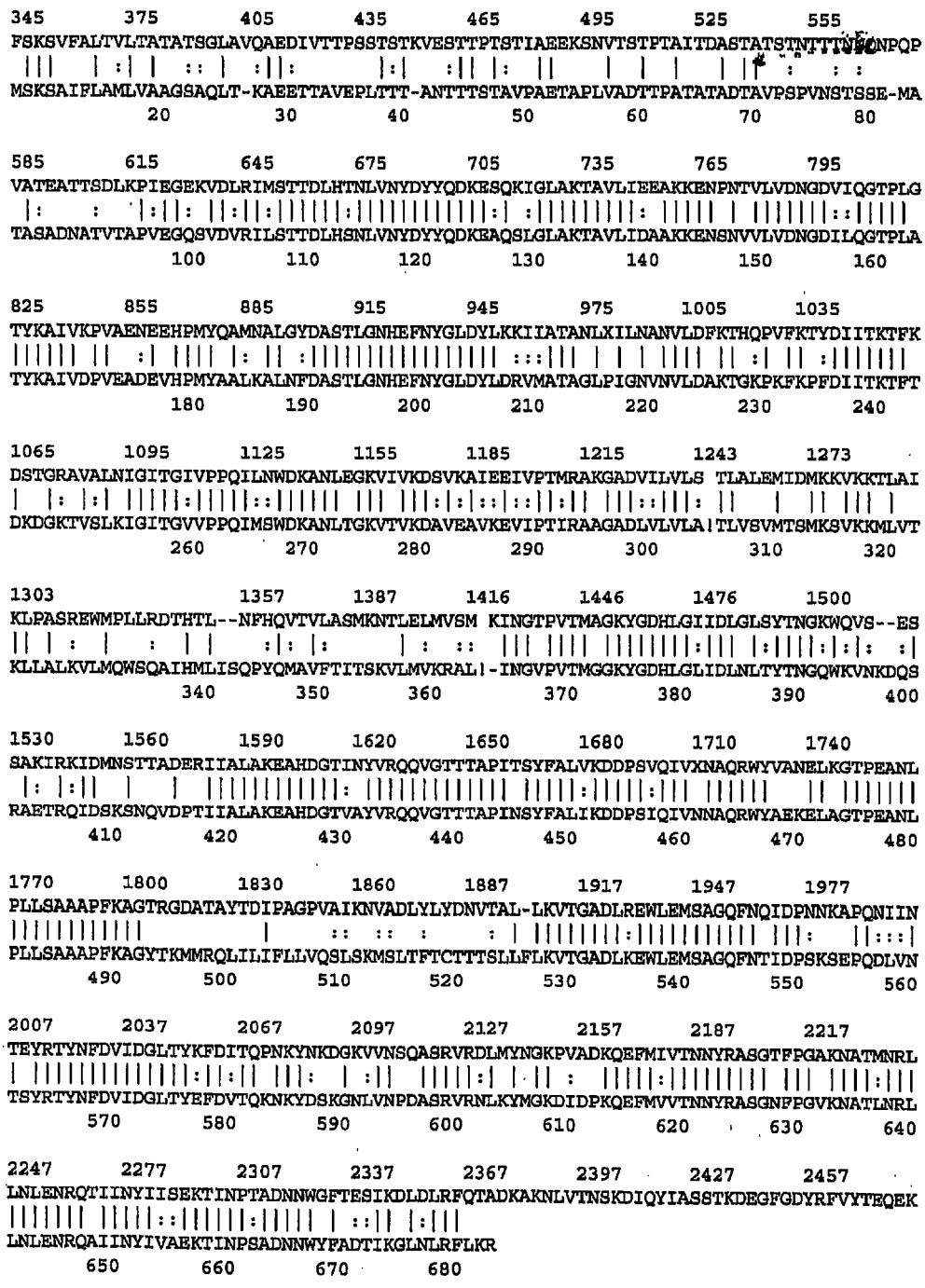
----- Final Results -----
               bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 769-773
```

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

```
ORF01378(298 - 2337 of 3000)
GP|6782402|emb|CAB70615.1||AJ133440(1 - 680 of 683) cyclo-nucleotide phosphodiesterase,
putative {Strept
ococcus dysgalactiae subsp. equisimilis}
%Match = 38.3
%Identity = 59.0  %Similarity = 72.3
Matches = 403  Mismatches = 181  Conservative Sub.s = 91

105       135       165       195       225       255       285       315
LFYHFLT*K*KKLEAQKELXTK*MCLTKLSFINKRLFLV*SLKIIRK*D*LNVFNKL**FL *DNIHVMF*WRRFMSKHY
                                                                      |:| |
                                                                     MMTKGY
```



SEQ ID 8586 (GBS53) was expressed in *E.coli* as a His-fusion product. The purified protein is shown in Figure 196, lane 9.

Example 502

A DNA sequence (GBSx0540) was identified in *S.agalactiae* <SEQ ID 1609> which encodes the amino acid sequence <SEQ ID 1610>. Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0296(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 503

A DNA sequence (GBSx0541) was identified in *S.agalactiae* <SEQ ID 1611> which encodes the amino acid sequence <SEQ ID 1612>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1504(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 10195> which encodes amino acid sequence <SEQ ID 10196> was also identified.

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

```
>GP:CAB12860 GB:Z99109 similar to glucanase [Bacillus subtilis]
 Identities = 212/345 (61%), Positives = 268/345 (77%), Gaps = 1/345 (0%)

Query: 30  SMETTLNYIKTLTSIPSPTGFTQTIMTYIIKELEAFGYSPIRTNKGGVMVSLKGKNDTKH 89
           S+  T+  IK L SIPSPTG T  ++ YI   L+ +    +R +KGG++ +L G++ ++H
Sbjct: 3   SVRKTMELIKELVSIPSPTGNTYEVINYIESLLKEWKVETVRNHKGGLIATLPGRDTSRH 62

Query: 90  RMITAHLDTLGAMVRAIKPDGRLKIDLIGGYTYNAIEGENCTIHLSKNGKEISGTALIHQ 149
           RM+TAH+DTLGAMV+ IK DGRLKIDLIGG+ YN+IEGE C I  + +GK  +GT L+HQ
Sbjct: 63  RMLTAHVDTLGAMVKEIKADGRLKIDLIGGFRYNSIEGEYCQIETA-SGKMYTGTILMHQ 121

Query: 150 TSVHVYKDAGTAERNQTNMEIRLDEKVTTADETRALGIQVGDFISFDPRTIITDSGFIKS 209
           TSVHVYKDAG AERNQ NMEIRLDE V    +T  LGI VGDF+SFDPR  IT SGFIKS
Sbjct: 122 TSVHVYKDAGKAERNQENMEIRLDEPVHCRKDTEELGIGVGDFVSFDPRVEITSSGFIKS 181

Query: 210 RYLDDKVSAGILMELLSVYKKEDIQLPYTTHFYFSAFEELGHGANSSIPNETVEYLAVDM 269
           R+LDDK S  +L+ L+   + EDI+LPYTTHF  S  EE+G+G NS+IP ETVEYLAVDM
Sbjct: 182 RHLDDKASVALLLRLIHEIQTEDIELPYTTHFLISNNEEIGYGGNSNIPPETVEYLAVDM 241

Query: 270 GAMGDDQETDEYTVSICVKDASGPYHYELRQHLVSLAENNNIPYKLDIYPYYGSDASAAM 329
           GA+GD Q TDEY+VSICVKDASGPYHY+LR+HLV LAE ++I YKLDIYPYYGSDASAA+
Sbjct: 242 GAIGDGQATDEYSVSICVKDASGPYHYQLRKHLVQLAEKHHIDYKLDIYPYYGSDASAAI 301

Query: 330 RAGAEVKHALLGAGIESSHSYERTHIDSIQATELLVDAYLKSNMV 374
           ++G ++ H L+G GI++SH++ERTH  S++ T  L+  Y++S MV
Sbjct: 302 KSGHDIVHGLIGPGIDASHAFERTHKSSLRHTAKLLYYYVQSPMV 346
```

There is also homology to SEQ ID 424.

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

Example 504

A DNA sequence (GBSx0542) was identified in *S.agalactiae* <SEQ ID 1613> which encodes the amino acid sequence <SEQ ID 1614>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3157(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:AAF11472 GB:AE002031 conserved hypothetical protein [Deinococcus radiodurans]
 Identities = 55/150 (36%), Positives = 85/150 (56%), Gaps = 2/150 (1%)

Query: 5   LIIIRGNSASGKSTIAKQLQAELGENTLLLSQDYLRREMLGTKDGENTTTIPLLINLLNY 64
           LI++RGNS SGKS++A+ L+   G      + QDYLRR +L   D       I L+   + Y
Sbjct: 23  LIVLRGNSGSGKSSVARALRERFGYGLAWVEQDYLRRVLLREHDVAGGKNIGLIETNVRY 82

Query: 65  GYHNCSYIILEGILRSDWYTPVWKHILKHNPNNTYAYYYDLSFQETVKRHSTRLKSLEFG 124
                S  +LEGIL S  Y P+ + +  H      + +Y+DL F+ETV+RH+TR ++ +FG
Sbjct: 83  CLSAGSVTVLEGILFSRHYGPMLERL--HADFGGHWFYFDLPFEETVRRHATRPQAADFG 140

Query: 125 EDSLARWWLEKDFLKEIPEKILTKAMSLED 154
               +  W+  +D L  + E+++  A SL D
Sbjct: 141 VQDMQAWFQARDVLPFVQEQLIGPASSLAD 170
```

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 505

A DNA sequence (GBSx0543) was identified in *S.agalactiae* <SEQ ID 1615> which encodes the amino acid sequence <SEQ ID 1616>. This protein is predicted to be periplasmic-iron-binding protein BitC. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.46   Transmembrane    9 -  25 (   5 -  30)

----- 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>
```

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

```
>GP:AAD18094 GB:U75349 periplasmic-iron-binding protein BitA
           [Brachyspira hyodysenteriae] (ver 2)
 Identities = 114/331 (34%), Positives = 184/331 (55%), Gaps = 3/331 (0%)

Query: 11  YILLVVSIIFISVFTYSISQPSKLLPPKELVILSPNSQAILTGTIPAFEEKY-GIKVKLI 69
           +I+  + ++ +++F    S  SK      LVI   +   ++   F+ K   I V+++
Sbjct: 4   FIIFCMLMLSMTLFYSCSSGDSK--NANSLVIYCSHPLDLMNTILDDFKAKNPDINVEVV 61

Query: 70  QGGTGQLIDRLSKEGKQLKADIFFGGNYTQFESHKALFESYVSKNVHTVIPDYIHPSDTA 129
             GTG+L+ R+   E      D+ +GG     +S   LFE+Y S N   ++ + +
Sbjct: 62  TAGTGELLKRVEAEKMNPLGDVLWGGTLNSVKSKTDLFENYTSTNEANILDEFKNTEGPF 121
```

```
Query: 130 TPYTINGSVLIVNNELAKGLTIKSYEDLLQPSLKGKIAFADPNTSSSAFSQLTNILLAKG 189
           T ++   S+L+VN  LA  + I+ YEDLL P LKGKIA ADP+ SSSAF  L N+L A G
Sbjct: 122 TRFSAIPSILMVNTNLAGNIKIEGYEDLLNPELKGKIAAADPSASSSAFEHLVNMLYAMG 181

Query: 190 GYTNPKAWNYVKKLQHNINAIKSSSSSEVYQSVAEGKMIVGLTYEDPSVNLQKSGANVSI 249
                K W+YV+KL  N++    S SS VY+ VA+G+  VGLTYE+P ++   SG+ V +
Sbjct: 182 KGDPEKGWDYVQKLCANLDGKLLSGSSAVYKGVADGEYTVGLTYEEPGISYMSSGSPVKV 241

Query: 250 VYPTEGTVFVPSSVAIIKNAPSMKEAKLFINFMLSLDVQNAFGQSTSNRPIRKDAQTSNG 309
           +Y  EG +  P  V IIK   +++ AK FI++ +SLD QN   +  S R IR DA  ++
Sbjct: 242 IYMKEGVISKPDGVYIIKGGKNLENAKKFIDYCVSLDAQNMLVEKLSRRSIRSDAVVTDM 301

Query: 310 MKALKDIATLKEDYRYVTKHKGQILKTYNRI 340
           +K + +I ++ ++   V + + + L  +  I
Sbjct: 302 VKPMSEIYSITDNADVVEESRQKWLDKFKDI 332
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1617> which encodes the amino acid sequence <SEQ ID 1618>. Analysis of this protein sequence reveals the following:

```
      Possible site: 33

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.16   Transmembrane    9 -  25 (   4 -  33)

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

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

```
>GP:AAB95371 GB:U75349 periplasmic-iron-binding protein BitC
           [Brachyspira hyodysenteriae]
 Identities = 115/324 (35%), Positives = 177/324 (54%), Gaps = 8/324 (2%)

Query: 15  VIIILAIVNVAMYIF-----SSSKKDSAKELVILTPNSQTILTGTIPAFEEKY-GVKVRL 68
           +++I   + ++++IF     S S   S   LVI  P+    +  + F+ K  G+ V +
Sbjct: 4   IVLIFTSLLLSVFIFYSCSSSESGAQSGNSLVIYCPHPLEFINPLVDDFKAKNPGINVDI 63

Query: 69  IQGGTGQLIDQL-GRKDKPLNADIFFGGNYTQFESHKDLFESYVSPQVSTVISDYQLPSH 127
           I  GTG+L+ ++   KD PL  DI +GG  +  +   DLFESY S     +   Y+
Sbjct: 64  IAAGTGELLKRVESEKDNPLG-DILWGGTISMAKPKIDLFESYTSTNEENIAEIYKNTEG 122

Query: 128 RATPYTINGSVLIVNNELARGLHITSYEDLLQPALKGKIAFADPNSSSSAFSQLTNILLA 187
             T  T   S+L+VN  LA  + I  YEDLL P LKGKIAFADP++SSS+F  L N+L A
Sbjct: 123 ALTRCTAVPSILMVNTNLAGDIKIEGYEDLLNPELKGKIAFADPSASSSSFEHLVNMLYA 182

Query: 188 KGGYTNADAWAYMKRLLVNMNSIRATSSSEVYQSVAEGKMIVGLTYEDPCINLQKSGANV 247
            G       W Y+ +L  N++    + SS VY+ VA+G+  VGLT+E+   N   +G+ V
Sbjct: 183 IGKGDPEKGWDYVSKLCANLDGKLLSGSSAVYKGVADGEYTVGLTFEEGGANYVSAGSPV 242

Query: 248 SIVYPKEGTVFVPSSVAIIKHAPNMTEAKLFINFMLSRDVQNAFGQSTSNRPIRQDAQTS 307
            +VY KEG +  P  + IIK+A N+  AK F+++  S D Q       + R +R D   S
Sbjct: 243 KLVYMKEGVIIKPDGIYIIKNAKNLENAKKFVDYATSYDAQKTITDKLNRRSVRGDLPPS 302

Query: 308 HDMKALETIATLKEDYAYVTKHKK 331
             +++++TI  + +D A V ++K+
Sbjct: 303 AILQSVDTINVITDDEAVVDQNKQ 326
```

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

```
 Identities = 257/345 (74%), Positives = 295/345 (85%), Gaps = 1/345 (0%)

Query: 1   MKEKQSKRLIYILLVVSIIFISVFTYSISQPSKLLPPKELVILSPNSQAILTGTIPAFEE 60
           +K K+   L ++L+++ +  ++V  Y  S  SK    KELVIL+PNSQ ILTGTIPAFEE
Sbjct: 2   LKLKRKWLLSFLLVIIILAIVNVAMYIFSS-SKKDSAKELVILTPNSQTILTGTIPAFEE 60
```

```
Query: 61   KYGIKVKLIQGGTGQLIDRLSKEGKQLKADIFFGGNYTQFESHKALFESYVSKNVHTVIP 120
            KYG+KV+LIQGGTGQLID+L ++ K L ADIFFGGNYTQFESHK LFESYVS  V TVI
Sbjct: 61   KYGVKVRLIQGGTGQLIDQLGRKDKPLNADIFFGGNYTQFESHKDLFESYVSPQVSTVIS 120

Query: 121  DYIHPSDTATPYTINGSVLIVNNELAKGLTIKSYEDLLQPSLKGKIAFADPNTSSSAFSQ 180
             DY  PS  ATPYTINGSVLIVNNELA+GL I SYEDLLQP+LKGKIAFADPN+SSSAFSQ
Sbjct: 121  DYQLPSHRATPYTINGSVLIVNNELARGLHITSYEDLLQPALKGKIAFADPNSSSSAFSQ 180

Query: 181  LTNILLAKGGYTNPKAWNYVKKLQHNINAIKSSSSSEVYQSVAEGKMIVGLTYEDPSVNL 240
            LTNILLAKGGYTN  AW Y+K+L  N+N+I+++SSSEVYQSVAEGKMIVGLTYEDP +NL
Sbjct: 181  LTNILLAKGGYTNADAWAYMKRLLVNMNSIRATSSSEVYQSVAEGKMIVGLTYEDPCINL 240

Query: 241  QKSGANVSIVYPTEGTVFVPSSVAIIKNAPSMKEAKLFINFMLSLDVQNAFGQSTSNRPI 300
            QKSGANVSIVYP EGTVFVPSSVAIIK+AP+M EAKLFINFMLS DVQNAFGQSTSNRPI
Sbjct: 241  QKSGANVSIVYPKEGTVFVPSSVAIIKHAPNMTEAKLFINFMLSRDVQNAFGQSTSNRPI 300

Query: 301  RKDAQTSNGMKALKDIATLKEDYRYVTKHKGQILKTYNRIRRNAD 345
            R+DAQTS+ MKAL+ IATLKEDY YVTKHK +I+ TYN++R+  +
Sbjct: 301  RQDAQTSHDMKALETIATLKEDYAYVTKHKKKIVATYNQLRQRLE 345
```

SEQ ID 1616 (GBS263) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 4; MW 63kDa).

The GBS263-GST fusion product was purified (Figure 205, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 301), 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 506

A DNA sequence (GBSx0544) was identified in *S.agalactiae* <SEQ ID 1619> which encodes the amino acid sequence <SEQ ID 1620>. This protein is predicted to be response regulator. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4733(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:AAF31452 GB:AF221126 putative response regulator [Streptococcus pneumoniae]
 Identities = 85/252 (33%), Positives = 147/252 (57%), Gaps = 17/252 (6%)

Query: 2    YRLLIVEDEHLIRKWLRYAIDYQSLNILVVGEAKDGKEGAQLIQEEQPDIVLSDINMPIM 61
            Y +LIVEDE+L+R+ L   ++  + ++ ++G+A++G++  +LIQ++ PDI+L+DINMP +
Sbjct: 3    YTILIVEDEYLVRQGLTKLVNVAAYDMEIIGQAENGRQAWELIQKQVPDIILTDINMPHL 62

Query: 62   TAFDMFEATKGQSYAK---IILSGYADFPNAQSAIHYGVLEFLTKPLEKQALIDCLKTIM 118
               +    + ++Y +   + L+GY DF  A SA+  GV ++L KP  +Q + + L  I
Sbjct: 63   NGIQLASLVR-ETYPQVHLVFLTGYDDFDYALSAVKLGVDDYLLKPFSRQDIEEMLGKIK 121

Query: 119  ARIE-EHKEKHLQEHTELYLPLPQANDQVPEVIKDMLAWIHSHFHGKIVISQLAHDLGYS 177
              +++ E KE+ LQ+     L   +    + + I+  LA         +  +  LA DLG+S
Sbjct: 122  QKLDKEEKEEQLQD-----LLTNRFEGNMAQKIQSHLA------DSQFSLKSLASDLGFS 170

Query: 178  ESYLYTVTKKHLHITLSDYINQYRINQAIQLMFREPDLMVYQIAEAVGIYDYRYFDRVFK 237
             +YL ++ KK L +   DY+ + R+ QA +L+    DL +Y+IAE VG  D  YF + FK
```