```
Query: 242 AQIAEELGLE 251
           I +ELG+E
Sbjct: 244 DWIKQELGIE 253
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5787> which encodes the amino acid sequence <SEQ ID 5788>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2417(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 = 225/255 (88%), Positives = 241/255 (94%)

Query: 1   MSVLEIKNLHVSIEDKEILKGLNLTLKTGEIAAIMGPNGTGKSTLSAAIMGNPNYEVTAG 60
           MS+LEI NLHVSIE KEILKG+NLTLKTGE+AAIMGPNGTGKSTLSAAIMGNPNYEVT G
Sbjct: 1   MSILEINNLHVSIEGKEILKGVNLTLKTGEVAAIMGPNGTGKSTLSAAIMGNPNYEVTQG 60

Query: 61  EILFDGEDILELEVDERARLGLFLAMQYPSEVPGITNAEFIRAAMNAGKADDDKISIRQF 120
           +IL DG +IL+LEVDERARLGLFLAMQYPSE+PGITNAEF+RAAMNAGKAD+DKIS+R F
Sbjct: 61  QILLDGVNILDLEVDERARLGLFLAMQYPSEIPGITNAEFMRAAMNAGKADEDKISVRDF 120

Query: 121 ITKLDEKMELLGMKEEMAERYLNEGFSGGEKKRNEILQLLMLEPKFALLDEIDSGLDIDA 180
           ITKLDEKM LLGMKEEMAERYLNEGFSGGEKKRNEILQLLMLEPKFALLDEIDSGLDIDA
Sbjct: 121 ITKLDEKMALLGMKEEMAERYLNEGFSGGEKKRNEILQLLMLEPKFALLDEIDSGLDIDA 180

Query: 181 LKVVSKGVNEMRGEGFGAMIITHYQRLLNYITPDKVHVMMDGKVVLSGGPELAVRLEKEG 240
           LKVVSKGVNEMRG+ FGAMIITHYQRLLNYITPD VHVMMDG++VLSG   LA RLEKEG
Sbjct: 181 LKVVSKGVNEMRGKDFGAMIITHYQRLLNYITPDLVHVMMDGRIVLSGDAALATRLEKEG 240

Query: 241 YAQIAEELGLEYKEE 255
           YA IA++LG+EYKEE
Sbjct: 241 YAGIAQDLGIEYKEE 255
```

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

Example 1862

A DNA sequence (GBSx1969) was identified in *S.agalactiae* <SEQ ID 5789> which encodes the amino acid sequence <SEQ ID 5790>. This protein is predicted to be RgpG (rfe). Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.10   Transmembrane  312 - 328 ( 308 - 336)
   INTEGRAL    Likelihood =-10.03   Transmembrane   15 -  31 (   6 -  41)
   INTEGRAL    Likelihood = -9.82   Transmembrane  205 - 221 ( 197 - 226)
   INTEGRAL    Likelihood = -8.60   Transmembrane  335 - 351 ( 329 - 358)
   INTEGRAL    Likelihood = -7.48   Transmembrane  257 - 273 ( 255 - 281)
   INTEGRAL    Likelihood = -5.52   Transmembrane   60 -  76 (  56 -  79)
   INTEGRAL    Likelihood = -5.31   Transmembrane  151 - 167 ( 148 - 171)
   INTEGRAL    Likelihood = -4.88   Transmembrane   91 - 107 (  90 - 108)
   INTEGRAL    Likelihood = -4.78   Transmembrane  184 - 200 ( 177 - 203)
   INTEGRAL    Likelihood = -3.13   Transmembrane  119 - 135 ( 119 - 135)
   INTEGRAL    Likelihood = -2.97   Transmembrane  229 - 245 ( 229 - 250)

----- Final Results -----
            bacterial membrane --- Certainty=0.5840(Affirmative) < succ>
```

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

A related GBS nucleic acid sequence <SEQ ID 8919> which encodes amino acid sequence <SEQ ID 8920> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      5.18
GvH: Signal Score (-7.5): -6.19
     Possible site: 15
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 9 value: -12.10 threshold:  0.0
   INTEGRAL    Likelihood =-12.10   Transmembrane  239 - 255 ( 235 - 263)
   INTEGRAL    Likelihood = -9.82   Transmembrane  132 - 148 ( 124 - 153)
   INTEGRAL    Likelihood = -8.60   Transmembrane  262 - 278 ( 256 - 285)
   INTEGRAL    Likelihood = -7.48   Transmembrane  184 - 200 ( 182 - 208)
   INTEGRAL    Likelihood = -5.31   Transmembrane   78 -  94 (  75 -  98)
   INTEGRAL    Likelihood = -4.88   Transmembrane   18 -  34 (  17 -  35)
   INTEGRAL    Likelihood = -4.78   Transmembrane  111 - 127 ( 104 - 130)
   INTEGRAL    Likelihood = -3.13   Transmembrane   46 -  62 (  46 -  62)
   INTEGRAL    Likelihood = -2.97   Transmembrane  156 - 172 ( 156 - 177)
   PERIPHERAL  Likelihood = 12.63     284
modified ALOM score:   2.92

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5840(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:BAA82114 GB:AB022909 RgpG [Streptococcus mutans]
 Identities = 266/382 (69%), Positives = 317/382 (82%)

Query: 10  TIEYIFVLIGAFLLSIILTPIIRVISLKVGAVDKPNARRINKVPMPSSGGLAIFLSFVVT 69
           T++++ VLI   L S++LTP++R  +L+VGAVD PNARRINKVPMPS+GGLAI +SFV+
Sbjct: 7   TLKFVLVLIATLLTSLVLTPLVRFFALRVGAVDNPNARRINKVPMPSAGGLAIIISFVIA 66

Query: 70  TLFFMPMAASRHFIEVSYFHYILPVIIGGLVVTTTGFIDDIFELRPRYKMLGIIIAAIII 129
           TL  MPM          SYF YILPV++G LV+  TGFIDD++EL P+ K LGI++ A+II
Sbjct: 67  TLALMPMILKTQIGGKSYFEYILPVVLGALVIALTGFIDDVYELSPKIKFLGILLGAVII 126

Query: 130 WKFTHFRFDSFKIPIGGPLLEFGPILTFFLTVLWIISITNAINLIDGLDGLVSGVSIISL 189
           W FT FRFDSFKIP GGP+L F P L+FFLT+LW+++ITNA+NLIDGLDGLVSGVS+ISL
Sbjct: 127 WIFTDFRFDSFKIPFGGPMLHFNPFLSFFLTILWVVAITNAVNLIDGLDGLVSGVSMISL 186

Query: 190 ATMAVVSYFFLPKIDFFLTLTIVILIASIVGFFPYNYHPAIIYLGDAGALFIGFMIGVLS 249
            TM +VSYFFL   D FLTLTI +LI +I GFFPYNYHPAIIYLGD GALFIGFMI VLS
Sbjct: 187 TTMGLVSYFFLYDTDIFLTLTIFVLIFAIAGFFPYNYHPAIIYLGDTGALFIGFMISVLS 246

Query: 250 LQGLKNSTAVAVITPVIILGVPILDTAVAIVRRKLSGKKISEADKMHLHHRLLSMGFTHR 309
           LQGLKN+TAVAV+TP+I+LGVPI+DT VAI+RR LSG+K  EAD MHLHHRLL+MGFTHR
Sbjct: 247 LQGLKNATAVAVVTPIIVLGVPIVDTTVAIIRRTLSGQKFYEADNMHLHHRLLAMGFTHR 306

Query: 310 GAVLVVYGIAIIFSLIALLLNVSSRIGGIFLLLALLLAMEIFIEGLNIWGENRTPLFNLL 369
           GAVLVVYGIA+ FSL++LLLNVSSR+GGI L++ +  A+EIFIEGL IWG  RTPLF LL
Sbjct: 307 GAVLVVYGIAMFFSLVSLLLNVSSRLGGILLMIGVAFALEIFIEGLEIWGPKRTPLFRLL 366

Query: 370 KFIGNSDYRQSVIAKYSDKHQK 391
           FIGNSDYRQ V+AKY  K +K
Sbjct: 367 AFIGNSDYRQEVVAKYRRKKKK 388
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5791> which encodes the amino acid sequence <SEQ ID 5792>. Analysis of this protein sequence reveals the following:

```
    Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.28   Transmembrane    9 -  25 (   1 -  33)
   INTEGRAL    Likelihood = -8.17   Transmembrane  201 - 217 ( 198 - 221)
   INTEGRAL    Likelihood = -7.64   Transmembrane  308 - 324 ( 305 - 329)
   INTEGRAL    Likelihood = -7.17   Transmembrane   55 -  71 (  51 -  74)
   INTEGRAL    Likelihood = -7.06   Transmembrane  145 - 161 ( 138 - 170)
   INTEGRAL    Likelihood = -6.58   Transmembrane  260 - 276 ( 251 - 278)
   INTEGRAL    Likelihood = -6.21   Transmembrane  180 - 196 ( 172 - 198)
   INTEGRAL    Likelihood = -5.95   Transmembrane  331 - 347 ( 330 - 353)
   INTEGRAL    Likelihood = -5.68   Transmembrane   87 - 103 (  82 - 104)
   INTEGRAL    Likelihood = -3.93   Transmembrane  113 - 129 ( 112 - 133)
   INTEGRAL    Likelihood = -2.60   Transmembrane  233 - 249 ( 232 - 250)

----- Final Results -----
               bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:BAA82114 GB:AB022909 RgpG [Streptococcus mutans]
 Identities = 289/381 (75%), Positives = 334/381 (86%), Gaps = 1/381 (0%)

Query: 5   TIDYVLVLIGALLMSLFLTPLVRFLAFRVGAVDNPNARRVNKVPMPTSGGLAIFMSFLVA 64
           T+ +VLVLI  LL SL LTPLVRF A RVGAVDNPNARR+NKVPMP++GGLAI +SF++A
Sbjct: 7   TLKFVLVLIATLLTSLVLTPLVRFFALRVGAVDNPNARRINKVPMPSAGGLAIIISFVIA 66

Query: 65  SLGLIPIASKGAMFFGQTYFSYILPVVIGATVITLTGFLDDLYELSPKLKMFGILIGAVI 124
           +L L+P+  K     G++YF YILPVV+GA VI LTGF+DD+YELSPK+K  GIL+GAVI
Sbjct: 67  TLALMPMILK-TQIGGKSYFEYILPVVLGALVIALTGFIDDVYELSPKIKFLGILLGAVI 125

Query: 125 VWAFTDFKFDSFKIPFGGPLLVFGPFLTLFLTVLWIVSITNAINLIDGLDGLVSGVSIIS 184
           +W FTDF+FDSFKIPFGGP+L F PFL+ FLT+LW+V+ITNA+NLIDGLDGLVSGVS+IS
Sbjct: 126 IWIFTDFRFDSFKIPFGGPMLHFNPFLSFFLTILWVVAITNAVNLIDGLDGLVSGVSMIS 185

Query: 185 LVTMAIVSYFFLPQKDFFLTLTILVLISAIAGFFPYNYHPAMIYLGDTGALFIGFMIGVL 244
           L TM +VSYFFL   D FLTLTI VLI AIAGFFPYNYHPA+IYLGDTGALFIGFMI VL
Sbjct: 186 LTTMGLVSYFFLYDTDIFLTLTIFVLIFAIAGFFPYNYHPAIIYLGDTGALFIGFMISVL 245

Query: 245 SLQGLKNSTAVAVVTPVIILGVPIMDTIVAIIRRSLSGQKFYEPDKMHLHHRLLSMGFTH 304
           SLQGLKN+TAVAVVTP+I+LGVPI+DT VAIIRR+LSGQKFYE D MHLHHRLL+MGFTH
Sbjct: 246 SLQGLKNATAVAVVTPIIVLGVPIVDTTVAIIRRTLSGQKFYEADNMHLHHRLLAMGFTH 305

Query: 305 RGAVLVVYGITMLFSLISLLLNVSSRIGGVLLMLGLLFGLEVFIEGLEIWGEKRTPLFNL 364
           RGAVLVVYGI M FSL+SLLLNVSSR+GG+LLM+G+ F LE+FIEGLEIWG KRTPLF L
Sbjct: 306 RGAVLVVYGIAMFFSLVSLLLNVSSRLGGILLMIGVAFALEIFIEGLEIWGPKRTPLFRL 365

Query: 365 LKFIGNSDYRQAMLLKWKEKK 385
           L FIGNSDYRQ ++ K++ KK
Sbjct: 366 LAFIGNSDYRQEVVAKYRRKK 386
```

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

```
 Identities = 282/384 (73%), Positives = 334/384 (86%), Gaps = 1/384 (0%)

Query: 6   MIPFTIEYIFVLIGAFLLSIILTPIIRVISLKVGAVDKPNARRINKVPMPSSGGLAIFLS 65
           M  FTI+Y+ VLIGA L+S+ LTP++R ++ +VGAVD PNARR+NKVPMP+SGGLAIF+S
Sbjct: 1   MFSFTIDYVLVLIGALLMSLFLTPLVRFLAFRVGAVDNPNARRVNKVPMPTSGGLAIFMS 60

Query: 66  FVVTTLFFMPMAAS-RHFIEVSYFHYILPVIIGGLVVTTTGFIDDIFELRPRYKMLGIII 124
           F+V +L  +P+A+     F   +YF YILPV+IG  V+T TGF+DD++EL P+ KM GI+I
Sbjct: 61  FLVASLGLIPIASKGAMFFGQTYFSYILPVVIGATVITLTGFLDDLYELSPKLKMFGILI 120

Query: 125 AAIIIWKFTHFRFDSFKIPIGGPLLEFGPILTFFLTVLWIISITNAINLIDGLDGLVSGV 184
            A+I+W FT F+FDSFKIP GGPLL FGP LT FLTVLWI+SITNAINLIDGLDGLVSGV
Sbjct: 121 GAVIVWAFTDFKFDSFKIPFGGPLLVFGPFLTLFLTVLWIVSITNAINLIDGLDGLVSGV 180
```

```
Query: 185 SIISLATMAVVSYFFLPKIDFFLTLTIVILIASIVGFFPYNYHPAIIYLGDAGALFIGFM 244
           SIISL TMA+VSYFFLP+ DFFLTLTI++LI++I GFFPYNYHPA+IYLGD GALFIGFM
Sbjct: 181 SIISLVTMAIVSYFFLPQKDFFLTLTILVLISAIAGFFPYNYHPAMIYLGDTGALFIGFM 240

Query: 245 IGVLSLQGLKNSTAVAVITPVIILGVPILDTAVAIVRRKLSGKKISEADKMHLHHRLLSM 304
           IGVLSLQGLKNSTAVAV+TPVIILGVPI+DT VAI+RR LSG+K  E DKMHLHHRLLSM
Sbjct: 241 IGVLSLQGLKNSTAVAVVTPVIILGVPIMDTIVAIIRRSLSGQKFYEPDKMHLHHRLLSM 300

Query: 305 GFTHRGAVLVVYGIAIIFSLIALLLNVSSRIGGIFLLLALLLAMEIFIEGLNIWGENRTP 364
           GFTHRGAVLVVYGI ++FSLI+LLLNVSSRIGG+ L+L LL  +E+FIEGL IWGE RTP
Sbjct: 301 GFTHRGAVLVVYGITMLFSLISLLLNVSSRIGGVLLMLGLLFGLEVFIEGLEIWGEKRTP 360

Query: 365 LFNLLKFIGNSDYRQSVIAKYSDK 388
           LFNLLKFIGNSDYRQ+++ K+ +K
Sbjct: 361 LFNLLKFIGNSDYRQAMLLKWKEK 384
```

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

Example 1863

A DNA sequence (GBSx1970) was identified in *S.agalactiae* <SEQ ID 5793> which encodes the amino acid sequence <SEQ ID 5794>. This protein is predicted to be negative regulator of genetic competence. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3460(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 9483> which encodes amino acid sequence <SEQ ID 9484> was also identified.

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

```
>GP:BAA82113 GB:AB022909 negative regulator of genetic competence
           [Streptococcus mutans]
 Identities = 168/248 (67%), Positives = 205/248 (81%), Gaps = 9/248 (3%)

Query: 1   MEMKQISETTLKITISMEDLEDRGMELKDFLIPQEKTEEFFYSVMDELDLPENFKNSGML 60
           MEMKQISETTLKITISMEDLE+RGMELKDFLIPQEKTEEFFY+VMDELDLPENFK SGML
Sbjct: 1   MEMKQISETTLKITISMEDLEERGMELKDFLIPQEKTEEFFYTVMDELDLPENFKGSGML 60

Query: 61  SFRVTPKKDRIDVFVTKSELSKDLNLEELADLGDISKMSPEDFFKTLEQSMLEKGDTDAH 120
           SFRVTP+ DRIDVFVTKSE++K+LNLE+L+D  DISKMSPEDFF TLE++M EKGD  A
Sbjct: 61  SFRVTPRNDRIDVFVTKSEINKNLNLEDLSDFDDISKMSPEDFFNTLEETMREKGDAAAL 120

Query: 121 AKLAEIENMMDKATQEVVEENVSEEQPEKEVETIGYVHYVFDFDNIEAVVRFSQTIDFPI 180
            KLAEIE   ++ TQ+  E+  ++E+ +       YVH+V DF NI+ V+ F++T+D+ +
Sbjct: 121 DKLAEIEKREEEKTQQ--EKGETKEKRD-------YVHFVLDFPNIQQVISFAKTVDYDV 171

Query: 181 EASELYKNGKGYHMTILLDLENQPSYFANLMYARMLEHANVGTKTRAYLKEHSIQLIHDD 240
           EASEL+K    YHMT+LL+LE++P Y+A+LM+ARMLEHA  GTKTRAYL EH +QLI  D
Sbjct: 172 EASELFKESDAYHMTVLLNLEDKPDYYADLMFARMLEHAGRGTKTRAYLLEHGVQLIKAD 231

Query: 241 AISKLQMI 248
           A+ +LQMI
Sbjct: 232 ALQELQMI 239
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5795> which encodes the amino acid sequence <SEQ ID 5796>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3307(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 = 171/253 (67%), Positives = 209/253 (82%), Gaps = 2/253 (0%)

Query: 1   MEMKQISETTLKITISMEDLEDRGMELKDFLIPQEKTEEFFYSVMDELDLPENFKNSGML 60
           MEMKQISETTLKITISM+DLE+RGMELKDFLIPQEKTEEFFYSVMDELDLP+NFK+SGML
Sbjct: 3   MEMKQISETTLKITISMDDLEERGMELKDFLIPQEKTEEFFYSVMDELDLPDNFKDSGML 62

Query: 61  SFRVTPKKDRIDVFVTKSELSKDLNLEELADLGDISKMSPEDFFKTLEQSMLEKGDTDAH 120
           SFRVTP+KDR+DVFVTKSE++KD+NLE+LA+ GD+S+M+PEDFFK+LEQSM EKGD  AH
Sbjct: 63  SFRVTPRKDRLDVFVTKSEINKDINLEDLAEFGDMSQMTPEDFFKSLEQSMREKGDVKAH 122

Query: 121 AKLAEIENMMDKATQEVV--EENVSEEQPEKEVETIGYVHYVFDFDNIEAVVRFSQTIDF 178
            KL +IE +M+   +  +  +   E E + YVHYV DF  I   V F++TIDF
Sbjct: 123 EKLEKIEEIMEDVVEATLANQSEAADPSTNHESEPLDYVHYVLDFSTITEAVAFAKTIDF 182

Query: 179 PIEASELYKNGKGYHMTILLDLENQPSYFANLMYARMLEHANVGTKTRAYLKEHSIQLIH 238
            IEASELYK    YHMTILLD++ QPSYFAN+MYAR++EHAN G+KTRAYL+EH +QL+
Sbjct: 183 SIEASELYKGSNCYHMTILLDVQQQPSYFANVMYARLIEHANPGSKTRAYLQEHGLQLML 242

Query: 239 DDAISKLQMIEMG 251
           D A+ +LQ IE+G
Sbjct: 243 DGAVEQLQKIELG 255
```

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

Example 1864

A DNA sequence (GBSx1971) was identified in *S.agalactiae* <SEQ ID 5797> which encodes the amino acid sequence <SEQ ID 5798>. This protein is predicted to be BacA (bacA). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.02   Transmembrane  115 - 131 ( 111 - 135)
   INTEGRAL    Likelihood = -8.97   Transmembrane  227 - 243 ( 219 - 247)
   INTEGRAL    Likelihood = -7.86   Transmembrane   48 -  64 (  44 -  69)
   INTEGRAL    Likelihood = -7.27   Transmembrane  263 - 279 ( 260 - 279)
   INTEGRAL    Likelihood = -7.22   Transmembrane   87 - 103 (  85 - 107)
   INTEGRAL    Likelihood = -3.50   Transmembrane    2 -  18 (   1 -  19)

----- Final Results -----
            bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAD50462 GB:AF169967 BacA [Flavobacterium johnsoniae]
 Identities = 101/275 (36%), Positives = 165/275 (59%), Gaps = 22/275 (8%)

Query: 7   LKALFLGVVEGVTEWLPVSSTGHLILVQEFMKLNQSKSFVEMFNIVIQLGAIMAVIVIYF 66
```

```
            L+A+ L V+EG+TE+LPVSSTGH+I+  F  +   + F ++F IVIQLGAI++V+V+YF
Sbjct: 4    LQAIVLAVIEGITEFLPVSSTGHMIIASSFFGIAH-EDFTKLFTIVIQLGAILSVVVLYF 62

Query: 67   KRLNPFQPGKSAREIRLTWQLWLKVVIACIPSILIALPFDNWFEAHFNFMIPIAIALIFY 126
            KR   FQ          T   + K+++A IP++++ L   ++ +       + +A++L+
Sbjct: 63   KRF--FQ----------TLDFYFKLLVAFIPAVVLGLLLSDFIDGLLENPVTVAVSLLIG 110

Query: 127  GFVFI----WVEKRNAHLKPQVTELASMSYKTAFLIGCFQVLSIVPGTSRSGATILGAII 182
            G + +    W    NA   Q      ++Y  A  IG FQ ++++PG SRSGA+I+G +
Sbjct: 111  GLILLKVDEWFNNPNAAETSQ-----KITYLQALKIGLFQCIAMIPGVSRSGASIVGGMS 165

Query: 183  IGTSRSVAADFTFFLAIPTMFGYSGLKAVKYFLDGNVLSLDQSLILLVASLTAFVVSLYV 242
               SR+ AA+F+FFLA+PTM G +  K   Y+  G  LS DQ  IL++ ++ AF+V+L
Sbjct: 166  QKLSRTTAAEFSFFLAVPTMLGATVKKCYDYYKAGFELSHDQVNILIIGNVVAFIVALLA 225

Query: 243  IRFLTDYVKRHDFTIFGKYRIVLGSLLILYWLVVH 277
            I+   ++ ++ F +FG YRI+ G +L+L   +H
Sbjct: 226  IKTFISFLTKNGFKVFGYYRIIAGIILLLIHFFIH 260
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5799> which encodes the amino acid sequence <SEQ ID 5800>. Analysis of this protein sequence reveals the following:

```
     Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.30   Transmembrane  225 - 241 ( 219 - 247)
   INTEGRAL    Likelihood = -9.24   Transmembrane  115 - 131 ( 109 - 135)
   INTEGRAL    Likelihood = -7.64   Transmembrane   48 -  64 (  44 -  69)
   INTEGRAL    Likelihood = -7.43   Transmembrane   87 - 103 (  85 - 108)
   INTEGRAL    Likelihood = -5.15   Transmembrane  263 - 279 ( 262 - 279)
   INTEGRAL    Likelihood = -3.82   Transmembrane    2 -  18 (   1 -  19)

----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAD50462 GB:AF169967 BacA [Flavobacterium johnsoniae]
 Identities = 102/269 (37%), Positives = 169/269 (61%), Gaps = 14/269 (5%)

Query: 7    LKAIFFGIIEGITEWLPVSSTGHLILVQEFIRLNQDKAFIEMFNIVIQLGAIIAVMLIYF 66
            L+AI   +IEGITE+LPVSSTGH+I+   F  +  +  F ++F IVIQLGAI++V+++YF
Sbjct: 4    LQAIVLAVIEGITEFLPVSSTGHMIIASSFFGIAHED-FTKLFTIVIQLGAILSVVVLYF 62

Query: 67   ERLNPFQPGKTAREVQLTWQLWLKVVIACIPSILIAVPLDNWFEAHFYFMVPIAIALIVY 126
            +R   FQ          T   + K+++A IP++++ L ++ +       V +A++L++
Sbjct: 63   KRF--FQ----------TLDFYFKLLVAFIPAVVLGLLLSDFIDGLLENPVTVAVSLLIG 110

Query: 127  GIAFIWIEKRNAQQEPAVTELARMSYKTAFFIGCFQVLSIVPGTSRSGATILGAIILGTS 186
            G+  + +++        A T   +++Y  A  IG FQ ++++PG SRSGA+I+G +  S
Sbjct: 111  GLILLKVDEWFNNPNAAETS-QKITYLQALKIGLFQCIAMIPGVSRSGASIVGGMSQKLS 169

Query: 187  RTVAADFTFFLAIPTMFGYSGLKAVKFFLDGHHLDFAQVLILLVASLTAFVVSLLAIRFL 246
            RT AA+F+FFLA+PTM G +  K   ++  G  L   QV IL++ ++ AF+V+LLAI+
Sbjct: 170  RTTAAEFSFFLAVPTMLGATVKKCYDYYKAGFELSHDQVNILIIGNVVAFIVALLAIKTF 229

Query: 247  TDYVKKHDFTIFGKYRIVLGSLLLIYSFF 275
              ++ K+ F +FG YRI+ G +LL+  FF
Sbjct: 230  ISFLTKNGFKVFGYYRIIAGIILLLIHFF 258
```

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

```
 Identities = 227/272 (83%), Positives = 253/272 (92%)

Query: 1    MLIIELLKALFLGVVEGVTEWLPVSSTGHLILVQEFMKLNQSKSFVEMFNIVIQLGAIMA 60
            MLIIELLKA+F G++EG+TEWLPVSSTGHLILVQEF++LNQ K+F+EMFNIVIQLGAI+A
```

```
Sbjct: 1   MLIIELLKAIFFGIIEGITEWLPVSSTGHLILVQEFIRLNQDKAFIEMFNIVIQLGAIIA 60

Query: 61  VIVIYFKRLNPFQPGKSAREIRLTWQLWLKVVIACIPSILIALPFDNWFEAHFNFMIPIA 120
           V++IYF+RLNPFQPGK+ARE++LTWQLWLKVVIACIPSILIA+P DNWFEAHF FM+PIA
Sbjct: 61  VMLIYFERLNPFQPGKTAREVQLTWQLWLKVVIACIPSILIAVPLDNWFEAHFYFMVPIA 120

Query: 121 IALIFYGFVFIWVEKRNAHLKPQVTELASMSYKTAFLIGCFQVLSIVPGTSRSGATILGA 180
           IALI YG  FIW+EKRNA  +P VTELA MSYKTAF IGCFQVLSIVPGTSRSGATILGA
Sbjct: 121 IALIVYGIAFIWIEKRNAQQEPAVTELARMSYKTAFFIGCFQVLSIVPGTSRSGATILGA 180

Query: 181 IIIGTSRSVAADFTFFLAIPTMFGYSGLKAVKYFLDGNVLSLDQSLILLVASLTAFVVSL 240
           II+GTSR+VAADFTFFLAIPTMFGYSGLKAVK+FLDG+ L   Q LILLVASLTAFVVSL
Sbjct: 181 IILGTSRTVAADFTFFLAIPTMFGYSGLKAVKFFLDGHHLDFAQVLILLVASLTAFVVSL 240

Query: 241 YVIRFLTDYVKRHDFTIFGKYRIVLGSLLILY 272
             IRFLTDYVK+HDFTIFGKYRIVLGSLL++Y
Sbjct: 241 LAIRFLTDYVKKHDFTIFGKYRIVLGSLLLIY 272
```

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

Example 1865

A DNA sequence (GBSx1972) was identified in *S.agalactiae* <SEQ ID 5801> which encodes the amino acid sequence <SEQ ID 5802>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.65   Transmembrane  494 - 510 ( 488 - 519)
   INTEGRAL    Likelihood = -8.01   Transmembrane  263 - 279 ( 256 - 288)
   INTEGRAL    Likelihood = -5.95   Transmembrane   25 -  41 (  20 -  43)
   INTEGRAL    Likelihood = -4.94   Transmembrane  475 - 491 ( 473 - 493)

----- 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 9481> which encodes amino acid sequence <SEQ ID 9482> was also identified.

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

```
>GP:AAB99606 GB:U67598 M. jannaschii predicted coding region MJ1577
          [Methanococcus jannaschii]
 Identities = 41/172 (23%), Positives = 78/172 (44%), Gaps = 19/172 (11%)

Query: 479 LISFVVIIYTLFLNYFTYFCIYLLLFGVILLLNKIIFMMTRKISNGYIVTEDGASRVYQW 538
           +IS ++ ++  F+  ++     +    ++ ++  II  +T     G          ++ +W
Sbjct: 442 VISILLAVFLYFIPKYSQTFNEVFYLSIVFVVQNIILALTPTSLFGRWKANYYKEKL-EW 500

Query: 539 TSFRNMLRDIKSFDRSELESIVLWNRILVYATLFGYADRVEKALR-VNQIDIPERFANID 597
            +F+N L ++     +   E I +W   L+Y T  G  D+V +A++ +N  ++   +  I
Sbjct: 501 DAFKNFLSNLAMIKKYSPEDISIWKDWLIYGTALGVGDKVVEAMKSLNLSELVADYVIIH 560

Query: 598 SHQFAISVNQSSNHFSTITEDVSHASNFSVNSGGSSGGFSGGGG--GGGGGA 647
           S+  ++  +  S + ST                GS GGF  GGG  GGGGGA
Sbjct: 561 SNYDSMKTSVDSVYSSTT--------------GSGGGFGAGGGFGGGGGGA 597
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5803> which encodes the amino acid sequence <SEQ ID 5804>. Analysis of this protein sequence reveals the following:

```
    Possible site: 21
>>> Seems to have a cleavable N-term signal seq.
```

```
   INTEGRAL    Likelihood = -7.91   Transmembrane  486 - 502 ( 483 - 508)
   INTEGRAL    Likelihood = -5.89   Transmembrane  465 - 481 ( 460 - 483)
   INTEGRAL    Likelihood = -2.18   Transmembrane  244 - 260 ( 241 - 260)

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

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

```
>GP:AAB99606 GB:U67598 M. jannaschii predicted coding region MJ1577
          [Methanococcus jannaschii]
 Identities = 59/263 (22%), Positives = 106/263 (39%), Gaps = 14/263 (5%)

Query: 369 FLDMAFGNKVTLPVDQLFSQYHYDADTIKQLKKTYKGKKLEQEVRQSSEQVIKAMKKASA 428
           ++ +  G K+ +    L +   Y++D +K L K  K    + E  +S  Q  K+ K
Sbjct: 346 YIKIMNGGKIEILKTDLENLDVYESDVMKFLMKYSKNNVFDPEYIKSLAQKYKSSKDKLK 405

Query: 429 AITNNVLETIKKLNLPDTYRQMTPA--EKRKSNSVQGLGCLLLILNSGLLIYLAIKESGL 486
           + +   E  K +  P    ++  A  E R    +  L  + ++L   L        ++
Sbjct: 406 KLKD---ELDKIMEYPRYSSKVVNAFLETRGKKIIIALLVISILLAVFLYFIPKYSQTFN 462

Query: 487 ALIYLALMVLTMCLGFYISLKLDQYKKLGIETPEGGVRLHQWQSFKNMIRDIDKFEDVAI 546
            + YL+++ +        I L L      G            +W +FKN + ++   +  +
Sbjct: 463 EVFYLSIVFVVQ----NIILALTPTSLFGRWKANYYKEKLEWDAFKNFLSNLAMIKKYSP 518

Query: 547 EGLVVWNRVLVYATLFGYAKKVERYLKVHRIALPEVYQAVRPGELSMVMYATTPTFVSSL 606
           E + +W   L+Y T  G   KV   +K   ++     + V    +    Y +  T V S+
Sbjct: 519 EDISIWKDWLIYGTALGVGDKVVEAMKSLNLS-----ELVADYVIIHSNYDSMKTSVDSV 573

Query: 607 SSATTSSNFSVSSGGGISGGGGG 629
            S+TT S      +GGG  GGGGG
Sbjct: 574 YSSTTGSGGGFGAGGGFGGGGGG 596
```

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

```
 Identities = 241/635 (37%), Positives = 372/635 (57%), Gaps = 18/635 (2%)

Query: 22  MKKCFLAICLALSFFMVSVQADEVDYNIPHYEGNLTIHNDNSADFTEKVTYQFDSSYNGQ 81
           MKK  + + L  S    + ++A +VDY+I +YEG L +  +N+A F +KVTYQFD+SYNGQ
Sbjct: 1   MKKILMTLVLCFSLLGIRIKAADVDYSITNYEGQLLLSKENTARFEQKVTYQFDTSYNGQ 60

Query: 82  YVTLGTAGKLPDNFDINNKPQVEVSINGKVRKVSYQIEDLEDGYRLKVFNGGEAGDTVKV 141
           Y++LG  G LP   F I+ KP+VEV   NG+   VS +  DL DGYRLK++N G+AGD V V
Sbjct: 61  YISLGRTGHLPAGFAIDQKPKVEVYQNGQQVPVSQEFSDLGDGYRLKLYNAGQAGDKVDV 120

Query: 142 NVQWKLKNVLFMHKDVGELNWIPISDWDKTLEKVDFWISTDKKVALSRLWGHLGYL-KTP 200
            V W+L ++L  ++DV ELNW PISDWDKTLEKV   ++T   +  S LW H GY  K P
Sbjct: 121 KVIWQLHHLLTAYQDVAELNWTPISDWDKTLEKVSLTVTTPTDIQDSNLWAHRGYYQKKP 180

Query: 201 PKIRQNNNRYHLTAFNVNKRLEFHGYWDRSYF--NLPTNSKNNYKKKIEYQEKMIERHGF 258
             +++ N+RY + A NV+ +LE H YWD+       P +   + K KI   E  I R
Sbjct: 181 QVLKEGNSRYQINAKNVSGQLELHAYWDKKALLGKEPVDVSTSKKNKIVALETKISRRRT 240

Query: 259 ILSFLLRILLPSFFIIVTLFISIRVFLFRKKVNKYGQFPKEHHLYEAPEDLSPLELTQSI 318
           +L  L   ++P   +   L+  I+    +K+ N+Y       H YE PEDLSPL LTQ+I
Sbjct: 241 LLQLLFGKVIPLVEVGFLLWQLIQFTRLKKQFNRYHLANHTDHSYEVPEDLSPLVLTQAI 300

Query: 319 YSMSFKNFQ---DEEKKTHL---ISQEQLIQSILLDLIDRKVL----NYDDNLLSLANLD 368
           Y  SF        E +K  +   ++ E L+Q+ LLDLID+KVL          L ++ LD
Sbjct: 301 YGQSFAYLSPTASESQKLLIPKGVTFEALVQATLLDLIDQKVLLLTKEEGKAYLEISQLD 360

Query: 369 RASDAEIDFIEFAFADSTSLKPDQLFSNYQFSYKETLRELKKQHKASDLQTQMRRRGSNA 428
           R +D E   F++ AF +  +L  DQLFS Y +   +T+++LKK +K   L+ ++R+
Sbjct: 361 RVTDEEAAFLDMAFGNKVTLPVDQLFSQYHYD-ADTIKQLKKTYKGKKLEQEVRQSSEQV 419

Query: 429 LSRITRLTRLISKDNINSLRRKGISSPYRKMSSEESKELSRLKRFSYLSPLISFVVIIYT 488
```

```
            +  + + +  I+ + + ++++  +   YR+M+  E ++ + ++    L  +++  ++IY
Sbjct: 420 IKAMKKASAAITNNVLETIKKLNLPDTYRQMTPAEKRKSNSVQGLGCLLLILNSGLLIY- 478

Query: 489 LFLNYFTYFCIYLLLFGVILLLNKIIFMMTRKISNGYIVTEDGASRVYQWTSFRNMLRDI 548
           L +       IYL L  + + L   I +   +     I T +G  R++QW SF+NM+RDI
Sbjct: 479 LAIKESGLALIYLALMVLTMCLGFYISLKLDQYKKLGIETPEGGVRLHQWQSFKNMIRDI 538

Query: 549 KSFDRSELESIVLWNRILVYATLFGYADRVEKALRVNQIDIPERFANIDSHQFAISVNQS 608
             F+   +E +V+WNR+LVYATLFGYA +VE+ L+V++I +PE +   +  ++ + +  +
Sbjct: 539 DKFEDVAIEGLVVWNRVLVYATLFGYAKKVERYLKVHRIALPEVYQAVRPGELSMVMYAT 598

Query: 609 SNHFSTITEDVSHASNFSVNSGGSSGGFSGGGGGG 643
           +  F +     + +SNFSV+SG   GG SGGGGGG
Sbjct: 599 TPTFVSSLSSATTSSNFSVSSG---GGISGGGGGG 630
```

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 8921> and protein <SEQ ID 8922> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:     10.29
GvH: Signal Score (-7.5): 3.11
     Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 3 value:  -8.65 threshold:  0.0
   INTEGRAL    Likelihood = -8.65   Transmembrane  475 - 491 ( 469 - 500)
   INTEGRAL    Likelihood = -8.01   Transmembrane  244 - 260 ( 237 - 269)
   INTEGRAL    Likelihood = -4.94   Transmembrane  456 - 472 ( 454 - 474)
   PERIPHERAL  Likelihood =  2.28     540
 modified ALOM score:   2.23

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no homology with any sequences in the databases.

Example 1866

A DNA sequence (GBSx1973) was identified in *S.agalactiae* <SEQ ID 5805> which encodes the amino acid sequence <SEQ ID 5806>. This protein is predicted to be glutamine-binding periplasmic protein/glutamine transport system perme. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.86   Transmembrane  301 - 317 ( 295 - 324)
   INTEGRAL    Likelihood = -6.05   Transmembrane  479 - 495 ( 473 - 496)
   INTEGRAL    Likelihood = -0.59   Transmembrane  369 - 385 ( 369 - 385)

----- Final Results -----
               bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:BAA17584 GB:D90907 glutamine-binding periplasmic protein
          [Synechocystis sp.]
 Identities = 147/534 (27%), Positives = 256/534 (47%), Gaps = 75/534 (14%)
```

```
Query: 4   ILLSLFTALLITFGGMTSIQADEYLRVGMEAAYAPFNWTQNDNTNGAVPIEGTDQYANGY 63
           +LL++   LL  F  ++     + + V  E  + PF  T          E T Q   G+
Sbjct: 24  VLLAIAIPLLPAFSQVSR----QTIIVATEPTFPPFEMTD----------EATGQLT-GF 68

Query: 64  DVQVAKKLAKKLNKKVVVVKTKWEGLVPALTSGKLDMIIAGMSPTEERKKEINFSKPYYI 123
           DV + + + +     V +    ++G++PAL S  +   I+ ++ T ER + ++FS PY+
Sbjct: 69  DVDLIQAIGEAAQVTVDIQGYPFDGIIPALQSNTVGAAISAITITPERAQSVSFSSPYFK 128

Query: 124 SEPTLVVNAEGKYTNAKNISDFKNAKVTAQQGVYLYNLIDQINGVKKEVAMGDFNQLRQA 183
           S   L +  +      KN+ D +  ++    G     +  + G K    + +F+ +  A
Sbjct: 129 S--VLAIAVQDGNDTIKNLKDLEGKRLAVAIGTTGAMVATNVPGAK----VTNFDSITSA 182

Query: 184 VE---SGVVDAYVSERPDATSAQTANPKLKMIELHQGFKTSDADTNISVGMRKGDNRINQ 240
           ++   +G  DA +++RP    A   +  L+ +++     + D    I++ +      INQ
Sbjct: 183 LQELVNGNADAVINDRPVLLYA-IKDAGLRNVKISADVGSEDY-YGIAMPLAP-PGEINQ 239

Query: 241 VNQVL-----ESISRDKQIALMDKMIKEQ---------PSV-----------KKEKNGK 274
             +VL      + I       A+ +K   E+          PS+             + + N
Sbjct: 240 TREVLNQGLFQIIENGTYNAIYEKWFGEKNPPPLPLVAPSLVGKVGTAQSLTERSQANPN 299

Query: 275 PNFFEQMATILKNNGSQFLRGTATTLLISMVGTIVGLFIGLLIGVFRTAPKSDNKLKAAL 334
            NF   + T+ +N       +G+  T+L++       GL  G  + +   A  SD
Sbjct: 300 DNF---LITLFRN----LFKGSILTVLLTAFSVFFGLIGGTGVAI---ALISD------- 342

Query: 335 QKLLGWLLNIYIEVFRGTPMIVQSMVIYYGTAQAF-----GVSLDRTLAAIFIVSINTGA 389
            K L  +  IY+E FRGTPM+VQ  +IY+G    F     G+++DR  AAI  +S+N A
Sbjct: 343 IKPLQLIFRIYVEFFRGTPMLVQLFIIYFGLPALFKEIGLGITIDRFPAAIIALSLNVAA 402

Query: 390 YMSEIVRGGIFSVDKGQFEAATALGFTHGQTMRKIVLPQVVRNILPATGNEFVINIKDTS 449
           Y++EI+RGGI S+D+GQ+EA  +LG +  QTM++++ PQ  R ILP  GNEF+  IKDTS
Sbjct: 403 YLAEIIRGGIQSIDQGQWEACESLGMSPWQTMKEVIFPQAFRRILPPLGNEFITLIKDTS 462

Query: 450 VLNVISVVELYFSGNTVATQTYQYFQTFTIIAIIYFILTFTVTRILRYIEKRFD 503
           +  VI   EL+  G  +    TY+ F+ +  +A++Y +LT   + + +++E   D
Sbjct: 463 LTAVIGFQELFREGQLIVATTYRAFEVYIAVALVYLLLTTISSFVFKWLENYMD 516
```

There is also homology to SEQ ID 1194.

A related GBS gene <SEQ ID 8923> and protein <SEQ ID 8924> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 2
McG: Discrim Score:       6.23
GvH: Signal Score (-7.5): 0.11
     Possible site: 24
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 3 value:  -8.86 threshold:  0.0
   INTEGRAL    Likelihood = -8.86   Transmembrane  301 - 317 ( 295 - 324)
   INTEGRAL    Likelihood = -6.05   Transmembrane  479 - 495 ( 473 - 496)
   PERIPHERAL  Likelihood =  1.32     441
 modified ALOM score:   2.27

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
34.3/57.3% over 462aa
                                                          Synechocystis PCC6803
  EGAD|48193| glutamine-binding periplasmic protein/glutamine transport system permease
protein Insert characterized
```

```
  GP|1652664|dbj|BAA17584.1||D90907 glutamine-binding periplasmic protein {Synechocystis
sp.} Insert characterized
  PIR|S77250|S77250 hypothetical protein - Synechocystis sp. (strain PCC 6803) Insert
characterized

ORF01242(454 - 1809 of 2148)
EGAD|48193|sll1270(54 - 516 of 530) glutamine-binding periplasmic protein/glutamine
transport system permease protein {Synechocystis PCC6803}GP|1652664|dbj|BAA17584.1||D90907
glutamine-binding periplasmic protein {Synechocystis sp.}PIR|S77250|S77250 hypothetical
protein - Synechocystis sp. (strain PCC 6803)
%Match = 12.3
%Identity = 34.2  %Similarity = 57.2
Matches = 128  Mismatches = 149  Conservative Sub.s = 86

204       234       264       294       324       354       384       414
PSFVCIPF*HKNTINRFQ*DNDIEIDLVFR*NRRK*LIGGC*MKKILLSLFTALLITXGGMTSIQADEYLRVGMEAAYAP

                                          MKGMVKLGHWGKTWRYYLLLALGVLLAIAIPLLPAFSQVS
                                                  10        20        30        40

444       474       495       525       555       585       615       645
FNWTQNDNTNGAVPIEGTDQ---YANGYDVQVAKKLAKKLNKKVVVVKTKWEGLVPALTSGKLDMIIAGMSPTEERKKEI
                 | | ||:      |:|| : : : :     | :    ::|::||| | :   |: :: | || : :
RQTIIVATEPTFPPFEMTDEATGQLTGFDVDLIQAIGEAAQVTVDIQGYPFDGIIPALQSNTVGAAISAITITPERAQSV
       50        60        70        80        90       100       110       120

675       705       735       765       795       825       855       885
NFSKPYYISEPTLVVNAEGKYTNAKNISDFKNAKVTAQQGVYLYNLIDQINGVKKEVAMGDFNQLRQAVESGVVDAYVSE
:|| ||: |   : |  :|  |  ||: |::  ::     |     :   : | |        : |:: | :|  || :::
SFSSPYFKSVLAIAVQ-DGNDT-IKNLKDLEGKRLAVAIGTTGAMVATNVPGAKVTNFDSITSALQELV-NGNADAVIND
       130       140       150       160       170       180       190

903               957       987
RP-----------------------------------------DATSAQTANPKLK-MIELHQG-FKTSDADTNISV
||                                          |    |||    :| | :  || |
RPVLLYAIKDAGLRNVKISADV~~~~NPPFLPLVAPSLVGKVGTAQSLTERSQANPNDNFLITLFRNLFKGS--------
          210             270       280       290       300       310

1017      1047      1077      1107      1137      1167      1197      1227
GMRKGDNRINQVNQVLESISRDKQIALMDKMIKEQPSVKKEKNGKPNFFEQMATILKNNGSQFLRGTATTLLISMVGTIV
                                                                      ::::: |
-------------------------------------------------------------------ILTVLLTAF
                                                                         320

1257      1284      1314      1344      1374      1404      1419      1449
GLFIGLLIGV-FRTAPKSDNKLKAALQKLLGWLLNIYIEVFRGTPMIVQSMVIYYGTAQAF-----GVSLDRTLAAIFIV
 :| ||: |      |  ||        | |  :: ||:| ||||||:||  :||:|    |     |:::||  |||  :
SVFFGLIGGTGVAIALISD-------IKPLQLIFRIYVEFFRGTPMLVQLFIIYFGLPALFKEIGLGITIDRFPAAIIAL
             340               350       360       370       380       390

1479      1509      1539      1569      1599      1629      1659      1689
SINTGAYMSEIVRGGIFSVDKGQFEAATALGFTHGQTMRKIVLPQVVRNILPATGNEFVINIKDTSVLNVISVVELYFSG
|:|  ||::||:|||| |:|:||:||  :|| : |||:::::||  | |||  ||||:  |||||:  ||   ||:  |
SLNVAAYLAEIIRGGIQSIDQGQWEACESLGMSPWQTMKEVIFPQAFRRILPPLGNEFITLIKDTSLTAVIGFQELFREG
           410       420       430       440       450       460       470

1719      1749      1779      1809      1839      1869      1899      1929
NTVATQTYQYFQTFTIIAIIYFILTFTVTRILRYIEKRFDSDNYTTGANQLQV*EVGMTQAILEIKHLKKSYGSNEVLKD
  :   ||: |: :  :|::|::||   : :::::|   |        :
QLIVATTYRAFEVYIAVALVYLLLTTISSFVFKWLENYMDPIGRAKKKAKAATA
           490       500       510       520       530
```

There is also homology to SEQ ID 5804.

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

Example 1867

A DNA sequence (GBSx1974) was identified in *S.agalactiae* <SEQ ID 5807> which encodes the amino acid sequence <SEQ ID 5808>. This protein is predicted to be ATP-binding. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3208(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:CAB73160 GB:AL139076 putative glutamine transport ATP-binding
          protein [Campylobacter jejuni]
 Identities = 132/241 (54%), Positives = 178/241 (73%), Gaps = 1/241 (0%)

Query: 5   ILEIKHLKKSYGSNEVLKDISLSVNKGEVISIIGSSGSGKSTFLRSINLLEEPSGGEILY 64
           ++E+K+L+K YG  EVLK+I+ +++KG+VI+IIG SG GKSTFLR IN LE    GEIL
Sbjct: 1   MIEVKNLQKKYGELEVLKNINTTISKGDVIAIIGPSGGGKSTFLRCINRLELADSGEILI 60

Query: 65  HGHNVLEKGYDLNNYREKLGMVFQSFNLFENLNILENAIVAQTTVLKRERQEAEKIAKEN 124
           +  N+L+K  D+N  R+K+ MVFQ FNLF N N++EN  +          ++EA K AK
Sbjct: 61  NKQNILDKEIDINKIRQKVSMVFQHFNLFANKNVMENLCLTPIKTGILSQEEAIKKAKLL 120

Query: 125 LNAVGMTEQYWKAKPKQLSGGQKQRVAIARALSVNPEAILFDEPTSALDPEMVGEVLKTM 184
           L  VG+ ++      P +LSGGQKQR+AIAR+L +NP+ ILFDEPTSALDPEM+GEVL  M
Sbjct: 121 LAKVGLADKE-NIMPHKLSGGQKQRIAIARSLMMNPDVILFDEPTSALDPEMIGEVLSIM 179

Query: 185 QDLAKSGLTMIIVTHEMEFAKEVSDRVIFMDKGIIAEQGTPKQLFENPTQERTKEFLQRFL 245
           +D+AK GLTM++VTHEM FA+ V++R+ FMDKG IA    +PK++FENP+ ER +EFL + L
Sbjct: 180 KDVAKEGLTMLVVTHEMGFARNVANRIFFMDKGKIAVDASPKEVFENPSNERLREFLNKVL 240
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2157> which encodes the amino acid sequence <SEQ ID 2158>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1170(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 = 212/246 (86%), Positives = 237/246 (96%)

Query: 1   MTQAILEIKHLKKSYGSNEVLKDISLSVNKGEVISIIGSSGSGKSTFLRSINLLEEPSGG 60
           M+ +I+EIK+LKKSYGSNEVLKDISLSVNKGEVISIIGSSGSGKST LRSINLLEEPS G
Sbjct: 24  MSNSIIEIKNLKKSYGSNEVLKDISLSVNKGEVISIIGSSGSGKSTLLRSINLLEEPSAG 83

Query: 61  EILYHGHNVLEKGYDLNNYREKLGMVFQSFNLFENLNILENAIVAQTTVLKRERQEAEKI 120
           +IL+HG +VL + Y+L +YREKLGMVFQSFNLFENLN+LENAIVAQTTVLKR+R +AE+I
Sbjct: 84  QILFHGEDVLAEHYNLTHYREKLGMVFQSFNLFENLNVLENAIVAQTTVLKRDRAQAEQI 143

Query: 121 AKENLNAVGMTEQYWKAKPKQLSGGQKQRVAIARALSVNPEAILFDEPTSALDPEMVGEV 180
           AKENLNAVGMTEQYW+AKPKQLSGGQKQRVAIARALSVNPEA+LFDEPTSALDPEMVGEV
Sbjct: 144 AKENLNAVGMTEQYWQAKPKQLSGGQKQRVAIARALSVNPEAMLFDEPTSALDPEMVGEV 203

Query: 181 LKTMQDLAKSGLTMIIVTHEMEFAKEVSDRVIFMDKGIIAEQGTPKQLFENPTQERTKEF 240
           LKTMQDLAKSGLTMIIVTHEMEFA++VSDR+IFMDKG+I E+G+P+Q+FENPTQ+RTKEF
Sbjct: 204 LKTMQDLAKSGLTMIIVTHEMEFARDVSDRIIFMDKGLITEEGSPQQIFENPTQDRTKEF 263
```

```
Query: 241 LQRFLK 246
           LQRFLK
Sbjct: 264 LQRFLK 269
```

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

Example 1868

A DNA sequence (GBSx1976) was identified in *S.agalactiae* <SEQ ID 5809> which encodes the amino acid sequence <SEQ ID 5810>. This protein is predicted to be hypersensitive-induced response protein. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-17.94   Transmembrane    4 -  20 (   1 -  28)

----- Final Results -----
               bacterial membrane --- Certainty=0.8175(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 9479> which encodes amino acid sequence <SEQ ID 9480> was also identified.

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

```
>GP:AAF68390 GB:AF236374 hypersensitive-induced response protein
           [Zea mays]
 Identities = 127/275 (46%), Positives = 174/275 (63%), Gaps = 1/275 (0%)

Query: 19  ITSLYVVKQQTVAIIERFGKYQKTATSGIHIRVPLGIDKIAARVQLRLLQSEIIVETKTK 78
           I  L  V Q TVAI E FGK+ +    G H        +IA  + LR+ Q ++  ETKTK
Sbjct: 4   ILGLVQVDQSTVAIKENFGKFSEVLEPGCHFLPWCIGQQIAGYLSLRVRQLDVRCETKTK 63

Query: 79  DNVFVTLNIATQYRVNENNVTDAYYKLIKPEAQIKSYIEDALRSSVPKLTLDELFEKKDE 138
           DNVFVT+  + QYR   +  +DA+YKL      QI+SY+ D +R++VPKL LD+ FE+K+E
Sbjct: 64  DNVFVTVVASVQYRALADKASDAFYKLSNTREQIQSYVFDVIRATVPKLGLDDAFEQKNE 123

Query: 139 IALEVQHQVAEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQRKRVAAQELANADKI 198
           IA  V+ ++ + MSTYGY IV+TLI  +EPD  VK++MNEINAA R RVAA E A A+KI
Sbjct: 124 IAKAVEEELEKAMSTYGYQIVQTLIVDIEPDDRVKRAMNEINAAARMRVAASEKAEAEKI 183

Query: 199 KIVTAAEAEAEKDRLHGVGIAQQRKAIVDGLADSIQELKDANVTLTEEQIMSILLTNQYL 258
             + AE EAE   L GVGIA+QR+AIVDGL DS+    +     T + IM ++L  QY
Sbjct: 184 LQIKKAEGEAESKYLAGVGIARQRQAIVDGLRDSVLAFSENVPGTTAKDIMDMVLVTQYF 243

Query: 259 DTLNTF-AINGNQTIFLPNNPEGVEDIRTQVLSAL 292
           DT+    A + + ++F+P+ P  V+D+  Q+   L
Sbjct: 244 DTMREIGASSKSSSVFIPHGPGAVKDVSAQIRDGL 278
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5811> which encodes the amino acid sequence <SEQ ID 5812>. Analysis of this protein sequence reveals the following:

```
     Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.06   Transmembrane    5 -  21 (   1 -  29)

----- 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:AAF68390 GB:AF236374 hypersensitive-induced response protein
          [Zea mays]
 Identities = 126/273 (46%), Positives = 174/273 (63%), Gaps = 3/273 (1%)

Query: 23  LYVVRQQSVAIVERFGRYQKTATSGIHIRLPFGI-DKIAARVQLRLLQSEIIVETKTKDN 81
           L  V Q +VAI E FG++ +    G H  LP+ I  +IA  + LR+ Q ++  ETKTKDN
Sbjct: 7   LVQVDQSTVAIKENFGKFSEVLEPGCHF-LPWCIGQQIAGYLSLRVRQLDVRCETKTKDN 65

Query: 82  VFVTLNVATQYRVNEQNVTDAYYKLMKPESQIKSYIEDALRSSVPKLTLDELFEKKDEIA 141
           VFVT+  + QYR      +DA+YKL      QI+SY+ D +R++VPKL LD+ FE+K+EIA
Sbjct: 66  VFVTVVASVQYRALADKASDAFYKLSNTREQIQSYVFDVIRATVPKLGLDDAFEQKNEIA 125

Query: 142 LEVQHQVAEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQRKRVAAQELANADKIKI 201
             V+ ++ + MSTYGY IV+TLI  +EPD  VK++MNEINAA R RVAA E A A+KI
Sbjct: 126 KAVEEELEKAMSTYGYQIVQTLIVDIEPDDRVKRAMNEINAAARMRVAASEKAEAEKILQ 185

Query: 202 VTAAEAEAEKDRLHGVGIAQQRKAIVDGLAESIQELKEANISLNEEQIMSILLTNQYLDT 261
           +  AE EAE   L GVGIA+QR+AIVDGL +S+    E       + IM ++L  QY DT
Sbjct: 186 IKKAEGEAESKYLAGVGIARQRQAIVDGLRDSVLAFSENVPGTTAKDIMDMVLVTQYFDT 245

Query: 262 LNTFAAKG-NQTLFLPNTPSGVEDIRTQVLSAL 293
           +    A   + ++F+P+ P  V+D+  Q+   L
Sbjct: 246 MREIGASSKSSSVFIPHGPGAVKDVSAQIRDGL 278
```

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

```
 Identities = 254/291 (87%), Positives = 278/291 (95%)

Query: 5   IILTVILVLVIVLLITSLYVVKQQTVAIIERFGKYQKTATSGIHIRVPLGIDKIAARVQL 64
           I +   +++++ ++ ++LYVV+QQ+VAI+ERFG+YQKTATSGIHIR+P GIDKIAARVQL
Sbjct: 6   IFIAFGVIVILAIVASTLYVVRQQSVAIVERFGRYQKTATSGIHIRLPFGIDKIAARVQL 65

Query: 65  RLLQSEIIVETKTKDNVFVTLNIATQYRVNENNVTDAYYKLIKPEAQIKSYIEDALRSSV 124
           RLLQSEIIVETKTKDNVFVTLN+ATQYRVNE NVTDAYYKL+KPE+QIKSYIEDALRSSV
Sbjct: 66  RLLQSEIIVETKTKDNVFVTLNVATQYRVNEQNVTDAYYKLMKPESQIKSYIEDALRSSV 125

Query: 125 PKLTLDELFEKKDEIALEVQHQVAEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQR 184
           PKLTLDELFEKKDEIALEVQHQVAEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQR
Sbjct: 126 PKLTLDELFEKKDEIALEVQHQVAEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQR 185

Query: 185 KRVAAQELANADKIKIVTAAEAEAEKDRLHGVGIAQQRKAIVDGLADSIQELKDANVTLT 244
           KRVAAQELANADKIKIVTAAEAEAEKDRLHGVGIAQQRKAIVDGLA+SIQELK+AN++L
Sbjct: 186 KRVAAQELANADKIKIVTAAEAEAEKDRLHGVGIAQQRKAIVDGLAESIQELKEANISLN 245

Query: 245 EEQIMSILLTNQYLDTLNTFAINGNQTIFLPNNPEGVEDIRTQVLSALKTR 295
           EEQIMSILLTNQYLDTLNTFA  GNQT+FLPN P GVEDIRTQVLSALKT+
Sbjct: 246 EEQIMSILLTNQYLDTLNTFAAKGNQTLFLPNTPSGVEDIRTQVLSALKTK 296
```

SEQ ID 5810 (GBS231) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 7; MW 60.9kDa).

GBS231d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 5-7; MW 59kDa) and in Figure 239 (lane 11; MW 59kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 9; MW 34kDa) and in Figure 183 (lane 6; MW 34kDa). Purified GBS231d-GST is shown in Figure 246, lane 8.

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

Example 1869

A DNA sequence (GBSx1977) was identified in *S.agalactiae* <SEQ ID 5813> which encodes the amino acid sequence <SEQ ID 5814>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2305(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 9291> which encodes amino acid sequence <SEQ ID 9292> was also identified.

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

```
>GP:CAB13457 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 259/514 (50%), Positives = 350/514 (67%), Gaps = 9/514 (1%)

Query: 1   MGMTMENGAKEVSDKPATTVGEVGQILSKGVLMGARGNSGVITSQLFRGFGQSIKDKEEL 60
           M ++M +GA+EV       +G+VG  LSKG+LMGARGNSGVI SQLFRGF ++I+ K+E+
Sbjct: 46  MNLSMTSGAREVEQMDTDDIGKVGSALSKGLLMGARGNSGVILSQLFRGFSKNIETKKEI 105

Query: 61  TGQDLAHAFQNGVEVAYKAVMKPVEGTILTVSRGAATAALKKAEETDDAVEVMRATLKGA 120
              + A A Q GV++AYKAVMKPVEGTILTV++ AA  A+  AE+  D   +M A  + A
Sbjct: 106 NALEFAAALQAGVDMAYKAVMKPVEGTILTVAKDAAKKAMILAEKETDITALMTAVTEEA 165

Query: 121 KRALAKTPDMLPVLKEVGVVDSGGQGLVFIYEGFLSALTGEYIASEDFKATPATMTEMVN 180
           + +L +TP++LPVLKEVGVVDSGG+GL+ +YEGFL++L GE +     KA   ++ +MV+
Sbjct: 166 EASLNRTPELLPVLKEVGVVDSGGKGLLCVYEGFLASLKGETVPQ---KAVLPSLDDMVS 222

Query: 181 AEHHKAVVGHVATEDIKYGYCTEVMVGLKQGPTYVKEFNYEEFQQGYLSNLGDSLLVVNDD 240
           AEHHK+    + TEDI++G+CTEVMV L Q     +EF+   F+  LS  GDSLLV+ D+
Sbjct: 223 AEHHKSAQSMMNTEDIEFGFCTEVMVRLDQTK---REFDEGTFRQDLSQFGDSLLVIADE 279

Query: 241 EIVKVHVHTEDPGLVMQEGLKYGSLVKVKVENMRNQHDA---QMQKVEVEETVKETKEYG 297
            + KVH+H E+PG V+     YG L+K+K+ENMR QH +   Q  K    ET    + YG
Sbjct: 280 SLAKVHIHAEEPGNVLNYAQHYGELIKIKIENMREQHTSIISQESKPADNETPPAKQPYG 339

Query: 298 IIAVVAGDGLAEIFKSQGVDYIISGGQTMNPSTEDIVKAIEKVNARNVIILPNNKNIFMA 357
           I+ V  G+G+A++FKS G   +I GGQTMNPSTEDIV A++ VNA  V ILPNN NI MA
Sbjct: 340 IVTVAMGEGIADLFKSIGASVVIEGGQTMNPSTEDIVDAVKSVNADTVFILPNNSNIIMA 399

Query: 358 AQSAADVVDIPAAVVETRTVPQGFTSLLAFDPAKSLETNVADMTNSLSDVISGSVTLAVR 417
           A  AA VVD    V+  +TVPQG ++LLAF+P +  E N A+M +++  V SG VT +VR
Sbjct: 400 ANQAASVVDEQVFVIPAKTVPQGMSALLAFNPDQEAEANEANMLSAIQQVKSGQVTFSVR 459

Query: 418 DTTIDGLEIHENDILGMVDGKILVSTPDMEKALKDTFDKMIDEDSEIVTIYVGEDGKQAL 477
           DT IDG +I + D +G+++G I+ ++ +   A K    +MI ED EIVTI  GED  Q
Sbjct: 460 DTHIDGKDIKKGDFMGILNGTIIGTSENQLSAAKMLLSEMIGEDDEIVTILYGEDASQEE 519

Query: 478 AETLSEYLEETYEDVEVEIHQGDQPVYPYLMSVE 511
           AE L  +L E YE++EVEIH G QP+Y Y++S E
Sbjct: 520 AEQLEAFLSEKYEEIEVEIHNGKQPLYSYIVSAE 553
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5635> which encodes the amino acid sequence <SEQ ID 5636>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1816(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 = 434/511 (84%), Positives = 475/511 (92%)

Query: 1   MGMTMENGAKEVSDKPATTVGEVGQILSKGVLMGARGNSGVITSQLFRGFGQSIKDKEEL 60
           M MTM+NGAKEV+DKPA+TVGEVGQ+LSKG+LMGARGNSGVITSQLFRGFGQSIK K+EL
Sbjct: 44  MSMTMDNGAKEVADKPASTVGEVGQMLSKGLLMGARGNSGVITSQLFRGFGQSIKGKDEL 103

Query: 61  TGQDLAHAFQNGVEVAYKAVMKPVEGTILTVSRGAATAALKKAEETDDAVEVMRATLKGA 120
           TG+DLA AFQ GVEVAYKAVMKPVEGTILTVSRGAATAALKKA+ TDDAVEVM+A L GA
Sbjct: 104 TGKDLAQAFQVGVEVAYKAVMKPVEGTILTVSRGAATAALKKADLTDDAVEVMQAALDGA 163

Query: 121 KRALAKTPDMLPVLKEVGVVDSGGQGLVFIYEGFLSALTGEYIASEDFKATPATMTEMVN 180
           K ALAKTPD+LPVLKEVGVVDSGGQGLVFIYEGFLSAL G+Y+ S DFKATPA M+EM+N
Sbjct: 164 KGALAKTPDLLPVLKEVGVVDSGGQGLVFIYEGFLSALNGDYVTSADFKATPANMSEMIN 223

Query: 181 AEHHKAVVGHVATEDIKYGYCTEVMVGLKQGPTYVKEFNYEEFQGYLSNLGDSLLVVNDD 240
           AEHHK+VVGHVATEDI YGYCTE+MV LKQGPTYVKEFNY+EFQGYLS LGDSLLVVNDD
Sbjct: 224 AEHHKSVVGHVATEDITYGYCTEIMVALKQGPTYVKEFNYDEFQGYLSGLGDSLLVVNDD 283

Query: 241 EIVKVHVHTEDPGLVMQEGLKYGSLVKVKVENMRNQHDAQMQKVEVEETVKETKEYGIIA 300
           EIVKVHVHTEDPGLVMQEGLKYGSL+K+KV+NMRNQH+AQ+QK +VE+   E K++G+IA
Sbjct: 284 EIVKVHVHTEDPGLVMQEGLKYGSLIKIKVDNMRNQHEAQVQKTDVEKNKAEVKDFGLIA 343

Query: 301 VVAGDGLAEIFKSQGVDYIISGGQTMNPSTEDIVKAIEKVNARNVIILPNNKNIFMAAQS 360
           VVAG+GL+EIFK+QGVDY+ISGGQTMNPSTEDIVKAIE VNA+ VIILPNNKNIFMAAQS
Sbjct: 344 VVAGEGLSEIFKAQGVDYVISGGQTMNPSTEDIVKAIEAVNAKQVIILPNNKNIFMAAQS 403

Query: 361 AADVVDIPAAVVETRTVPQGFTSLLAFDPAKSLETNVADMTNSLSDVISGSVTLAVRDTT 420
           AA+VVDIPAAVV TRTVPQGFTSLLAFDP+KSLE NVADM+ SLSDV+SGSVTLAVRDTT
Sbjct: 404 AAEVVDIPAAVVATRTVPQGFTSLLAFDPSKSLEDNVADMSTSLSDVVSGSVTLAVRDTT 463

Query: 421 IDGLEIHENDILGMVDGKILVSTPDMEKALKDTFDKMIDEDSEIVTIYVGEDGKQALAET 480
           IDGLEIHEND LGMVDGKI+VS PDME  LK  F+KMIDEDSEIVTI+VGE+G Q LAE
Sbjct: 464 IDGLEIHENDFLGMVDGKIIVSNPDMEATLKAAFEKMIDEDSEIVTIFVGEEGDQDLAEE 523

Query: 481 LSEYLEETYEDVEVEIHQGDQPVYPYLMSVE 511
           L+ YL ETYEDVEVEIHQGDQPVYPYLMSVE
Sbjct: 524 LAGYLGETYEDVEVEIHQGDQPVYPYLMSVE 554
```

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

Example 1870

A DNA sequence (GBSx1978) was identified in *S.agalactiae* <SEQ ID 5815> which encodes the amino acid sequence <SEQ ID 5816>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4771(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 1871

A DNA sequence (GBSx1979) was identified in *S.agalactiae* <SEQ ID 5817> which encodes the amino acid sequence <SEQ ID 5818>. This protein is predicted to be proliferating-cell nucleolar antigen P120. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3774(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 9345> which encodes amino acid sequence <SEQ ID 9346> was also identified.

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

```
>GP:AAC74905 GB:AE000278 putative nucleolar proteins [Escherichia
          coli K12]
 Identities = 87/229 (37%), Positives = 128/229 (54%), Gaps = 8/229 (3%)

Query: 63  GKSIEHTTGLVYSQEPAAQ--IVAQIAEPQEGMKVLDLAAAPGGKTTHLLSYLNNTGLLV 120
           G + EH +GL Y QE ++   + A  A+      +V+D+AAAPG KTT + + +NN G ++
Sbjct: 89  GSTAEHLSGLFYIQEASSMLPVAALFADGNAPQRVMDVAAAPGSKTTQISARMNNEGAIL 148

Query: 121 SNEISNKRSKILVENVERFGARNVIVTNESSQRLAKCFNSFFDLIVFDGPCSGEGMFRKD 180
           +NE S  R K+L  N+ R G  NV +T+   +          FD I+ D PCSGEG+ RKD
Sbjct: 149 ANEFSASRVKVLHANISRCGISNVALTHFDGRVFGAAVPEMFDAILLDAPCSGEGVVRKD 208

Query: 181 PQAIQYWHKDYPTECAQLQRDILKEAIKMLAHGGILVYSTCTWSPEENEEVVNWLLQEY- 239
           P A++ W  +   E A  QR+++  A   L  GG LVYSTCT + EENE V  WL + Y
Sbjct: 209 PDALKNWSPESNQEIAATQRELIDSAFHALRPGGTLVYSTCTLNQEENEAVCLWLKETYP 268

Query: 240 ---DYLELVDIPKLNGMVEGINVPQVARMYPHHFQGEGQFVAKLRDTRS 285
              ++L L D+    G  + +        ++P  +  EG FVA+LR T++
Sbjct: 269 DAVEFLPLGDL--FPGANKALTEEGFLHVFPQIYDCEGFFVARLRKTQA 315
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5819> which encodes the amino acid sequence <SEQ ID 5820>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2316(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 = 213/311 (68%), Positives = 254/311 (81%), Gaps = 3/311 (0%)

Query: 1   MKLPNEFIEKYQTILKDEAEAFFDSFEQKPISAYRTNPLKEKQLDFPNAIPSTPWGHYGK 60
           M LP EFI  YQ IL  E E F  SF Q+P++A+R NPLK +   F + IP+T WG+YGK
Sbjct: 2   MSLPKEFINTYQAILGKELEDFLASFNQEPVNAFRINPLKNQLKTFEHPIPNTLWGYYGK 61

Query: 61  ISGKSIEHTTGLVYSQEPAAQIVAQIAEPQEGMKVLDLAAAPGGKTTHLLSYLNNTGLLV 120
           +SGKS EH +GLVYSQEPAAQ+VAQ+A PQ+G +VLDLAAAPGGK+THLL+YL+NTGLLV
Sbjct: 62  LSGKSPEHVSGLVYSQEPAAQMVAQVAAPQKGSRVLDLAAAPGGKSTHLLAYLDNTGLLV 121

Query: 121 SNEISNKRSKILVENVERFGARNVIVTNESSQRLAKCFNSFFDLIVFDGPCSGEGMFRKD 180
           SNEIS KRSK+LVEN+ERFGARNV+VTNES+ RLAK F+ +FD IVFDGPCSGEGMFRKD
Sbjct: 122 SNEISKKRSKVLVENIERFGARNVVVTNESADRLAKVFSHYFDTIVFDGPCSGEGMFRKD 181
```

```
Query: 181 PQAIQYWHKDYPTECAQLQRDILKEAIKMLAHGGILVYSTCTWSPEENEEVVNWLLQEYD 240
           P AIQYWH  YP ECA+LQ+ IL++A+ ML  GG L+YSTCTW+PEENE+VV WLL+ Y
Sbjct: 182 PDAIQYWHHGYPAECAKLQKSILEDALAMLKPGGELIYSTCTWAPEENEDVVQWLLETYT 241

Query: 241 YLELVDIPKLNGMVEGINVPQVARMYPHHFQGEGQFVAKLRDTRSKEAQKIKPKAQKIN- 299
           +LELVD+PKLNGMV GI +P+ ARMYPH +QGEGQFVAKL+D R +E Q  K KA K N
Sbjct: 242 FLELVDVPKLNGMVSGIGLPETARMYPHRYQGEGQFVAKLKDKR-QEGQSTKLKAPKSNL 300

Query: 300 -KMQLQLWQQF 309
            K QL+LW+ F
Sbjct: 301 IKDQLRLWKMF 311
```

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

Example 1872

A DNA sequence (GBSx1980) was identified in *S.agalactiae* <SEQ ID 5821> which encodes the amino acid sequence <SEQ ID 5822>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4111(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:AAC24940 GB:AF012285 unknown [Bacillus subtilis]
 Identities = 86/240 (35%), Positives = 133/240 (54%), Gaps = 10/240 (4%)

Query: 6   DFAKQLVYKAGQFIKSEMQNTFDVEEKSRFDDLVTSLDKKTQKLLIQEIIQHYPDDNILA 65
           + AK+ + +AG  I   M  +  +E KS  +DLVT++DK+T+K  I  I + +P   IL
Sbjct: 9   EIAKKWIREAGARITQSMHESLTIETKSNPNDLVTNIDKETEKFFIDRIQETFPGHRILG 68

Query: 66  EE---DBVRSPIAQGNVWVLDPIDGTVNFIVQKDNFAVMLAYYEEGVGQFGIIYDVMADI 122
           EE   D + S   +G VW++DPIDGT+NF+ Q+ NFA+ +  +E G G+ G+IYDV+ D
Sbjct: 69  EEGQGDKIHS--LEGVVWIIDPIDGTMNFVHQQRNFAISIGIFENGEGKIGLIYDVVHDE 126

Query: 123 LYSGGGHFDVYANDKKIVPFQECPLERCLLGVNSAMYAEN----DCGIAHLASETLGVRI 178
           LY        Y N+ K+ P +E  +E  +L +N+    EN        +A L     G R
Sbjct: 127 LYHAFSGRGAYMNETKLAPLKETVIEEAILAINATWVTENRRIDQSVLAPLVKRVRGTRS 186

Query: 179 YGGAGISMAKVMQGKLLAYFSY-IQPWDYAAAKIMGETLGFTLLTLDGEEPNYSTRQKVM 237
           YG A + +A V  G++ AY +  + PWDYAA  ++   +G T  T++GE   +     V+
Sbjct: 187 YGSAALELANVAAGRIDAYITMRLAPWDYAAGCVLLNEVGGTYTTIEGEPFTFLENHSVL 246
```

A related GBS nucleic acid sequence <SEQ ID 10937> which encodes amino acid sequence <SEQ ID 10938> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5823> which encodes the amino acid sequence <SEQ ID 5824>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1843(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 = 155/253 (61%), Positives = 205/253 (80%)

Query: 1   MDAKFDFAKQLVYKAGQFIKSEMQNTFDVEEKSRFDDLVTSLDKKTQKLLIQEIIQHYPD 60
           ++ K+ FA+Q++ +AG FIKS+M    D++ K++FDDLVT++D++TQ+LL+  I Q YP
Sbjct: 8   LETKYAFARQIIKEAGLFIKSKMSEQLDIQVKTQFDDLVTNVDQETQQLLMDRIHQTYPC 67

Query: 61  DNILAEEDBVRSPIAQGNVWVLDPIDGTVNFIVQKDNFAVMLAYYEEGVGQFGIIYDVMA 120
           D ILAEE++VR PI QGNVWV+DPIDGTVNFIVQ   FAVM+AYYE+G+GQFG+IYDVMA
Sbjct: 68  DAILAEENDVRHPINQGNVWVIDPIDGTVNFIVQGSQFAVMIAYYEQGIGQFGLIYDVMA 127

Query: 121 DILYSGGGHFDVYANDKKIVPFQECPLERCLLGVNSAMYAENDCGIAHLASETLGVRIYG 180
           D L +GGG F+V  N  K+  +QE PLER L+G N+ M+A ND  +AHL ++TLGVR+YG
Sbjct: 128 DQLLAGGGDFEVTLNGDKLPAYQEKPLERSLIGCNAGMFARNDRNLAHLIAKTLGVRVYG 187

Query: 181 GAGISMAKVMQGKLLAYFSYIQPWDYAAAKIMGETLGFTLLTLDGEEPNYSTRQKVMFLP 240
           GAGI M KVM+ +LLAYFS+IQPWDYAAAK++G+ LG+ LLT+DG EP++ TRQK+MF+P
Sbjct: 188 GAGICMVKVMKQELLAYFSFIQPWDYAAAKVLGDKLGYVLLTIDGYEPDFQTRQKIMFVP 247

Query: 241 KSKLNLIQSYLTK 253
           K +L  I S+LTK
Sbjct: 248 KCQLTRIASFLTK 260
```

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

Example 1873

A DNA sequence (GBSx1981) was identified in *S.agalactiae* <SEQ ID 5825> which encodes the amino acid sequence <SEQ ID 5826>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4131(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:AAC24938 GB:AF012285 unknown [Bacillus subtilis]
 Identities = 33/78 (42%), Positives = 50/78 (63%)

Query: 13 YSYPLDPSWNTEDITKVLRFLNQVEHAYENSIKVDDLLDSYKEFKKVVKSKAQEKQIDRE 72
          Y YP++  W TE+   V+ F  QVE AYE     ++LL +Y+ FK++V  KA+EK++  E
Sbjct: 3  YQYPMNEDWTTEEAVDVIAFFQQVELAYEKGADREELLKAYRRFKEIVPGKAEEKKLCGE 62

Query: 73 FQRTSGYSTYQAVKAAQQ 90
          F+  S YS Y+ VK A++
Sbjct: 63 FEEQSTYSPYRTVKQARE 80
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5827> which encodes the amino acid sequence <SEQ ID 5828>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4442(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 = 59/91 (64%), Positives = 70/91 (76%)

Query: 9   ISSNYSYPLDPSWNTEDITKVLRFLNQVEHAYENSIKVDDLLDSYKEFKKVVKSKAQEKQ 68
           +S NY YPLD SW+TE+I+ VL FLN+VE AYE  +    LLDSYK +K +VKSKAQEKQ
Sbjct: 5   MSGNYYYPLDLSWSTEEISSVLHFLNKVELAYEKKVDAKQLLDSYKTYKTIVKSKAQEKQ 64

Query: 69  IDREFQRTSGYSTYQAVKAAQQQAKGFISLG 99
           IDR+FQ+ SGYSTYQ VK A+   KGF SLG
Sbjct: 65  IDRDFQKVSGYSTYQVVKKAKAIEKGFFSLG 95
```

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

Example 1874

A DNA sequence (GBSx1982) was identified in *S.agalactiae* <SEQ ID 5829> which encodes the amino acid sequence <SEQ ID 5830>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence (or aa 1-18)

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0952(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:AAF21893 GB:AF103794 unknown [Listeria monocytogenes]
 Identities = 74/126 (58%), Positives = 101/126 (79%)

Query: 1   MITLFLSPSCTSCRKARAWLSKHEVAFEEHNIITSPLNKEELLQILSFTENGTEDIISTR 60
           M+TL+ SPSCTSCRK+RAWL +H++ ++E NI + PL+ +E+ +IL  TE+GT++IISTR
Sbjct: 1   MVTLYTSPSCTSCRKSRAWLEEHDIPYKERNIFSEPLSLDEIKEILRMTEDGTDEIISTR 60

Query: 61  SKVFQKLAIDVDELSTSSLMELISENPSLLRRPIILDKKRMQIGFNEDEIRAFLPRDYRK 120
           SK FQKL +D+D L    L ELI +NP LLRRPII+D+KR+Q+G+NEDEIR FLPR  R
Sbjct: 61  SKTFQKLNVDLDSLPLQQLFELIQKNPGLLRRPIIIDEKRLQVGYNEDEIRRFLPRRVRT 120

Query: 121 QELKQA 126
            +L++A
Sbjct: 121 YQLREA 126
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5831> which encodes the amino acid sequence <SEQ ID 5832>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0511(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 = 112/134 (83%), Positives = 127/134 (94%)

Query: 1   MITLFLSPSCTSCRKARAWLSKHEVAFEEHNIITSPLNKEELLQILSFTENGTEDIISTR 60
           M+TLFLSPSCTSCRKARAWL KHEV F+EHNIITSPL+++EL+ ILSFTENGTEDIISTR
Sbjct: 1   MVTLFLSPSCTSCRKARAWLVKHEVDFQEHNIITSPLSRDELMSILSFTENGTEDIISTR 60

Query: 61  SKVFQKLAIDVDELSTSSLMELISENPSLLRRPIILDKKRMQIGFNEDEIRAFLPRDYRK 120
           SKVFQKL IDV+ELS S L++LI++NPSLLRRPII+D+KRMQIGFNEDEIRAFL RDYRK
Sbjct: 61  SKVFQKLDIDVEELSISDLIDLIAKNPSLLRRPIIMDQKRMQIGFNEDEIRAFLSRDYRK 120
```

```
Query: 121 QELKQATIRAEIEG 134
           QEL+QATI+AEIEG
Sbjct: 121 QELRQATIKAEIEG 134
```

SEQ ID 5830 (GBS232) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 10; MW 16.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 2; MW 42kDa).

GBS232-GST was purified as shown in Figure 207, lane 7.

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

Example 1875

A DNA sequence (GBSx1983) was identified in *S.agalactiae* <SEQ ID 5833> which encodes the amino acid sequence <SEQ ID 5834>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> 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 DNA sequence was identified in *S.pyogenes* <SEQ ID 5835> which encodes the amino acid sequence <SEQ ID 5836>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1768(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 = 210/308 (68%), Positives = 252/308 (81%)

Query: 1   MKIHYINDYKDIQAKEDCVLVLGYFDGLHLGHKALFDKAKKIATEKNLKIVVLTFNETPR 60
           M+I YI DY+DI  ++D VL+LGYFDGLH GHKALFDKA+++A ++ LK+VV TF E+P+
Sbjct: 1   MEIEYIKDYRDINQEDDTVLILGYFDGLHRGHKALFDKAREVANKEGLKVVVFTFTESPK 60

Query: 61  LTFARFQPELLLHLTSPEKRSEKFQEYGVDELYLMNFTSHFSKVSSDLFIKKYIYGLRAK 120
           L F+RF PELLLH+T P+KR EKF +YGV++LYL++FTS FSKVSSD FI  YI  L+AK
Sbjct: 61  LAFSRFSPELLLHITYPKKRYEKFADYGVNKLYLVDFTSKFSKVSSDHFITHYIKNLKAK 120

Query: 121 AAVVGFDYKFGHNRTSGDYLARNFKGPVYIIDEISEGGEKISSTRIRQLITEGNVEKANQ 180
             VVGFDYKFGHNRT  DYL RNF+G VY I+EI E   KIS+T IR+LI EGNV KAN
Sbjct: 121 HIVVGFDYKFGHNRTDSDYLTRNFEGQVYTIEEIKEDHRKISATWIRKLIQEGNVVKANH 180

Query: 181 LLGYEFSTCGMVVHGDARGRTIGFPTANLAPINRTYLPADGVYISNVLINGKYYRAMTSI 240
           LLGY+ ST G VVHGDARGRTIGFPTANLAPI+ TYLPADGVY++NV++  K YR+MTS+
Sbjct: 181 LLGYDLSTRGRVVHGDARGRTIGFPTANLAPIDNTYLPADGVYVTNVIVANKIYRSMTSL 240

Query: 241 GKNITFGGTELRLEANIFDFDGDIYGETIEIFWLKRIREMVKFNGIDDLVKQLKKDKEIA 300
           GKN+TFGG ELRLE NIFDFD +IYGE IEI WL +IR+M KF GI+DL  +L+ DK  A
Sbjct: 241 GKNVTFGGKELRLEVNIFDFDEEIYGEIIEIVWLDKIRDMEKFEGIEDLTDRLEYDKRTA 300

Query: 301 LNWKKDSQ 308
```

```
          LNWKKDS+
Sbjct: 301 LNWKKDSK 308
```

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

Example 1876

A DNA sequence (GBSx1984) was identified in *S.agalactiae* <SEQ ID 5837> which encodes the amino acid sequence <SEQ ID 5838>. This protein is predicted to be tRNA pseudouridine 5S synthase (truB). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2576(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 9817> which encodes amino acid sequence <SEQ ID 9818> was also identified.

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

```
>GP:BAB06129 GB:AP001515 tRNA pseudouridine 5S synthase [Bacillus halodurans]
 Identities = 145/283 (51%), Positives = 191/283 (67%), Gaps = 12/283 (4%)

Query: 2   ITGIINLKKEAGMTSHDAVFKLRKILHTKKIGHGGTLDPDVVGVLPIAVGKATRVIEYMT 61
           +TGI+ L K  GMTSHD V KLR++L TKK+GH GTLDPDV GVLP+ +G AT+V +YM+
Sbjct: 3   MTGILPLAKPRGMTSHDCVAKLRRLLKTKKVGHTGTLDPDVYGVLPVCIGHATKVAQYMS 62

Query: 62  ESGKIYEGEITLGYATSTEDSSGEVISRTPLTQSDLSEDVVDHAMKSFTGPITQVPPMYS 121
           +  K YEGE+T+G++T+TED SG+ +  T   Q    E VVD  + +F G I Q+PPMYS
Sbjct: 63  DYPKAYEGEVTVGFSTTTEDRSGDTVE-TKTIQQPFVEAVVDQVLATFVGEIKQIPPMYS 121

Query: 122 AVKVNGKKLYEYARSGEEVERPKRQITISEFRRTSPLYFEKGICRFSFYVSCSKGTYVRT 181
           AVKV GK+LYEYAR+G  VERP+R +TI    R S + +E+G+CRF F VSCSKGTYVRT
Sbjct: 122 AVKVRGKRLYEYARAGITVERPERTVTIFSLERMSDIVYEEGVCRFRFNVSCSKGTYVRT 181

Query: 182 LAVDLGIKLGYASHMSFLKRTSSAGLSITQSLTLEEINEKYKQ-EDFSFLLPIEYGVLDL 240
           LAVD+G  LGY +HMS L RT S   S+ +  T  E+ E+ +Q E  S LLPIE  +LD+
Sbjct: 182 LAVDIGKALGYPAHMSDLVRTKSGPFSLEECFTFTELEERLEQGEGSSLLLPIETAILDI 241

Query: 241 PKVNLTEEDKVEISYGR----------RILLENEADTLAAFYE 273
           P+V + +E + +I +G           R  + NE   L A Y+
Sbjct: 242 PRVQVNKEIEEKIRHGAVLPQKWFNHPRFTVYNEEGALLAIYK 284
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5839> which encodes the amino acid sequence <SEQ ID 5840>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2698(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 = 201/295 (68%), Positives = 246/295 (83%), Gaps = 2/295 (0%)
```

```
Query: 1   MITGIINLKKEAGMTSHDAVFKLRKILHTKKIGHGGTLDPDVVGVLPIAVGKATRVIEYM 60
           MI GIINLKKEAGMTSHDAVFKLRK+L  KKIGHGGTLDPDVVGVLPIAVGKATRVIEYM
Sbjct: 1   MINGIINLKKEAGMTSHDAVFKLRKLLQEKKIGHGGTLDPDVVGVLPIAVGKATRVIEYM 60

Query: 61  TESGKIYEGEITLGYATSTEDSSGEVISRTPLTQSDLSEDVVDHAMKSFTGPITQVPPMY 120
           TE+GK+YEG++TLGY+T+TED+SGEV++R+ L  + L+E++VD  M +F G ITQ PPMY
Sbjct: 61  TEAGKVYEGQVTLGYSTTTEDASGEVVARSSL-PAVLTEELVDQTMTTFLGKITQTPPMY 119

Query: 121 SAVKVNGKKLYEYARSGEEVERPKRQITISEFRRTSPLYF-EKGICRFSFYVSCSKGTYV 179
           SAVKVNG+KLYEYAR+GE VERP+R++TIS F RTSPL F E G+CRFSF V+CSKGTYV
Sbjct: 120 SAVKVNGRKLYEYARAGESVERPRREVTISLFERTSPLNFTEDGLCRFSFKVACSKGTYV 179

Query: 180 RTLAVDLGIKLGYASHMSFLKRTSSAGLSITQSLTLEEINEKYKQEDFSFLLPIEYGVLD 239
           RTLAVDLG  LG  SHMSFL+R++SAGL++  + TL EI +   +++ SFLLPIEYGV D
Sbjct: 180 RTLAVDLGRALGVESHMSFLQRSASAGLTLETAYTLGEIADMVSKQEMSFLLPIEYGVAD 239

Query: 240 LPKVNLTEEDKVEISYGRRILLENEADTLAAFYENRVIAILEKRGNEFKPHKVLL 294
           LPK+ + + +  EIS+GRR+ L ++   LAAF+  +VIAILEKR  E+KP KVL+
Sbjct: 240 LPKMVIDDTELTEISFGRRLSLPSQEPLLAAFHGEKVIAILEKRDQEYKPKKVLI 294
```

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

Example 1877

A DNA sequence (GBSx1985) was identified in *S.agalactiae* <SEQ ID 5841> which encodes the amino acid sequence <SEQ ID 5842>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2776(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 9819> which encodes amino acid sequence <SEQ ID 9820> was also identified.

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

```
>GP:CAB12871 GB:Z99109 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 39/145 (26%), Positives = 68/145 (46%), Gaps = 7/145 (4%)

Query: 3   MKIRTATLDDSEKLVPLYQELG----YAISLSEIQSILKVILTHSDYGFLIAEDNGKLLA 58
           M IR A   D+  + PL+ +       A  L   ++ LK  L + +   LIAE+NG+ +
Sbjct: 1   MNIRQAKTSDAAAIAPLFNQYREFYRQASDLQGAEAFLKARLENHESVILIAEENGEFIG 60

Query: 59  FVGYHKLYFFEKSGTYYRILALVVNEKHRRKGIASQLINHVKQLAKTDGSEVLALNSSLK 118
           F   +  +        Y +  L V    R KG   +L++  K  A  +G++ L L +  +
Sbjct: 61  FTQLYPTFSSVSMKRIYILNDLFVVPHARTKGAGGRLLSAAKDYAGQNGAKCLTLQT--E 118

Query: 119 EYRQEAYHFYENLGFKKVSTGFSYY 143
            + ++A   YE  G+++  TGF +Y
Sbjct: 119 HHNRKARSLYEQNGYEE-DTGFVHY 142
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5843> which encodes the amino acid sequence <SEQ ID 5844>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0962(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 = 37/126 (29%), Positives = 64/126 (50%), Gaps = 16/126 (12%)

Query: 18  PLYQE-----LGYAISLSEIQSILKVILTHSDYGFLIA--EDNGKLLAFVG---YHKLYF 67
           P+ QE     LGY +SL  ++   + ++    + FL    +D  +LL +V    Y  LY
Sbjct: 11  PMLQEINAKALGYLVSLDLLERQYERLIEDCHHYFLAYADKDTNQLLGYVHAERYETLY- 69

Query: 68  FEKSGTYYRILALVVNEKHRRKGIASQLINHVKQLAKTDGSEVLALNSSLKEYRQEAYHF 127
              +     +L L V   ++R+GI S L+  ++  A+ +G   + LNS+   +R+EA+ F
Sbjct: 70  ---ASDGLNLLGLAVLPAYQRRGIGSALLRALESQARQEGIAFIRLNSA--SHRKEAHAF 124

Query: 128 YENLGF 133
           Y NL +
Sbjct: 125 YRNLDY 130
```

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

Example 1878

A DNA sequence (GBSx1986) was identified in *S.agalactiae* <SEQ ID 5845> which encodes the amino acid sequence <SEQ ID 5846>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1659(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif 28-30
```

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

```
>GP:AAF30776 GB:AE002133 conserved hypothetical [Ureaplasma
          urealyticum]
 Identities = 106/440 (24%), Positives = 206/440 (46%), Gaps = 65/440 (14%)

Query: 13  FAINESEYHQLLEQIRGDAFDKEVSERLEKERLILGEQAKNQLQEVVVE-KDKEIAKLQY 71
           F  N+ +Y++L++Q     +D      LEK+R  L E+ KN+  + +    KD +  K
Sbjct: 71  FLANDRDYNELVKQ----RYD------LEKQRDELKEKLKNEGNKAIAHFKDSDEYKNLI 120

Query: 72  KVKQFLIEKDNLLKDNEYQLAEQLNQKDMMLRD---------LENQIDRLRLEHENSLQEA 123
           K ++ +   +  ++ NE    +++   ++ L+          L+N I +  ++ +N+ + A
Sbjct: 121 KAQEKINSLNKTIESNEQSYKKEIENIELKLKSQFDEETKSLKNTIAKQEIKLDNAEKMA 180

Query: 124 LTKVERE-------RDAIQNQLHIQ--------------------EKEKDLALASVKSDY 156
           +   +          +D I   + I+                    E +K + +  ++S
Sbjct: 181 IINFKESNEYQKIIKDKIDLDIEIEKLKFAIQAHEDNMKAAKENWESKKIVEIKELESKK 240

Query: 157 EVQLKAANEQVEFYKNFKAQQSTKAVGESLEHYAETEFNKVRHLAFPNAYFEKDNTLSSR 216
           + ++    E +E  K  K+  + K VGE LE + + +F++    + P+  F K N
Sbjct: 241 DKEIHKLTESIEQLKREKSS-NVKLVGEELEQWLKNKFDETYSFSCPDMTFTKINEAID- 298

Query: 217 GSKGDFIY------REKDENDLEFL-SIMFEMKNESDDTIKKHKNEDFFKELDKDRREKS 269
           G K DF+        +E   +D + + S   E K E  D  K  KN    +K+LD+DR  +
Sbjct: 299 GKKADFLLEFFDFGKEMSNDDKKLIFSATIEAKTEFFDNQKGTKNSAHYKKLDQDRINQK 358

Query: 270 CEYAVLVTMLEADNDYYNTGIVDVSHKYPKMYVIRPQFFIQLIGILRNAALNTLKYKQEL 329
            EYA+LVT LE ++ +    ++   ++Y  M+ +RPQ+FI L+ ++RN A  TLK K
Sbjct: 359 SEYAILVTELEPEDHF----VIKKINEYKNMFAVRPQYFIPLVDMIRNFA--TLKAKINS 412
```

```
Query: 330 ALMKEQNIDITHFEEDLDIFKNAFAKN-YNSASKNFQKAIDEIDKSIKRMEAV-KAALTT 387
           +++ +  D   EE+LD  K     N     +   +K ID+    IK+ E++ ++A
Sbjct: 413 QIIRYE--DRAKIEENLDELKKDIVDNTLKYINDKTKKIIDDSKAIIKKAESIEESAEDI 470

Query: 388 SENQLRLANNKLDDVSVKKL 407
              +L     K+++++++K+
Sbjct: 471 INKKLNTLKKKINELTIRKI 490
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5847> which encodes the amino acid sequence <SEQ ID 5848>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3192(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 = 310/445 (69%), Positives = 352/445 (78%), Gaps = 22/445 (4%)

Query: 1   MNEIKCPHCGTAFAINESEYHQLLEQIRGDAFDKEVSERLEKERLILGEQAKNQLQEVVV 60
           MNEIKCPHC T F INESEY QLLEQ+RG AFD+E+ +RL  E  +L E+AK+QL EVV
Sbjct: 1   MNEIKCPHCHTLFTINESEYSQLLEQVRGQAFDEELKKRLINEIALLEEKAKHQLHEVVA 60

Query: 61  EKDKEIAKLQYKVKQF----------LIEKDNLL----------KDNEYQLAEQLNQK 98
           +K+  I  L  +++Q            L +KD L+            N  +LA QL +K
Sbjct: 61  KKETAITSLTNQLEQIEKEQAYLRQEELAKKDQLIASLEAKLDKLASQNALELANQLAEK 120

Query: 99  DMMLRDLENQIDRLRLEHENSLQEALTKVERERDAIQNQLHIQEKEKDLALASVKSDYEV 158
           D  +  L NQ+D+L LE + + Q  L  +E+ERD I+NQL +Q KE +L+LASV+SDYE
Sbjct: 121 DKEVVSLTNQLDKLALEKDATFQSKLATIEKERDGIKNQLALQAKESELSLASVRSDYEA 180

Query: 159 QLKAANEQVEFYKNFKAQQSTKAVGESLEHYAETEFNKVRHLAFPNAYFEKDNTLSSRGS 218
           QLKAANEQVEFYKNFKAQQSTKA+GESLE YAETEFNKVR  AFPNA F KDN LSSRGS
Sbjct: 181 QLKAANEQVEFYKNFKAQQSTKAIGESLELYAETEFNKVRSYAFPNASFVKDNQLSSRGS 240

Query: 219 KGDFIYREKDENDLEFLSIMFEMKNESDDTIKKHKNEDFFKELDKDRREKSCEYAVLVTM 278
           KGD+IYRE D N +E LSIMFEMKNE+D T  KHKN DFFKELDKDRREK CEYAVLV+M
Sbjct: 241 KGDYIYREVDANGVEILSIMFEMKNEADTTKTKHKNSDFFKELDKDRREKDCEYAVLVSM 300

Query: 279 LEADNDYYNTGIVDVSHKYPKMYVIRPQFFIQLIGILRNAALNTLKYKQELALMKEQNID 338
           LEADNDYYNTGIVDVSH+Y KMYV+RPQ FIQLIGILRNAALN+L YKQELAL+KEQNID
Sbjct: 301 LEADNDYYNTGIVDVSHEYQKMYVVRPQLFIQLIGILRNAALNSLHYKQELALVKEQNID 360

Query: 339 ITHFEEDLDIFKNAFAKNYNSASKNFQKAIDEIDKSIKRMEAVKAALTTSENQLRLANNK 398
           ITHFEEDLD FKNAFAKNY SAS NF+KAIDEIDKSIKRME VK  LTTSENQLRLANNK
Sbjct: 361 ITHFEEDLDQFKNAFAKNYQSASNNFKKAIDEIDKSIKRMEEVKRFLTTSENQLRLANNK 420

Query: 399 LDDVSVKKLTRKNPTMKAKFDALKD 423
           L+DVSVKKLTR+NPTM+ KF+ALKD
Sbjct: 421 LEDVSVKKLTRQNPTMREKFEALKD 445
```

SEQ ID 5846 (GBS304) was expressed in *E.coli* as a His-fusion product. The purified protein is shown in Figure 206, lane 7.

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

Example 1879

A DNA sequence (GBSx1987) was identified in *S.agalactiae* <SEQ ID 5849> which encodes the amino acid sequence <SEQ ID 5850>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1845(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 5851> which encodes the amino acid sequence <SEQ ID 5852>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2492(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 = 113/180 (62%), Positives = 141/180 (77%)

Query: 16  LSELVDCFKGKAVPSKAEAGDIRIINLSDMSPLGIDYHNLRTFQDEQRSLLKYLLQEGDV 75
           L  +VDCFKGKAV SK   GD+ +INLSDM  LGI YH LRTFQ ++R LL+YLL++GDV
Sbjct: 18  LGTVVDCFKGKAVSSKVVPGDVGLINLSDMGTLGIQYHQLRTFQMDRRQLLRYLLEDGDV 77

Query: 76  LIASKGTVKKVAIFEEQDYPVVASANITILRPTQHIRGYYLKLFFDSEEGQQALENANKG 135
           LIASKGT+KKV +F +Q+  VVAS+NIT+LRP + +RGYY+K F DS  GQ  L+ A+ G
Sbjct: 78  LIASKGTLKKVCVFHKQNRDVVASSNITVLRPQKLLRGYYIKFFLDSPIGQALLDVADHG 137

Query: 136 KAVMNISTKELLNIAIPSIPLFRQDYLIQRYKQGLNDYKRKIARAEQEWERIQNDIRQQL 195
           K V+N+STKELL+I IP IPL +QDYLI  Y +GL DY RK+ RAEQEWE IQN+I++ L
Sbjct: 138 KDVINLSTKELLDIPIPVIPLVKQDYLINHYLRGLTDYHRKLNRAEQEWEYIQNEIQKGL 197
```

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

Example 1880

A DNA sequence (GBSx1988) was identified in *S.agalactiae* <SEQ ID 5853> which encodes the amino acid sequence <SEQ ID 5854>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.43   Transmembrane   62 -  78 (  55 -  82)
   INTEGRAL    Likelihood = -2.87   Transmembrane  130 - 146 ( 130 - 150)
   INTEGRAL    Likelihood = -1.28   Transmembrane   37 -  53 (  37 -  53)

----- Final Results -----
             bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9347> which encodes amino acid sequence <SEQ ID 9348> 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 = 38/139 (27%), Positives = 64/139 (45%), Gaps = 5/139 (3%)

Query: 15  SASVEILCRGWLLPVSATKYSKIVSVSISSIFFGLLHSANNHVSLISIFNLCL-FGLFLS 73
           +A+ E++ RG L  +        +++ ++ + FGL+H  N   +L     + +  G  L+
Sbjct: 143 AATEEVVFRGVLFRIIEEHIGTYLALGLTGLVFGLMHLLNEDATLWGALAIAIEAGFMLA 202

Query: 74  LYVILKGNIWGACGIHGAWNCVQGSVFGIEVSGEPMLSNSLVHVKTYGADWISGGKFGVE 133
                N+W   G+H  WN   G VF   VSG    S  L+     G   ++GG FG E
Sbjct: 203 AAYAATRNLWLTIGVHFGWNFAAGGVFSTVVSGNGD-SEGLLDATMSGPKLLTGGDFGPE 261

Query: 134 GSMIT---SIVLIVACYWL 149
           GS+ +    ++L +   WL
Sbjct: 262 GSVYSVGFGVLLTLVFLWL 280
```

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 1881

A DNA sequence (GBSx1989) was identified in *S.agalactiae* <SEQ ID 5855> which encodes the amino acid sequence <SEQ ID 5856>, which is a methylase gene homolog. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2192(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 264-266
```

A related GBS nucleic acid sequence <SEQ ID 9929> which encodes amino acid sequence <SEQ ID 9930> was also identified.

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

```
>GP:BAA87672 GB:AB016260 Hypothetical gene, methylase gene homolog
          [Agrobacterium tumefaciens]
 Identities = 358/1238 (28%), Positives = 595/1238 (47%), Gaps = 99/1238 (7%)

Query: 1072 KEVARIKGMVDIRNAYQEVIAIQRYYDYDKETFNHLLGKLNRTYDSFVKHYGYLNSAV-- 1129
          K V  I+ ++ IR+A +EV+  Q          +  L  +L    + SFV+ +G +N
Sbjct: 497  KHVRIIRKLIPIRDAVREVLKAQEL----DRPWKDLQVRLRVAWSSFVRDFGPINHTTVS 552

Query: 1130 ----------------NRNLFDSDDKYSLLASLEDESL--DPSGKSVIYTKSLAFEKAL 1170
                            N   F  D    L+AS+ED  L  D +     I+T     E+ +
Sbjct: 553  ITEDPESGETRESHRRPNLQPFADDPDCWLVASIEDYDLENDTAKPGAIFT-----ERVI 607

Query: 1171 VRPEKEVKKVHTALDALNSSLADGRGVDFAYMMSIYQVESQMTLIEELGDLIMPDPEKYL 1230
             P   V  + +A DAL   L +   VD  ++  +   +    ++ ELG  I  DP
Sbjct: 608  SPPAPPV--ITSAADALAVVLNERGRVDLDHIAELLHRDPD-DVVAELGSAIFRDP---- 660
```

```
Query: 1231 NGELTYVSRQDFLSGDVVTKLEVVDLFVKQDNQDFNWSHYAGLLEAIKPARITLADIDYR 1290
            + ++      +LSG V  KL+V +      D      ++      L  ++P  +  +DI  R
Sbjct: 661  -ADGSWQMADAYLSGPVRDKLKVAEAAAALDPV---YNRNVTALAGVQPVDLRPSDITAR 716

Query: 1291 IGSRWIPLAVYGKFAQETFMGKAYELSDQ-EVATVLEVSPIDGVITYQSKFAYTYSNATD 1349
           +G+ WIP A     F +E  MG     +    E+A+     +  G +       A T    TD
Sbjct: 717  LGAPWIPAADVVAFVKE-MMGTDIRIHHMPELASWTVEARQLGYLA-----AGTSEWGTD 770

Query: 1350 RSLGVPASRYDSGRKIFENLLNSNQPTITKQVVEGDKKKNVTDVEKTTVLRAKETHLQEL 1409
           R              ++  + LNS  P I    + +GD ++ V +V  T    + K    +++
Sbjct: 771  RR---------HAGELLSDALNSRVPQIFDTIRDGDSERRVLNVVDTEAAKEKLHKIKDA 821

Query: 1410 FQGFVAKYPEVQQMIEDTYNRLYNRTVSKSYDGSHLTIDGLAQNISLRPHQKNAIQRIVE 1469
           FQ ++   P+    +   YN  +N    + + G HL + G +    L  HQK  I RI+
Sbjct: 822  FQRWIWSDPDRTDRLARVYNDRFNNIAPRKFSGDHLNLPGASGAFVLYGHQKRGIWRIIS 881

Query: 1470 EKRALLAHEVGSGKTLTMLGAGFKLKELGMVHKPLYVVPSSLTAQFGQEIMKFFPTKKVY 1529
                LAH VG+GKT+TM  +  + + LG++ K + VVP    AQ  +E +  +PT ++
Sbjct: 882  SGSTYLAHAVGAGKTMTMAASIMEQRRLGLIAKAMQVVPGHCLAQAAREFLALYPTARIL 941

Query: 1530 VTTKKDFAKAKRKQFVSRIITGDYDAIVIGDSQFEKIPMSREKQVTYINDKLEQLREIKL 1589
           V  + +F+K KR +F+SR   T  +DAI+I  S F  I +     I+D+LE     + L
Sbjct: 942  VADETNFSKDKRARFLSRAATATWDAIIITHSAFRFIGVPAAFESQMIHDELELYETLLL 1001

Query: 1590 GSDSDYTV--KEAERSIKGLEHQLEELQKLERDTFIEFENLGIDFLFVDEAHHFKNIRPI 1647
              + + V  K  ER  +GL+ +LE L     +D  +     +G+D + VDEA   F+ +
Sbjct: 1002 KVEDEDRVSRKRLERLKEGLQERLEALST-RKDDLLTIAEIGVDQIIVDEAQEFRKLSFA 1060

Query: 1648 TGLGNVAGITNTTSKKNVDMEMKVRQVQAEHGDRNVVFATGTPVSNSISELFTMMDYIQP 1707
           T +  + G+     S++  D+ +K R ++  +  R +V A+GTP++N++  E+F++    +
Sbjct: 1061 TNMSTLKGVDPNGSQRAWDLYVKSRFIETINPGRALVLASGTPITNTLGEMFSVQRLMGH 1120

Query: 1708 DVLERYLVSNFDSWVGAFGNIENSMELAPTGDKYQPKKRFKKFVNLPELMRIYKETADI- 1766
             LE    + FD+W   FG+     +EL P+G KY+P  RF  FVN+PEL+ +++  AD+
Sbjct: 1121 AALEERGLHEFDAWASTFGDTTTELELQPSG-KYKPVSRFASFVNVPELIAMFRSFADVV 1179

Query: 1767 ---QTSDMLDLP-VPEAKIIAVESELTQAQKYYLEELVKRSDAIKSGS--VDPSRDNMLK 1820
                + + +P +    +  V S+ TQA K++   L +R  AI+      P  D +L
Sbjct: 1180 MPADLREYVKVPAISTGRRQIVTSKPTQAFKHHQMVLAERIKAIEERERPPQPGDDILLS 1239

Query: 1821 ITGEARKLAIDMRLIDPTYSLSDNQKILQVVDNVERIYRDGAGDK----------AT 1867
           +  + R  AID+RL+D        + K+  +V N  RI++  AG                A
Sbjct: 1240 VITDGRHAAIDLRLVDADNDNEPDNKLNNLVSNAFRIWKATAGSVYLRHDSKPFEVPGAA 1299

Query: 1868 QMIFSDIGTPK-SKEEGFDVYNELKDLFVDRGIPKEEIAFVHDANTDEKKNSLSRKVNSG 1926
           QMIFSD+GT    K  GF  Y  ++D  +  G+P  EIAF+ D    E K  L   V +G
Sbjct: 1300 QMIFSDLGTISVEKTRGFSAYRWIRDELIRLGVPASEIAFMQDFKKSEAKQRLFGDVRAG 1359

Query: 1927 EVRILMASTEKGGTGLNVQSRMKAVHYLDVPWRPSDIVQRNGRLIRQGNMHQEVDIYHYI 1986
            VR L+ S+E  GTG+NVQ R+KA+H+LDVPW  PS I  QR GR++RQGN H EVDI+ Y
Sbjct: 1360 RVRFLIGSSETMGTGVNVQLRLKALHHLDVPWLPSQIEQREGRIVRQGNQHDEVDIFAYA 1419

Query: 1987 TKGSFDNYLWQTQENKLKYITQIMTSKDPVRSAEDIDE-QTMTASDFKALATGNPYLKLK 2045
           T+GS D  +WQ  E K ++I    ++    +R  EDI E Q     +  KA+A+G+  L  K
Sbjct: 1420 TEGSLDATMWQNNERKARFIAAALSGDTSIRRLEDIGEGQANQFAMAKAIASGDQRLMQK 1479

Query: 2046 MELENELTVLENQKRAFNRSKDEYRHTISYSEKHLPIMEKRLSQYDKDIAQSLATKSQDF 2105
              LE ++  LE  + A    +  R  +  +E+ + +  +R+++  +DI + + T  +DF
Sbjct: 1480 AGLEADIARLERLRAAHIDDQHAVRRQLRDAERDIEVSTRRIAEIGQDITRLVPTTGEDF 1539

Query: 2106 VMRFDNQAMDNRAEAGDYLRK-LITYNRSETKEVRTLASFRGFDLKM-TTRGASEPLPET 2163
            M    +  R EAG  L K ++T  +   +  +AS  GF+L+    R  +  T
Sbjct: 1540 TMTVAGKDYSERKEAGRALMKEILTLVQLSPEGEAVIASIGGFELEYHGQRYGKDGYRYT 1599

Query: 2164 ISLMIVGDNQYTVALDLK-SDVGTIQRISNAIDHIIDDQEKTQELVKDLKDKLRVAKVEV 2222
           L   G + Y + L +  + +G + R+ +A+D    ++E+ ++ + D + +L   +
Sbjct: 1600 TMLKRTGAD-YEIELPVTVTPLGAVSRLEHALDDFDGERERYRQRLGDARRRLASYQSRG 1658
```

```
Query: 2223 DKVFPKEEDYQLVKAKYDVLAPLVEKEAEIEEIDAALA 2260
            +              +++     L EK  ++ E++ ALA
Sbjct: 1659 E------------GSEFAFAGELAEKHRQLAEVETALA 1684
 Identities = 99/271 (36%), Positives = 153/271 (55%), Gaps = 10/271 (3%)

Query: 607 RDKVETNIVAIRLVKNLEVEHRNASPSEQELLAKYVGWGG--LANEFFD-----DYNPKF 659
           +D+   NI AIRL   +E   R A+  EQE L ++ G+G   LAN  F      ++   +
Sbjct: 80  KDRARDNIAAIRLAAEIEASERPATREEQETLIRFTGFGASDLANGVFRRPGELEFRKGW 139

Query: 660 SKEREELKSLVTDKEYSDMKQSSLTAYYTDPSLIRQMWDKLERDGFTGGKILDPSMGTGN 719
            +   +L+  V + +Y+ + + +  A++T   ++R +W  L+R G+ GG++L+P +GTG
Sbjct: 140 DEIGSDLEDAVGETDYASLARCTQYAHFTPEFIVRAIWSGLQRLGWRGGRVLEPGIGTGL 199

Query: 720 FFAAMPKHLREKSELYGVELDTITGAIAKHLHPNSHIEIKGFETVAFNDNSFDLVISNVP 779
           F A MP+ LR+ S + GVELD +T  I + L P + I    F        SFDL I N P
Sbjct: 200 FPALMPEALRDLSHVTGVELDPVTACIVRLLQPRARILTGDFARTEL-PASFDLAIGNPP 258

Query: 780 FANIRIADNRYDRP--YMIHDYFVKKSLDLLHDGGQVAIISSTGTMDKRTENILQDIRET 837
           F++  +  +R  R     +HDYFV +S+DLL  G   A ++S+GTMDK      Q I  T
Sbjct: 259 FSDRTVRSDRAYRSLGLRLHDYFVARSIDLLKPGAFAAFVTSSGTMDKADSAARQHIATT 318

Query: 838 TEFLGGVRLPDSAFKAIAGTSVTTDMLFFQK 868
            + +  +RLP+ +F+A AGT V  D+LFF+K
Sbjct: 319 ADLIAAIRLPEGSFRADAGTDVVVDILFFRK 349
```

SEQ ID 5856 (GBS327N) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 8-10; MW 140kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 11-13; MW 115kDa) and in Figure 182 (lane 8; MW 115kDa).

Purified GBS327N-GST is shown in Figure 243, lane 5; Purified GBS327N-His is shown in Figure 235, lane 5.

GBS327C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 14; MW 73kDa).

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

Example 1882

A DNA sequence (GBSx1990) was identified in *S.agalactiae* <SEQ ID 5857> which encodes the amino acid sequence <SEQ ID 5858>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3656(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 1883

A repeated DNA sequence (GBSx1991) was identified in *S.agalactiae* <SEQ ID 5859> which encodes the amino acid sequence <SEQ ID 5860>. This protein is predicted to be giant membrane protein. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3698(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:AAG19662 GB:AE005054 calcium-binding protein homology; Cbp
          [Halobacterium sp. NRC-1]
 Identities = 22/43 (51%), Positives = 29/43 (67%), Gaps = 1/43 (2%)

Query: 9   KDSDQDGLTDAQELAL-GTDPQSVDTDGDGQADLEELQSGHSP 50
           +D+D DGL+D  E+ + GTDP   DTDGDG  D  EL++G  P
Sbjct: 198 RDTDDDGLSDGVEVRVAGTDPTERDTDGDGVDDAAELRAGSLP 240
```

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

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

Example 1884

A DNA sequence (GBSx1992) was identified in *S.agalactiae* <SEQ ID 5861> which encodes the amino acid sequence <SEQ ID 5862>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.39   Transmembrane 1609 -1625 (1609 -1625)
   INTEGRAL    Likelihood = -1.81   Transmembrane   30 -  46 (  29 -  46)

----- 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>

LPXTG motif 1600-1604
```

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

```
!GB:X57841 antigen I /II [Streptococcus sobrinus] (v...

>GP:CAA40973 GB:X57841 antigen I /II [Streptococcus sobrinus]
 Identities = 419/1436 (29%), Positives = 608/1436 (42%), Gaps = 310/1436 (21%)

Query: 23   KSKKYRTLCSVALGTMVTAVVAWGGTVAHADEVTTSV----DTTIQRTE--NPATNLPEA 76
           K K  RTL    LGT + A  A  G  A A+E +T+     DT +  TE  NPATNLP+
Sbjct: 23   KVKSGRTLSGALLGTAILASGA--GQKALAEETSTTSTSGGDTAVVGTETGNPATNLPDK 80

Query: 77   QPNP------------------VSEQTESMASTGQSNGAIAVTVPHDTVT-----QAVE 112
           Q NP                   V   T  +    +S     VTV  D         +  +
Sbjct: 81   QDNPSSQAETSQAQARQKTGAMSVDVSTSELDEAAKSPQEAGVTVSQDATVNKGTVEPSD 140

Query: 113  EAKAEGVSTVEDSPMDLGNTRSAVET--------------NQQIS-------------K 144
           EA  +     +D      + + A E                 NQ+I+              K
Sbjct: 141  EANQKEPEIKDDYSKQAADIQKATEDYKASVAANQAETDRINQEIAAKKAQYEQDLAANK 200
```

```
Query: 145  AD------------------ADTQKQVETINEVTK----TYKADKATYESNKARIEQEN 181
            A+                    A  QK +  I +        Y A K  Y+   AR++  N
Sbjct: 201  AEVERSLMRMRKPRPIYEAKLAQNQKDLAAIQQANSDSQAAYAAAKEAYDKEWARVQAAN 260

Query: 182  KELSQAYEGANQTGKETNAWVDTKVNDLKARYADADVTVKEQ-------VVSSGNGTSVL 234
                +AYE A       N  +  ++  ++ R A AD   K             +GN  +
Sbjct: 261  AAAKKAYEEALAANTAKNDQIKAEIEAIQQRSAKADYEAKLAQYEKDLAAAQAGNAANEA 320

Query: 235  DY----TNYGKAVETIQSTNEQAVADY----LTKKTKADDIVAKNQAIQKENEA------ 280
            DY      Y + +  +Q+ N  A   Y          K   I A+N+AIQ+  +A
Sbjct: 321  DYQAKKAAYEQELARVQAANAAAKQAYEQALAANSAKNAQITAENEAIQQNAQAKADYEA 380

Query: 281  -------GLANAKADNEAIERRNQAGQAAVDAEN---RAGQAAVDQANQEKQQLVSDRAA 330
                   LA A++ N A E   Q   AA + E    +A  AA  QA +++ Q  + + A
Sbjct: 381  KLAQYQKDLAAAQSGNAANEADYQEKLAAYEKELARVQAANAAAKQAYEQQVQQANAKNA 440

Query: 331  EIEAITKRNKEKEAAARKENEAIDAYNTKEMERYQRDLAEIS----------------- 372
            EI    +E+ A A+ + E   +    +E+ +Y++DLAE
Sbjct: 441  EITEANRAIRERNAKAKTDYELKLSKYQEELAQYKKDLAEYPAKLQAYQDEQAAIKAALA 500

Query: 373  -----KGEEGYISEALAQALNLNNGEPQAQHGAITRN---------------------- 404
                 K E+G +SE  AQ+L + + EP AQ   +T
Sbjct: 501  ELEKHKNEDGNLSEPSAQSL-VYDLEPNAQVALVTDGKLLKASALDEAFSHDEKNYNNHL 559

Query: 405  --PDQI----------ISTGDALLGGYSRILDSTGF----------FVYDMFKTGETLS 441
              PD +          +++   L G +    D  G+           F   + K G++ +
Sbjct: 560  LQPDNLNVTYLEQADDVASSVELFGNFG---DKAGWTTTVSNGAEVKFASVLLKRGQSAT 616

Query: 442  FNYQNLQHARFDGKKISRVTYDITNLVSPAG-----TNAVKLVVPNDPTEGFIAYRNDGN 496
              Y NL+++ ++GKKIS+V Y  T  V P       T  V L +  DPT G  A    G
Sbjct: 617  ATYTNLKNSYYNGKKISKVVYKYT--VDPDSKFQNPTGNVWLGIFTDPTLGVFASAYTGQ 674

Query: 497  GDWRTD---KMEFRVVAKYYLEDGSQVTFSKEKPGVFTHSSLNHNDIGLEYVKDSSGKFV 553
              +  T    K EF      +Y EDG+ + F      + + +SLN      +E  KD SG FV
Sbjct: 675  NEKDTSIFIKNEF----TFYDEDGNPIDFDN---ALLSVASLNREHNSIEMAKDYSGTFV 727

Query: 554  PINGSTVQVTN--------------EGLARSLGSNRASDLNLPEEWDTTSSRYAYKGAIV 599
             I+GS++   N               EG +      RAS+      WD+  +  ++ GA
Sbjct: 728  KISGSSIGEKNGMIYATDTLNFKKGEGGSLHTMYTRASEPG--SGWDSADAPNSWYGAGA 785

Query: 600  STVTSGNTY--------TVTFGQGDMPQNVGL--------SYWFALN------------- 630
              ++  N Y         T       +MPQ  G         + W++LN
Sbjct: 786  VRMSGPNNYITLGATSATNVLSLAEMPQVPGKDNTAGKKPNIWYSLNGKIRAVNVPKVTK 845

Query: 631  --TLPVARTVTPYSPKPHVTVEL-----EPIPEPITVTPDIYTPKTFTPEKPVTFT---- 679
                P      P  P   V  EL     EP  EP    TP    P    PEKPV  T
Sbjct: 846  EKPTPPVEPTKPDEPTYEVEKELVDLPVEPKYEP-EPTPPSKNPDQSIPEKPVEPTYEVE 904

Query: 680  ----PKPLDEVVQPSLTLTKVT-------LPVKPIPKELPTPP------------QVPTV 716
                P P++   +   T  T          PV+P  + LPTPP              VPTV
Sbjct: 905  KELEPAPVEPSYEKEPTPPQSTPDQEEPEKPVEPSYQSLPTPPVEPVYETVPGPVSVPTV 964

Query: 717  HYHAYRLTTTSEIMKEVVNSDQANLHEKTVAKDSTVIYPLTVDALSPNRAQTTSLIFEDY 776
             YH Y+L     + KE+ N D  ++ +  VAK STV + L    L   R +TTS +  D
Sbjct: 965  RYHYYKLAVQPGVTKEIKNQDDLDIDKTLVAKQSTVKFQLKTADLPAGRPETTSFVLMDP 1024

Query: 777  LPAGYLFDKETTQKENGNYVLSFDETKNFVTLTAKENLLQEVNKDLTQVYQLTAPKLYGS 836
            LP+GY  + E T+  +  +  S+D   + VT TA    L  +N+DLT+      P + G
Sbjct: 1025 LPSGYQLNLEATKVASPGFEASYDAMTHTVTFTATAETLAALNQDLTKAVATIYPTVVGQ 1084

Query: 837  VQNDGATYSNSYKLLLNKGTTNAYTVTSNVVTVRTPG-----DGETTTLITPDKNNENAD 891
            V NDGATY+N++ L++N     +AY + SN+V V TPG      D  +   ITP K N+N +
Sbjct: 1085 VLNDGATYTNNFTLMVN----DAYGIKSNIVRVTTPGKPNDPDNPSNNYITPHKVNKNEN 1140

Query: 892  GVLINDTVVALGTTNHYRLTWDLDQYKGDRSAKETIARGFFFVDDYPEEVLDVVENGTAI 951
            GV+I+   V  GTTN+Y LTWDLDQYKGD+SAKE I +GFF+VDDYPEE LD+  +   +
Sbjct: 1141 GVVIDGKSVLAGTTNYYELTWDLDQYKGDKSAKEIIQKGFFYVDDYPEEALDLRTDLIKL 1200
```

```
Query: 952  TTLDGQKVSGITVKNYASLNEAPKDLQDKLARAKITPTGAFQVFMPDDNQAFYDQYVQTG 1011
            T  +G+ V+G++V +YASL  AP  +QD L +A I P GAFQVF  DD QAFYD YV TG
Sbjct: 1201 TDANGKAVTGVSVADYASLEAAPAAVQDMLKKANIIPKGAFQVFTADDPQAFYDAYVVTG 1260

Query: 1012 TSLALLTKMTVKDSLYGQTKTYTNKAYQVDFGNGYETKEVTNTLVSPEPKKQ-NLNKDKV 1070
            T L ++T MTVK  +     +Y N+AYQ+DFGNGYE+  V N +    P+K   L  D
Sbjct: 1261 TDLTIVTPMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVVNNVPKINPEKDVTLTMDPA 1320

Query: 1071 D---INGKPMLVGTQNHYTLSWDLDQYRGIKADNSQIAQGFYFVDDYPE-----EALLPD 1122
            D   ++G+ + +    +Y L   +     I AD+++    + F DDY +
Sbjct: 1321 DSTNVDGQTIALNQVFNYRLIGGI-----IPADHAEELFEYSFSDDYDQTGDQYTGQYKA 1375

Query: 1123 EAAIQFVTSDGKTV-SGITVKSY--SQLLEAPKTLQAAFSKQKIQPKGAFQVFMPE 1175
             A +     DG  + +G  + SY  +Q+ EA   +   F +  ++       F  E
Sbjct: 1376 FAKVDLTLKDGTIIKAGTDLTSYTEAQVDEANGQIVVTFKEDFLRSVSVDSAFQAE 1431
 Identities = 209/442 (47%), Positives = 280/442 (63%), Gaps = 27/442 (6%)

Query: 1198 TVLETMLNSGKSY-ENVAYQVDFGQAYETNTVTNFVPK-----------VTPHKSNTNQ 1244
            TV+  +LN G +Y  N    V+     ++N V    P              +TPHK N N+
Sbjct: 1080 TVVGQVLNDGATYTNNFTLMVNDAYGIKSNIVRVTTPGKPNDPDNPSNNYITPHKVNKNE 1139

Query: 1245 EGISIDGKTVLPNTVNYYKIVLDYSQYKDMVVTDDVLAKGFYMVDDYPEEALTLNPDGIQ 1304
             G+ IDGK+VL  T NYY++  D  QYK      +++ KGF+ VDDYPEEAL L  D I+
Sbjct: 1140 NGVVIDGKSVLAGTTNYYELTWDLDQYKGDKSAKEIIQKGFFYVDDYPEEALDLRTDLIK 1199

Query: 1305 VLDKDGNRVSGISVSTYASLSEAPKVVQDAMAKRQFTPKGAIQVLSSDDPKVFYDTYVKT 1364
            + D +G  V+G+SV+ YASL  AP  VQD + K    PKGA QV ++DDP+ FYD YV T
Sbjct: 1200 LTDANGKAVTGVSVADYASLEAAPAAVQDMLKKANIIPKGAFQVFTADDPQAFYDAYVVT 1259

Query: 1365 GQTLVVTLPMTVKNELTKTGGQYENTAYQIDFGLAYVTETVVNNVPKLDPQKDVVIDLSH 1424
            G  L +  PMTVK E+ KTGG YEN AYQIDFG  Y +  VVNNVPK++P+KDV + +
Sbjct: 1260 GTDLTIVTPMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVVNNVPKINPEKDVTLTMDP 1319

Query: 1425 KDA-SLDGKEVALHQTFNYRLVGAMIPSNRATDLFEYGFEDNYDEKHDEYNGVYRSYLMT 1483
             D+ ++DG+ +AL+Q FNYRL+G +IP++ A +LFEY F D+YD+  D+Y G Y+++
Sbjct: 1320 ADSTNVDGQTIALNQVFNYRLIGGIIPADHAEELFEYSFSDDYDQTGDQYTGQYKAFAKV 1379

Query: 1484 DVILKDGSVLKEGTEVTKYTLQQVDTENGLVSISFDKSFLETVSDDSAFQADVYLQMKRI 1543
            D+ LKDG+++K GT++T YT  QVD  NG + ++F + FL +VS DSAFQA+VYLQMKRI
Sbjct: 1380 DLTLKDGTIIKAGTDLTSYTEAQVDEANGQIVVTFKEDFLRSVSVDSAFQAEVYLQMKRI 1439

Query: 1544 AAGQVENTYLHTVNGYVISSNTVVTHTPQPEEPSPNQP--------TPPQPPIETIEPPV 1595
            A G   NTY++TVNG  SSNTV T TP+P++PSP  P          P Q       PP
Sbjct: 1440 AVGTFANTYVNTVNGITYSSNTVRTSTPEPKQPSPVDPKTTTTVVFQPRQGKAYQPAPPA 1499

Query: 1596 PASILPNTGEQES----LLGLI 1613
             A  LP TG+  +    LLGL+
Sbjct: 1500 GAQ-LPATGDSSNAYLPLLGLV 1520
 Identities = 100/210 (47%), Positives = 137/210 (64%), Gaps = 4/210 (1%)

Query: 1060 PKKQNLNKDKVDINGKPMLVGTQNHYTLSWDLDQYRGIKADNSQIAQGFYFVDDYPEEAL 1119
            P K N N++ V I+GK +L GT N+Y L+WDLDQY+G K+    I +GF++VDDYPEEAL
Sbjct: 1132 PHKVNKNENGVVIDGKSVLAGTTNYYELTWDLDQYKGDKSAKEIIQKGFFYVDDYPEEAL 1191

Query: 1120 LPDEAAIQFVTSDGKTVSGITVKSYSQLLEAPKTLQAAFSKQKIQPKGAFQVFMPEDPQA 1179
                 I+   ++GK V+G++V  Y+ L  AP  +Q    K  I PKGAFQVF  +DPQA
Sbjct: 1192 DLRTDLIKLTDANGKAVTGVSVADYASLEAAPAAVQDMLKKANIIPKGAFQVFTADDPQA 1251

Query: 1180 FFESYVTKGENITIVTPMTVLETMLNSGKSYENVAYQVDFGQAYETNTVTNFVPKVTPHK 1239
            F+++YV  G ++TIVTPMTV   M  +G SYEN AYQ+DFG  YE+N V N VPK+ P K
Sbjct: 1252 FYDAYVVTGTDLTIVTPMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVVNNVPKINPEK 1311

Query: 1240 SNT----NQEGISIDGKTVLPNTVNYYKIV 1265
             T      +  ++DG+T+  N V  Y+++
Sbjct: 1312 DVTLTMDPADSTNVDGQTIALNQVFNYRLI 1341
```

There is also homology to SEQ ID 598.

SEQ ID 5862 (GBS76) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 2; MW 17.4kDa). The GBS76-His fusion product was purified (Figure 196, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 294), 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 1885

A DNA sequence (GBSx1993) was identified in *S.agalactiae* <SEQ ID 5863> which encodes the amino acid sequence <SEQ ID 5864>. This protein is predicted to be abortive infection bacteriophage resistance protein (abiEi). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2765(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 9931> which encodes amino acid sequence <SEQ ID 9932> was also identified.

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

```
>GP:AAB52382 GB:U36837 AbiEi [Lactococcus lactis]
 Identities = 51/206 (24%), Positives = 90/206 (42%), Gaps = 23/206 (11%)

Query: 17  KNNGIVTNKDCKALGIPTIYLTRLEKEGIIFRVEKGIPLTQNGDYDEYYFFQYRFPKAIF 76
           K  G +  K  +  GI   YL +   +  +  V+KG+++  +   D  + FQ ++ KA+
Sbjct: 76  KYKGNIIRKIVRDEGISDYYLRKFVLKYNLTEVDKGVYIFPHKKKDSLFIFQQKYSKAVI 135

Query: 77  SYISALYLQQFTDEIPQYFDVTVPRGYRF----------------------NTPPANLNI 114
           S+ ++LYLQ   D IPQ   ++VP  Y                        N    N+ I
Sbjct: 136 SHETSLYLQDVIDYIPQKIQMSVPEKYNISRIQEPHENRLTSYNYVDINSNNIMDKNIPI 195

Query: 115 HFV-SKEYSELGMTTVPTPMGNNVRVYDFERIICDFVIHREKIDSELFVKTLQSYGNYPK 173
           + V +K  S   + TV + +G  +RV    R I D +    K + E+  + ++ Y
Sbjct: 196 NLVRNKSISPTQIETVNSFLGLPLRVTSIARSIVDVLKPSHKAEEEVKEQAIKYYLERFP 255

Query: 174 KNLAKLYEYATKMNTLEKVKQTLEVL 199
            N+ +L   A   N L++++  L +L
Sbjct: 256 DNIVRLKRIAKTQNVLKELEYYLILL 281
```

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 1886

A DNA sequence (GBSx1994) was identified in *S.agalactiae* <SEQ ID 5865> which encodes the amino acid sequence <SEQ ID 5866>. This protein is predicted to be abortive infection bacteriophage resistance protein (abiEii). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.12   Transmembrane  260 - 276 ( 259 - 277)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAB52383 GB:U36837 AbiEii [Lactococcus lactis]
 Identities = 76/276 (27%), Positives = 135/276 (48%), Gaps = 19/276 (6%)

Query: 14  SKNTGLTFNSVMTYYFLEVILKKLSQSSYSNHYIFKGGFLLSNVIGVESRSTVDIDFLFH 73
           ++N  +   +    Y  E  L +LS S Y   ++ KGGFL+     +  R+T D+D
Sbjct: 12  TRNDDIGIENYRIRYATERFLTRLSASQYKEKFVLKGGFLIGVTYNLSQRTTKDLDTALI 71

Query: 74  QITLSEETVKQQLKEIL-ADSEEGISFVIQSITTIKESDDYGGYRATISCQLE--NIKQV 130
                +++++ + EI   D E+ + F ++ +T+ ++   Y GYRA +       N +
Sbjct: 72  DFKSDAQSIERVITEICNIDLEDQVLFKLKELTSSQDMRIYPGYRAKLKMMFPDGNTRID 131

Query: 131 IHLDIATGDVVTPQPITYDYKAIFDE-----DNFPIIAYTIETILAEKLQTIYSRNFLNS 185
             LDI  GD +TP+         IF+E          ++AY  ETI AEKL+TI +R  +N+
Sbjct: 132 FDLDIGVGDRITPEAKKIKIPLIFNEVKGVEKQIEVLAYPKETIQAEKLETILTRGKVNT 191

Query: 186 RSKDFYDVYIL--SKLKKKDIDFNQLKNACQRTFSYRE-TELDFEKIIE-----LLERFK 237
           R KD+YD ++L   +     I F     A + T+ +R  T+   E++ E     L E  +
Sbjct: 192 RMKDYYDFHLLLTDQENSNSISFYY---AFKNTWEFRNPTQFIDEELFEDWLFILDEILE 248

Query: 238 SDPTQNQQWQNYSKKYSYTKGISLANVLDEMISLIT 273
           S   + + W NY K  +Y K +++ +++ E+   ++
Sbjct: 249 SKELKEKYWPNYIKDRNYAKHLNMDDIISEIKEFVS 284
```

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 1887

A DNA sequence (GBSx1995) was identified in *S.agalactiae* <SEQ ID 5867> which encodes the amino acid sequence <SEQ ID 5868>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1137(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 1888

A DNA sequence (GBSx1996) was identified in *S.agalactiae* <SEQ ID 5869> which encodes the amino acid sequence <SEQ ID 5870>. Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2782(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 1889

A DNA sequence (GBSx1997) was identified in *S.agalactiae* <SEQ ID 5871> which encodes the amino acid sequence <SEQ ID 5872>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.14   Transmembrane  310 - 326 ( 301 - 334)

----- Final Results -----
              bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAG38044 GB:AF295925 Orf28 [Streptococcus pneumoniae]
 Identities = 272/344 (79%), Positives = 307/344 (89%)

Query: 568 VYVNPAFYFPKVIQVQTTILPTIGQFGGDEFERAKAIYDYLKSKGATNQAIAAILGNWSV 627
           +YVNP FYFPKVIQ+QTTILP IGQFGGDEFERAK IY++LKS+GA+ QAIAAILGNWSV
Sbjct: 1   MYVNPQFYFPKVIQLQTTILPAIGQFGGDEFERAKHIYEFLKSQGASPQAIAAILGNWSV 60

Query: 628 ESSINPKRAEGDYLSPPVGATDSSWDDEGWLTLNGPTIYNGRYPNILKRGLGLGQWTDTA 687
           ESSINPKRAEGDYL+PPVG     WDDE WL + GP IY+G YPNIL RGLGLGQWTDTA
Sbjct: 61  ESSINPKRAEGDYLTPPVGVPIPPWDDESWLAIGGPAIYSGAYPNILHRGLGLGQWTDTA 120

Query: 688 DGSRRHTLLLEYAKGKHQKWYDLGLQLDFMLYGDSPYYTNWLKDFFKNSGSPASLAQLFL 747
           DGS RHT LL YA+ +++KWYDL LQLDFML+GDSPYY +WLKDFFKN+GS A+LAQLFL
Sbjct: 121 DGSTRHTALLNYARTQNKKWYDLDLQLDFMLHGDSPYYQSWLKDFFKNTGSAANLAQLFL 180

Query: 748 IYWEGNSGDKLLERQTRASEWYYQIEKGFSQPNGGTAQSDPKALEAVREDLFENSIPGGG 807
            YWEGNSGDKLLERQTRA+EWYYQIEKGFSQ NGG A+SDP++LE VR DL+++S+PGGG
Sbjct: 181 TYWEGNSGDKLLERQTRATEWYYQIEKGFSQTNGGQAKSDPQSLEGVRGDLYDHSVPGGG 240

Query: 808 DGMGYAYGQCTWGVAARINQLGLKLKGKNGEKIPIISTMGNGQDWVRTAASLGGETGTSP 867
           DGM YAYGQCTWGVAAR+NQLGLKLKG+NGEKI II+TMGNGQDWV T++SLGGETG++P
Sbjct: 241 DGMAYAYGQCTWGVAARMNQLGLKLKGRNGEKISIINTMGNGQDWVATSSSLGGETGSTP 300

Query: 868 QEGAILSFAGGGHGTPTEYGHVAFVEKVYPDGSFLISETNYNGN 911
           + GAI+SF GG HGTP  YGHVAFVEKVY DGSFL+SETNY GN
Sbjct: 301 RAGAIVSFVGGTHGTPASYGHVAFVEKVYDDGSFLVSETNYGGN 344
```

SEQ ID 5872 (GBS74d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 3 & 4; MW 95.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 5-7; MW 70.5kDa) and in Figure 179 (lane 9; MW 70.5kDa).

GBS74d-His was purified as shown in Figure 233, lane 7-8.

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

Example 1890

A DNA sequence (GBSx1998) was identified in *S.agalactiae* <SEQ ID 5873> which encodes the amino acid sequence <SEQ ID 5874>. This protein is predicted to be TrsE-like protein. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5526(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:AAG38042 GB:AF295925 Orf26 [Streptococcus pneumoniae]
 Identities = 618/782 (79%), Positives = 712/782 (91%), Gaps = 1/782 (0%)

Query: 1   MKKLKHSMKSK-TSSNDKKQKTKTQKQEISPSTVNTLAYQGLFQNGLMQVSPSYFSQTYL 59
           MK+  +++K + TS+ +KK++ K +K+E+ PST NTL+YQ L+QNGLMQV   YFSQ+YL
Sbjct: 3   MKRKSNTLKKQQTSTTNKKEEVKDKKEEVLPSTANTLSYQALYQNGLMQVKEDYFSQSYL 62

Query: 60  LGDVNYQTVGLDDKGAIVEKYSDLINSLDDKTNFQLTIFNQKVNLEKFRKSILYPLQEDG 119
           LGDVNYQTVGL+DKGAI+EKYSDLI SLDD+TNFQLTIFN+++NLEKFR S+LY  +EDG
Sbjct: 63  LGDVNYQTVGLEDKGAIIEKYSDLIKSLDDQTNFQLTIFNKRLNLEKFRHSVLYEEKEDG 122

Query: 120 FDTYRDELNRMMDANLEAGENNFSAVKFLSFGKSDQTPKLAFRSLSQIGEYFKSGFSEID 179
           +D+YR ELNRMM+ NL++GENNFSAVK +SFG+ D  PK A+RSLSQIGEYFKSGFSEID
Sbjct: 123 YDSYRKELNRMMNQNLDSGENNFSAVKLISFGRKDSNPKQAYRSLSQIGEYFKSGFSEID 182

Query: 180 VSLGLLGGEERVNVLADMLRGENHLPFSYKDLTLSGQSTKHFIAPTYLSFKHKNHIELDD 239
                L GEERVN+LADMLRGE+HLPFSY+DLT SGQ+T+HFIAP  L FK+KN+++++D
Sbjct: 183 ARFESLAGEERVNLLADMLRGEHHLPFSYRDLTRSGQTTRHFIAPNLLDFKNKNYLQIND 242

Query: 240 RLLQIVYVRDYGMELGDKFIRDLMQSDLEVMISLHAKGSTKSETMTKLRTKKTLMESQKI 299
           RLLQIVYVRDYGMELGD+FIRDLMQ DLE+++SLHA+ STKS+ M KLRTKKTLMESQKI
Sbjct: 243 RLLQIVYVRDYGMELGDQFIRDLMQGDLELIVSLHAQSSTKSDAMKKLRTKKTLMESQKI 302

Query: 300 GEQQKMARTGIYLEKVGHVLENNIDEAEALLQTMTQTGDKLFDTVFLIGVLADTEDQLKQ 359
           GEQQK+ARTGIYLEKVGHVLE+NIDEAE LL+TMT+TGDKLF TVFLIGV    E++LKQ
Sbjct: 303 GEQQKLARTGIYLEKVGHVLESNIDEAEELLKTMTETGDKLFQTVFLIGVFGQDEEELKQ 362

Query: 360 SLDIIKQVAGSNDMIIDNLTYMQEAAFNSLLPFGKNYLEGVSRSLLTSNIAVNAPWTSVD 419
           +LD ++QVAGSND++ID L YMQEAAFNSLLPFG ++LEGVSRSLLTSNIAVN+PWTSVD
Sbjct: 363 ALDTVQQVAGSNDLMIDKLPYMQEAAFNSLLPFGCDFLEGVSRSLLTSNIAVNSPWTSVD 422

Query: 420 IHDKGGKFYGINQISSNIISIDRGKLNTPSGLILGTSGAGKGMATKHEIISTKLKEADSD 479
           + D+ GK+YGINQISSNII+IDR  LNTPSGLILGTSGAGKGMATKHEII+TK+KE+  +
Sbjct: 423 LQDRSGKYYGINQISSNIITIDRSLLNTPSGLILGTSGAGKGMATKHEIITTKIKESGEN 482

Query: 480 TEIIIVDPENEYSIIGQAFGGESIDIAPDSTTFLNVLELSDENMDEDPVKVKSEFLLSWI 539
           TEIIIVDPE EYS+IG+ FGGE IDIAPDS T+LNVL+LS+ENMDEDPVKVKSEFLLS+I
Sbjct: 483 TEIIIVDPEAEYSVIGRTFGGEMIDIAPDSETYLNVLDLSEENMDEDPVKVKSEFLLSFI 542

Query: 540 GKLLDRKMDGREKSLIDRVTRLTYKHFDTPSLVEWVFVLSQQPEQEAKDLALDMELYVEG 599
           GKLLDRKMDGREKS+IDRVTRLTY+ F  PSL EWVFVLSQQPE+EA++LALDMELYVEG
Sbjct: 543 GKLLDRKMDGREKSIIDRVTRLTYQSFKEPSLEEWVFVLSQQPEEEAQNLALDMELYVEG 602

Query: 600 SLDIFSHRTNIKTDSHFLIYNVKKLGDELKQIALMVIFDQIWNRVVKNQKLGKKTWIYFD 659
           SLDIFSH+TNI+T S+FLIYNVKKLGDELKQIALMV+FDQIWNRVV+NQKLGKKTWIYFD
Sbjct: 603 SLDIFSHKTNIQTGSNFLIYNVKKLGDELKQIALMVVFDQIWNRVVRNQKLGKKTWIYFD 662
```

```
Query: 660 EMQLLLLDKYASDFFFKLWSRVRKYGAIPTGITQNVETLLLDANGRRIIANSEFMILLKQ 719
           E++LLLLDKY SDFFFKLWSRVRKYGA PTGITQNVETLLLD NGRRIIANSEFMILLKQ
Sbjct: 663 EIELLLLDKYPSDFFFKLWSRVRKYGASPTGITQNVETLLLDPNGRRIIANSEFMILLKQ 722

Query: 720 AKSDREELVHMLGLSKELEKYLVNPEKGAGLIKAGSTVVPFKNKIPQHTKLFDIMSTDPE 779
           AK+DREELV +LGLSKELEKYLVNPEKGAGLIKAGS VVPFKNKIPQ ++LFDIM +DP+
Sbjct: 723 AKNDREELVQLLGLSKELEKYLVNPEKGAGLIKAGSVVVPFKNKIPQGSQLFDIMRSDPD 782

Query: 780 KM 781
           KM
Sbjct: 783 KM 784
```

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

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

A related GBS gene <SEQ ID 8925> and protein <SEQ ID 8926> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:    -26.26
GvH: Signal Score (-7.5): -3.87
     Possible site: 55
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 0 value:   6.26 threshold:  0.0
   PERIPHERAL  Likelihood =  6.26     335
 modified ALOM score:  -1.75

*** Reasoning Step: 3

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5526(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear)
```

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

```
33.5/57.2% over 789aa
                                                  Enterococcus faecalis
  GP|8100663| TrsE-like protein Insert characterized

ORF01332(319 - 2628 of 2949)
GP|8100663|gb|AAF72347.1|AF192329_8|AF192329(2   -   791   of   799)   TrsE-like   protein
{Enterococcus faecalis
}
%Match = 20.7
%Identity = 33.4  %Similarity = 57.2
Matches = 259  Mismatches = 323  Conservative Sub.s = 184

210       240       270       300       330       360                387
SCYLGSIAPTIYHLKYTSSTVFIMN*RCQTAHLLEEKETNVKKLKHSMKSKTSSNDKKQKTKTQKQEI----------S
                                          | |      :| : |: |::||
                                        MSKKEIPRETEKTKLTRAQRKEIDAVIRKYKGDGR
                                                 10        20        30

414       444       474       504       534       564       594       624
PSTVN-TLAYQGLFQNGLMQVSPSYFSQTYLLGDVNYQTVGLDDKGAIVEKYSDLINSLDDKTNFQLTIFNQKVNLEKFR
| |   :: |: :: :|: :||| ||:   : |::||    | : || ||  || | :|   : |:: :|:||:  ::
PHTAQQSIPYEVMYPDGVCRVSPGVFSKCIEFADISYQLAQPDTQTAIFEKLCDLYNYVDASIHIQFSFLNRKVDPVQYA
             50        60        70        80        90       100       110

654       684       714       744       774       804       834       864
KSILYPLQEDGFDTYRDELNRMMDANLEAGENNFSAVKFLSFGKSDQTPKLAFRSLSQIGEYFKSGFSEIDVSLGLLGGE
||     | | ||  | |   ::   |  | |     |:|:|   :: | |   | :|| :   |  :     :: |
KSFEIAPQGDDFDDIRAEYTGILQKQLANGNNGMVKTKYLTFTIEAESVKAARARLKRIGFDLLGYFKSMGAVAHVMDGW
```

```
          130       140       150       160       170       180       190

894       921       951       981      1011      1041      1071      1101
ERVNVLADMLRGENHL-PFSYKDLTLSGQSTKHFIAPTYLSFKHKNHIELDDRLLQIVYVRDYGMELGDKFIRDLMQSDL
 ||:|:|  :   :  :  | :| |  || ||| ||||: | | :     :  :   : :::    || |  : |:: ::
ERLNLLHGVYHPDGEIFNFDWKWLAPSGLSTKDFIAPSSLCFGNAKTFGMGGKYGAVSFLQILSPELSDDMLADFLNTES
          210       220       230       240       250       260       270

1131      1161      1191      1221      1251      1281      1311      1341
EVMISLHAKGSTKSETMTKLRTKKTLMESQKIGEQQKMARTGIYLEKVGHVLENNIDEAEALLQTMTQTGDKLFDTVFLI
 |:::|| :   ::: :  :: | | ::: || ||:|  |:|  :: :   |    ::|: ||  :    ::||   ||:
GVLVNLHVQAIEQTKAIKTIKRKITDLDAMKIAEQKKAVRSGYDMDILPSDLATYGEDAKKLLTKLQTRNERLFQLTFLV
          290       300       310       320       330       340       350

1371      1401      1431      1461      1491      1521      1551
GVLADTEDQLKQSLDIIKQVAGSNDMIIDNLTYMQEAAFNSLLPFGKNYLEGVSRSLLTSNIAVNAPWTSVDIHDKGGK-
  :|||: :|   :     ||  ::  :  | | ||  : | ||:| | :  : ||| ||::||  |: : ::   |
LNVADTKQKLNNDVFQAAGVAQKHNCPLVRLDYQQEQGLASSLPLGVNQI-KIQRSLTTSSVAVFVPFVTQELFQGGAAM
          370       380       390       400       410       420       430

1608      1638      1668      1698      1728      1758      1788      1818
FYGINQISSNIISIDRGKLNTPSGLILGTSGAGKGMATKHEIISTKLKEADSDTEIIIVDPENEYSIIGQAFGGESIDIA
:|||| | |:| :|| :  |: | ||| |:|| |: | ||:|  |   |  :|.| ||| ||  : : : |: | ::
YYGINAKSRNMIMLDRKQARCPNALKLGTPGSGKSMSCKSEIVSVFLTTPD---DIFISDPEAEYYPLVKRLHGQVIRLS
          450       460       470       480       490       500       510

1848      1875      1905      1935      1959      1989      2019
PDSTTFLNVLELS-DENMDEDPVKVKSEFLLSWIGKLLDRK--MDGREKSLIDRVTRLTYKHFDTPSLVEWVFVLS----
| |  |:| |::: : : |::|: :||:|:||:   ::  |  ::  ||::|||  |: |: :       | : :||
PTSKDFVNPLDINLNYSEDDNPLALKSDFVLSFCELVMGGKNGLEAIEKTVIDRAVRVIYRPYLADPRPENMPILSDLHK
          530       540       550       560       570       580       590

2058      2088      2118      2148      2178      2208      2238      2268
---QQPEQEAKDLALDMELYVEGSLDIFSHRTNIKTDSHFLIYNVKKLGDELKQIALMVIFDQIWNRVVKNQKLGKKTWI
    |   ||  :|  ::||| |||::|:||||:   : :: :::|:|| :||:: :::: |||| ||  |:  || ||
ALLDQHVPEADRVAQALDLYVSGSLNVFNHRTNVDIGNRLVSFDIKELGKQLKKLGMLIVQDQIWGRVTANRSQGKATWY
          610       620       630       640       650       660       670

2298      2328      2358      2388      2418      2448      2478      2508
YFDEMQLLLLDKYASDFFFKLWSRVRKYGAIPTGITQNVETLLLDANGRRIIANSEFMILLKQAKSDREELVHMLGLSKE
: || :||| ::  : :  ::| | ||:| |||| ||||: ||       |: ||:|: || ||  ||: |   | || |
FADEFHLLLKEEQTAAYSAEIWKRFRKWGGIPTGATQNVKDLLSSPEIENILENSDFITLLNQASGDRKILAERLNLSTE
          690       700       710       720       730       740       750

2538      2568      2598      2628      2658      2688      2718      2748
LEKYLVNPEKGAGLIKAGSTVVPFKNKIPQHTKLFDIMSTDPEKMRT*DERG*KASQTG*AKLSKQLKISSYALSERS*D
 :||: | | | ||:   : |:|| | ||::|:|: ||:|    :
QQKYIDNSEPGEGLLIFENVVLPFTNPIPHNTQLYKIMTTRLNEVAGV
          770       780       790
```

A related GBS gene <SEQ ID 8927> and protein <SEQ ID 8928> were also identified. Analysis of this protein sequence reveals the following:

```
This protein might be involved in vancomycin resistance
```

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

```
>GP|8100663|gb|AAF72347.1|AF192329_8|AF192329 TrsE-like protein
          {Enterococcus faecalis}

 Score =  427 bits (1086), Expect = e-118
 Identities = 257/785 (32%), Positives = 431/785 (54%), Gaps = 28/785 (3%)

Query: 9   DKKQKTKTQKQEIS----------PSTVN-TLAYQGLFQNGLMQVSPSYFSQTYLLGDV 56
           +K + T+ Q++EI           P T   ++ Y+ ++ +G+ +VSP  FS+     D+
Sbjct: 11  EKTKLTRAQRKEIDAVIRKYKGDGRPHTAQQSIPYEVMYPDGVCRVSPGVFSKCIEFADI 70

Query: 57  NYQTVGLDDKGAIVEKYSDLINSLDDKTNFQLTIFNQKVNLEKFRKSILYPLQEDGFDTY 116
```

```
            +YQ    D + AI EK  DL N +D   + Q +  N+KV+  ++ KS     Q D FD
Sbjct: 71   SYQLAQPDTQTAIFEKLCDLYNYVDASIHIQFSFLNRKVDPVQYAKSFEIAPQGDDFDDI 130

Query: 117 RDELNRMMDANLEAGENNFSAVKFLSFGKSDQTPKLAFRSLSQIGEYFKSGFSEIDVSLG 176
            R E   ++   L  G N     K+L+F    ++ K A   L +IG      F  +
Sbjct: 131 RAEYTGILQKQLANGNNGMVKTKYLTFTIEAESVKAARARLKRIGFDLLGYFKSMGAVAH 190

Query: 177 LLGGEERVNVLADMLRGENHL-PFSYKDLTLSGQSTKHFIAPTYLSFKHKNHIELDDRLL 235
            ++ G ER+N+L  +   +  +  F +K L  SG STK FIAP+ L F +     +  +
Sbjct: 191 VMDGWERLNLLHGVYHPDGEIFNFDWKWLAPSGLSTKDFIAPSSLCFGNAKTFGMGGKYG 250

Query: 236 QIVYVRDYGMELGDKFIRDLMQSDLEVMISLHAKGSTKSETMTKLRTKKTLMESQKIGEQ 295
             + +++    EL D  + D + ++  V+++LH +   +++ +   ++ K T +++ KI EQ
Sbjct: 251 AVSFLQILSPELSDDMLADFLNTESGVLVNLHVQAIEQTKAIKTIKRKITDLDAMKIAEQ 310

Query: 296 QKMARTGIYLEKVGHVLENNIDEAEALLQTMTQTGDKLFDTVFLIGVLADTEDQLKQSLD 355
            +K  R+G  ++ +    L    ++A+ LL  +    ++LF   FL+  +ADT+ +L   +
Sbjct: 311 KKAVRSGYDMDILPSDLATYGEDAKKLLTKLQTRNERLFQLTFLVLNVADTKQKLNNDVF 370

Query: 356 IIKQVAGSNDMIIDNLTYMQEAAFNSLLPFGKNYLEGVSRSLLTSNIAVNAPWTSVDIHD 415
              VA  ++  +  L Y QE     S LP G N ++ + RSL TS++AV   P+ + ++
Sbjct: 371 QAAGVAQKHNCPLVRLDYQQEQGLASSLPLGVNQIK-IQRSLTTSSVAVFVPFVTQELFQ 429

Query: 416 KGGK-FYGINQISSNIISIDRGKLNTPSGLILGTSGAGKGMATKHEIISTKLKEADSDTE 474
             G   +YGIN  S N+I  +DR +   P+ L LGT G+GK M+ K EI+S  L   D  +
Sbjct: 430 GGAAMYYGINAKSRNMIMLDRKQARCPNALKLGTPGSGKSMSCKSEIVSVFLTTPD---D 486

Query: 475 IIIVDPENEYSIIGQAFGGESIDIAPDSTTFLNVLELS-DENMDEDPVKVKSEFLLSWIG 533
            I I DPE EY  + +   G+ I  ++P S  F+N L+++ + + D++P+  +KS+F+LS+ +
Sbjct: 487 IFISDPEAEYYPLVKRLHGQVIRLSPTSKDFVNPLDINLNYSEDDNPLALKSDFVLSFCE 546

Query: 534 KLLDRK--MDGREKSLIDRVTRLTYKHF-------DTPSLVEWVFVLSQQPEQEAKDLAL 584
             ++ K  ++  EK++IDR  R+ Y+ +          + P L +    L Q   EA  +A
Sbjct: 547 LVMGGKNGLEAIEKTVIDRAVRVIYRPYLADPRPENMPILSDLHKALLDQHVPEADRVAQ 606

Query: 585 DMELYVEGSLDIFSHRTNIKTDSHFLIYNVKKLGDELKQIALMVIFDQIWNRVVKNQKLG 644
             ++LYV GSL++F+HRTN+   + + +++K+LG +LK++ ++++ DQIW RV  N+  G
Sbjct: 607 ALDLYVSGSLNVFNHRTNVDIGNRLVSFDIKELGKQLKKLGMLIVQDQIWGRVTANRSQG 666

Query: 645 KKTWIYFDEMQLLLLDKYASDFFFKLWSRVRKYGAIPTGITQNVETLLLDANGRRIIANS 704
            K TW + DE  LLL ++  + +  ++W R RK+G IPTG TQNV+ LL       I+ NS
Sbjct: 667 KATWYFADEFHLLLKEEQTAAYSAEIWKRFRKWGGIPTGATQNVKDLLSSPEIENILENS 726

Query: 705 EFMILLKQAKSDREELVHMLGLSKELEKYLVNPEKGAGLIKAGSTVVPFKNKIPQHTKLF 764
            +F+ LL QA  DR+ L   L LS E +KY+ N E G GL+   + V+PF N IP +T+L+
Sbjct: 727 DFITLLNQASGDRKILAERLNLSTEQQKYIDNSEPGEGLLIFENVVLPFTNPIPHNTQLY 786

Query: 765 DIMST 769
             IM+T
Sbjct: 787 KIMTT 791
```

SEQ ID 8926 (GBS75) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 11; MW 89.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 20 (lane 6; MW 114.7kDa).

GBS75-GST was purified as shown in Figure 197, lane 8.

GBS329 was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 8; MW 89kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 174 (lane 2; MW 114kDa).

GBS329-GST was purified as shown in Figure 220, lanes 9 & 10.

Example 1891

A DNA sequence (GBSx1999) was identified in *S.agalactiae* <SEQ ID 5875> which encodes the amino acid sequence <SEQ ID 5876>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2442(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

Example 1892

A DNA sequence (GBSx2000) was identified in *S.agalactiae* <SEQ ID 5877> which encodes the amino acid sequence <SEQ ID 5878>. This protein is predicted to be DNA-directed RNA polymerase ii largest subunit. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4393(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 1893

A DNA sequence (GBSx2001) was identified in *S.agalactiae* <SEQ ID 5879> which encodes the amino acid sequence <SEQ ID 5880>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.92   Transmembrane  256 - 272 ( 250 - 277)
   INTEGRAL    Likelihood = -8.28   Transmembrane  216 - 232 ( 213 - 244)
   INTEGRAL    Likelihood = -8.12   Transmembrane  151 - 167 ( 148 - 191)
   INTEGRAL    Likelihood = -7.27   Transmembrane   57 -  73 (  54 -  80)
   INTEGRAL    Likelihood = -6.74   Transmembrane   93 - 109 (  88 - 111)
   INTEGRAL    Likelihood = -3.50   Transmembrane  172 - 188 ( 168 - 191)
   INTEGRAL    Likelihood = -2.76   Transmembrane  113 - 129 ( 110 - 130)

----- Final Results -----
            bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAG38039 GB:AF295925 Orf23 [Streptococcus pneumoniae]
 Identities = 71/86 (82%), Positives = 83/86 (95%)

Query: 37  VKSLADFNPTVWSYMTAITKGIMQPLGVAILAVVLVLEFSKMAKKIANSGGAMTFEAIAP 96
           +KSL+ +NPTVW+YM++ITK +MQPLGVAIL+VVL+LEFSKMAKKIANSGGAMTFEA+AP
Sbjct: 1   MKSLSSYNPTVWTYMSSITKSVMQPLGVAILSVVLILEFSKMAKKIANSGGAMTFEALAP 60

Query: 97  MIVSYIMVAVVITNTTVIVEAIIAIA 122
           M++SYIMVAVVITNTTVIVEAII IA
Sbjct: 61  MLISYIMVAVVITNTTVIVEAIIGIA 86
```

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 1894

A DNA sequence (GBSx2002) was identified in *S.agalactiae* <SEQ ID 5881> which encodes the amino acid sequence <SEQ ID 5882>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.54   Transmembrane   32 -  48 (  25 -  52)
   INTEGRAL    Likelihood = -4.09   Transmembrane   63 -  79 (  62 -  80)

----- Final Results -----
               bacterial membrane --- Certainty=0.4015(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 9933> which encodes amino acid sequence <SEQ ID 9934> was also identified. A related GBS nucleic acid sequence <SEQ ID 10777> which encodes amino acid sequence <SEQ ID 10778> was also identified.

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

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

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

Example 1895

A DNA sequence (GBSx2003) was identified in *S.agalactiae* <SEQ ID 5883> which encodes the amino acid sequence <SEQ ID 5884>. This protein is predicted to be TrsK-like protein (traK). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.38   Transmembrane   66 -  82 (  62 -  85)

----- 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:AAG38037 GB:AF295925 Orf21 [Streptococcus pneumoniae]
 Identities = 343/457 (75%), Positives = 385/457 (84%), Gaps = 24/457 (5%)
```

```
Query: 142 LIVIGGSGAGKTFRFVKPNLIQLNCSNIVVDPKDHLAEKTGKLFLENGYQVKVLDLVNMT 201
           + VIGGSG+GKTFRFVKPNLIQ+N SNIVVDPKDHLAEKTGKLFLE+GYQVKVLDLVNM
Sbjct: 1   MAVIGGSGSGKTFRFVKPNLIQMNSSNIVVDPKDHLAEKTGKLFLEHGYQVKVLDLVNMK 60

Query: 202 NSDGFNPFRYVETENDLNRMLTVYFNNTKGNGSRSDPFWDEASMTLVRAIASYLVDFYNP 261
           NSDGFNPFRY+ETENDLNRML VYFNNTKG+GSRSDPFWDEASMTLVRA+ASYLVDFYNP
Sbjct: 61  NSDGFNPFRYIETENDLNRMLAVYFNNTKGSGSRSDPFWDEASMTLVRALASYLVDFYNP 120

Query: 262 PGS----------------------SKQEQEARRKRGRYPAFSEIGKLIKLLSKGDNQDKS 300
           P +                        K+E E R+KRGR   F E     +     + KS
Sbjct: 121 PKTREQLIEESRLSQKEYQNLLKRQKKEVEERKKRGRLSKFCESQNSLNTYPRVKTR-KS 179

Query: 301 ILEVLFEDYAKKYGHENFTMRNWADFQNYKDKTLDSVIAVTTAKFALFNIQSVIDLTQRD 360
           +LE+LFE+YAKKYG ENFTMRNWADFQNYKDKTLDSVIAVTTAKFALFNIQSV+DLT+RD
Sbjct: 180 VLEILFENYAKKYGTENFTMRNWADFQNYKDKTLDSVIAVTTAKFALFNIQSVMDLTKRD 239

Query: 361 TMDLKTWGTQKTMVYLVIPDNDTTFRFLSAL-FFSTVFSTLTRQADV-DFKGQLPIHVRS 418
           T+D+KTWG +K+MVYLVIPDND+TFRFLSAL FF+  F T  + + +   + +LP+HVR
Sbjct: 240 TLDMKTWGQEKSMVYLVIPDNDSTFRFLSALLFFNPYFQTPNKTSQILMLRVRLPLHVRV 299

Query: 419 YLDEFANVGEIPDFAEQTSTVRSRNMSLVPILQNIAQLQGLYKEKEAWKTILGNCDSLLY 478
           YLDEFAN+GEIPDFAEQTSTVRSRNMSLVPILQNIAQLQGLYKEKEAWKTILGNCDSL+Y
Sbjct: 300 YLDEFANIGEIPDFAEQTSTVRSRNMSLVPILQNIAQLQGLYKEKEAWKTILGNCDSLVY 359

Query: 479 LGGNDEETFKFMSGLLGKQTVDVRSTSRSFGQTGSSSTSHQKIARDLMTADEVGTMKRDE 538
           LGGNDE+TFKFMSGLLGKQT+DVR+TSRSFGQTGS S SHQKIARDLMT DEVG MKR E
Sbjct: 360 LGGNDEDTFKFMSGLLGKQTIDVRNTSRSFGQTGSGSLSHQKIARDLMTPDEVGNMKRHE 419

Query: 539 CLVRIAGVPVFRTKKYFPLKHKHWKLLADKETDDRWW 575
           CLVRIA +PVF++KKY    KH +WK LA++ETD+R W
Sbjct: 420 CLVRIANMPVFKSKKYNSTKHPNWKYLANQETDERRW 456
```

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

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

A related GBS gene <SEQ ID 8929> and protein <SEQ ID 8930> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:       5.53
GvH: Signal Score (-7.5): -0.78
     Possible site: 34
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 1 value:  -7.38 threshold:  0.0
   INTEGRAL    Likelihood = -7.38   Transmembrane   66 -  82 (  62 -  85)
   PERIPHERAL  Likelihood =  1.75     338
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>
```

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

```
33.9/50.9% over 419aa
                                                          Lactococcus lactis
  GP|3582206| trsK protein (traK) Insert characterized
  PIR|T43089|T43089 transfer complex protein TrsK - plasmid pMRC01 Insert characterized

ORF00383(715 - 2004 of 2415)
```

```
GP|3582206|gb|AAC56002.1||AE001272(23 - 442 of 530) trsK protein (traK) {Lactococcus
lactis}PIR|T43089|T43089 transfer complex prote
in TrsK - Lactococcus lactis plasmid pMRC01
%Match = 10.1
%Identity = 33.8  %Similarity = 50.8
Matches = 141  Mismatches = 193  Conservative Sub.s = 71

519       549       579       609       639       669       699       729
SFLAFILGVLMMTLVYLYVSTGQKVYREGEEYGSARFGTSKEKRNFYSKNPFNDTILARDVRLTLLEKKKPQFDRNKNLI
                                                          |    :         |:|::
                                                        MNGTILGVLDNKIIYQDNTTKPNRNVM
                                                             10        20

759       789       816       846       876       906       936       966
VIGGSGAGKTFRFVKPNLIQLNCSNIVV-DPKDHLAEKTGKLFLENGYQVKVLDLVNMTNSDGFNPFRYVETENDLNRML
||||||: ||   |  ||     ::||| |||  | |||  : |  ||:| |::: || :|| :||| |:| :     :
VIGGSGSYKTQSVVITNLFNETKNSIVVTDPKGELYEKTAGIKLAQGYEVHVVNFANMAHSDRYNPFDYIERDIQAESVA
           40        50        60        70        80        90        100

996      1026      1056      1086      1116      1146      1176      1194
TVYFNNTKGNGSRSDPFWDEASMTLVRAIASYLVDFYNPPGSSKQEQEARRKRGRYPAFSEIGKLIKLLSKGD----NQD
|    :    | :    |      |::|:  :::                  : | |    :  :| :|    |     |:|
TKIVQSENAEGKK--DVWFSTQRQLLKALILFVM-----------------KERSPEQRNLAGVINVLQTFDSEPINKD
          120       130                          140       150       160

1221      1251      1281      1311      1341      1371      1401      1431
K-SILEVLFEDYAKKYGHENFTMRNWADFQNYKDKTLDSVIAVTTAKFALFNIQSVIDLTQRDTMDLKTWGTQKTMVYLV
: | |: ||   | |  |          |:  |     |:|:   |  : |  : | ::|      |:  | :| ::|::
ENSDLDNLF--LALKITHPARIAYELG-FKKAKGDMKASIISSLLATISKFTDEEVSNFTSISDFHLQDIGRKKIVLYVI
            180       190       200       210       220       230       240

1461      1491      1521      1551      1581      1611      1641      1671
IPDNDTTFRFLSALFFSTVFSTLTRQADVDFKGQLPIHVRSYLDEFANVGEIPDFAEQTSTVRSRNMSLVPILQNIAQLQ
||  | |:  :  |||| :|  | : |  :   :||  |   |||| |:|: | : |  :| |   : :  | | : |||
IPVMDNTYESFINLFFSQMFDELYKLASSN-GAKLPQEVDFILDEFVNLGKFPKYEEFLATCRGYGIGVTTICQTLTQLQ
             260       270       280       290       300       310       320

1701      1731      1761      1791      1809      1839      1869      1899
GLYKEKEAWKTILGNCDSLLYLGGNDEETFKFMSGLLGKQTVDVR----STSRSFGQTGSSSTSHQKIARDLMTADEVGT
 ||  ||  ::|||     : |  ::| | |: | |||| || |      ||| |   : | | |:   :| ||| ||:
SLY-GKEKAESILGNHAVKICLNASNEATAKYFSELLGKSTVKVETGSESTSHSKETSTSKSDSYSYTSRQLMTPDEIIR
                  340       350       360       370       380       390       400

1929      1956      1974      2004      2034      2064      2094      2124
MKRDECLVRIAGV-PVFRTK----KYFPLKHKHWKLLADKETDDRWWNYHINPLAKEEELDLSDYQIRDLSTETSLH**K
|   : |:      |:  ||    | ||       ||  :|       :          : :     :::::
MPDTQSLLIFTNQKPIKATKAFQFKLFPDADSKVKLEQNKYVGITSKSQLEKYNDLSVKWEEKLQSLKNITVTEEEEKDL
                  420       430       440       450       460       470       480
```

SEQ ID 5884 (GBS11d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 6; MW 61kDa) and in Figure 182 (lane 10; MW 61kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 12 (lane 5; MW 91.5kDa).

Example 1896

A DNA sequence (GBSx2004) was identified in *S.agalactiae* <SEQ ID 5885> which encodes the amino acid sequence <SEQ ID 5886>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4192(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 9935> which encodes amino acid sequence <SEQ ID 9936> was also identified.

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

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

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

Example 1897

A DNA sequence (GBSx2005) was identified in *S.agalactiae* <SEQ ID 5887> which encodes the amino acid sequence <SEQ ID 5888>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3391(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 1898

A DNA sequence (GBSx2006) was identified in *S.agalactiae* <SEQ ID 5889> which encodes the amino acid sequence <SEQ ID 5890>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.03   Transmembrane   68 -  84 (  64 -  90)
   INTEGRAL    Likelihood = -7.06   Transmembrane   33 -  49 (  27 -  50)
   INTEGRAL    Likelihood = -5.73   Transmembrane  106 - 122 ( 105 - 123)
   INTEGRAL    Likelihood = -4.46   Transmembrane    6 -  22 (   3 -  24)
   INTEGRAL    Likelihood = -2.13   Transmembrane  154 - 170 ( 154 - 170)
   INTEGRAL    Likelihood = -0.53   Transmembrane  180 - 196 ( 180 - 196)

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

A related GBS nucleic acid sequence <SEQ ID 9937> which encodes amino acid sequence <SEQ ID 9938> was also identified.

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

```
>GP:BAA11325 GB:D78257 ORF8 [Enterococcus faecalis]
 Identities = 35/102 (34%), Positives = 57/102 (55%), Gaps = 4/102 (3%)

Query: 90  TRNQAVLVQVGKQVPPIIFLLFL-VNASILEEIVYRQLLWEKLTF--PFEQIGVTSFLFV 146
           T N + L+++   V P++ +L L + A I+EEIV+R  +   L        I ++SFLF
Sbjct: 7   TANDSTLIKLFSGVSPVLVVLLLGIAAPIMEEIVFRGGIIGYLVENNALLAILISSFLFG 66
```

```
Query: 147 LSHGPNQLGSWLIYSCLGLTLAVVRLKT-DCMTAIALHLLWN 187
          + HGP    S+ +Y  +G+ L+V   KT D   +I++H L N
Sbjct: 67  IIHGPTNFISFGMYFFMGIILSVSYYKTKDLRVSISIHFLNN 108
```

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

A related GBS gene <SEQ ID 8931> and protein <SEQ ID 8932> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:      9.32
GvH: Signal Score (-7.5): -5.41
     Possible site: 45
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 6 value: -10.03 threshold:  0.0
   INTEGRAL    Likelihood =-10.03   Transmembrane   68 -  84 (  64 -  90)
   INTEGRAL    Likelihood = -7.06   Transmembrane   33 -  49 (  27 -  50)
   INTEGRAL    Likelihood = -5.73   Transmembrane  106 - 122 ( 105 - 123)
   INTEGRAL    Likelihood = -4.46   Transmembrane    6 -  22 (   3 -  24)
   INTEGRAL    Likelihood = -2.13   Transmembrane  154 - 170 ( 154 - 170)
   INTEGRAL    Likelihood = -0.53   Transmembrane  180 - 196 ( 180 - 196)
   PERIPHERAL  Likelihood =  1.38     131
 modified ALOM score:   2.51

*** Reasoning Step: 3

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

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

```
ORF01326(568 - 861 of 1188)
EGAD|148261|158156(7   -   108   of   120)   hypothetical   protein   (Enterococcus   faecalis)
GP|1402529|dbj|BAA11325.1||D78257 ORF8 (Enterococcus faecalis)
%Match = 5.9
%Identity = 34.7  %Similarity = 60.4
Matches = 35  Mismatches = 37  Conservative Sub.s = 26

303       333       363       393       423       453       483       513
Y*L*RFI*EVTMIRIVLFYLAIQLNGLLVSLFLKEYLTIEGIVLLQLVLLSVTCLEIARHKTVPLKIVGVQNRLSWLLLG

543       573       603       633       660       690       714       744
FVAMVAFAVFISFLFPVQTRNQAVLVQVGKQVPPIIFLLFL-VNASILEEIVYRQLLWEKLT--FPFEQIGVTSFLFVLS
                | | : |:::   | |:: :|:| : | |:||||:|  :   |     :. | ::|||| :
           MQGHTTTANDSTLIKLFSGVSPVLVVLLLGIAAPIMEEIVFRGGIIGYLVENNALLAILISSFLFGII
                    10        20        30        40        50        60

774       804       831       861       891       921       951       981
HGPNQLGSWLIYSCLGLTLAVVRLKT-DCMTAIALHLLWNSLAYVVTFL*YQNQECFRIMEAPYV**GIEKRGGHYVI*T
|||  : |: :|  :|: |:|   || |   :|::|:| |
HGPTNFISFGMYFFMGIILSVSYYKTKDLRVSISIHFLNNLFPAIAIAYGLI
          80        90       100       110       120
```

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

Example 1899

A DNA sequence (GBSx2007) was identified in *S.agalactiae* <SEQ ID 5891> which encodes the amino acid sequence <SEQ ID 5892>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
```

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2490(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 9939> which encodes amino acid sequence <SEQ ID 9940> was also identified.

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

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

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

Example 1900

A DNA sequence (GBSx2008) was identified in *S.agalactiae* <SEQ ID 5893> which encodes the amino acid sequence <SEQ ID 5894>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5298(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:AAC98423 GB:L29323 unknown [Streptococcus pneumoniae]
 Identities = 68/126 (53%), Positives = 88/126 (68%)

Query: 1   MNLLHKKSILDCTELEERIHQAETNQLLQKILSLPNFDCDFEVTFEDDYHKEMNDPLFYE 60
           M  L+K+SILDC ELE  +H AE  QL ++I  +PN+ C+FEVTF DDYHK+ N PLFYE
Sbjct: 1   MKALNKESILDCDELETELHDAEIKQLDEQIFLMPNYPCEFEVTFLDDYHKKHNYPLFYE 60

Query: 61  SNLHQISDFMETRDIKNGVDTLLTKDNHLAFRAFGENYSARGKEGILTTLVTVKCFGEGR 120
           S L  I +F+E++DIKNG D  +    +L F  +G+ Y A GKEGILTT VTVK F E +
Sbjct: 61  SYLQNIMEFLESQDIKNGADAFVDDHQNLVFVLYGQGYRAEGKEGILTTQVTVKAFDEDK 120

Query: 121 MPIDMS 126
            PI+ +
Sbjct: 121 KPINFA 126
```

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 1901

A DNA sequence (GBSx2009) was identified in *S.agalactiae* <SEQ ID 5895> which encodes the amino acid sequence <SEQ ID 5896>. This protein is predicted to be methyl transferase. Analysis of this protein sequence reveals the following:

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

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

```
bacterial cytoplasm --- Certainty=0.1209(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:AAC98421 GB:L29323 methyl transferase [Streptococcus pneumoniae]
 Identities = 323/449 (71%), Positives = 389/449 (85%), Gaps = 3/449 (0%)

Query: 1   MKFLDLFAGIGGFRLGMESQGHKCLGFCEIDKFARTSYKAMFNTEGEIEYHDIKEVTDHD 60
           M+F+DLF+GIGGFRLGMES GH+C+GFCEIDKFAR SYK++F TEGEIE+HDI++V+D +
Sbjct: 1   MRFIDLFSGIGGFRLGMESVGHECIGFCEIDKFARESYKSIFQTEGEIEFHDIRDVSDDE 60

Query: 61  FRQFRGQVDIICGGFPCQAFSLAGRRLGFEDTRGTLFFEIARAAKQIQPRFLFLENVKGL 120
           F++ RG+VD+ICGGFPCQAFS+AGRRLGFEDTRGTLFFEIARAAKQIQPRFLFLENVKGL
Sbjct: 61  FKKLRGKVDVICGGFPCQAFSIAGRRLGFEDTRGTLFFEIARAAKQIQPRFLFLENVKGL 120

Query: 121 LNHDEGRTFATILSTLDELGYDVEWQVLNSKDFQVPQNRERVFIIGHSRRYRSRFIFPLR 180
           LNHD+GRTF TIL+TLDELG+DVEWQ+LNSKDF VPQNRERVFIIGHSR+  +R  FP R
Sbjct: 121 LNHDKGRTFTTILTTLDELGFDVEWQMLNSKDFGVPQNRERVFIIGHSRKRGTRLGFPFR 180

Query: 181 RED---SPAHLERLGNINPSKHGLNGEVYLTSGLAPTLTRGKGEGAKIAIPVLTPDRLEK 237
           RE    +P  L+ LGN+NPSK G++G+VY + GLAPTL RGKGEG KIAIP +TPDRL+K
Sbjct: 181 REGQATNPETLKILGNLNPSKSGMSGKVYYSEGLAPTLVRGKGEGFKIAIPCMTPDRLDK 240

Query: 238 RQHGRRFKDNQDPMFTLTSQDKHGVVVAGNLPTSFDQTGRVFDISGLSPTLTTMQGGDKV 297
           RQ+GRRFKDNQ+PMFTL +QD+HG+VV G+LPTSF +TGRV+   GLSPTLTTMQGGDK+
Sbjct: 241 RQNGRRFKDNQEPMFTLNTQDRHGIVVVGDLPTSFKETGRVYGSEGLSPTLTTMQGGDKI 300

Query: 298 PKILLREELPFLKIKEATKTGYAKATLGDSVNLAYPDSTKRRGRVGKGISNTLTTSDNMG 357
           PKIL+ E + FLK++EATK GYA+A +GDS+NL  P S  RRGRVGKGI+NTLTTS  MG
Sbjct: 301 PKILIPEPIQFLKVREATKKGYAQAEIGDSINLERPSSQHRRGRVGKGIANTLTTSGQMG 360

Query: 358 VVVAALEYRQDKWYEVTGIVLEGKLYRLRIRRLTPRECFRLQGFPDWAYERAESVSSKSQ 417
           VVVA+ E     + Y+V G++++G+ YRLRIRR+TP+ECFRLQGFPDWA+E A  VSS SQ
Sbjct: 361 VVVASYEGEDKQVYQVAGVLIDGQFYRLRIRRITPKECFRLQGFPDWAFEAARKVSSNSQ 420

Query: 418 LYKQAGNSVTVTVIEAIAREFRRTEEEEK 446
           LYKQAGNSVTV VI AIA++ +  EE+++
Sbjct: 421 LYKQAGNSVTVPVIAAIAKKLKEVEEKDE 449
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2435> which encodes the amino acid sequence <SEQ ID 2436>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1725(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 = 60/75 (80%), Positives = 69/75 (92%)

Query: 1  MKFLDLFAGIGGFRLGMESQGHKCLGFCEIDKFARTSYKAMFNTEGEIEYHDIKEVTDHD 60
          MKFLDLFAGIGGFRLG+ +Q H+C+GFCEIDKFAR SYKA++ TEGEIE+HDI++VTD D
Sbjct: 4  MKFLDLFAGIGGFRLGLINQCHECIGFCEIDKFARQSYKAIYETEGEIEFHDIRQVTDQD 63

Query: 61 FRQFRGQVDIICGGF 75
          FRQ RGQVDIICGGF
Sbjct: 64 FRQLRGQVDIICGGF 78
```

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

Example 1902

A DNA sequence (GBSx2010) was identified in *S.agalactiae* <SEQ ID 5897> which encodes the amino acid sequence <SEQ ID 5898>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.71   Transmembrane    8 -  24 (   3 -  30)

----- Final Results -----
               bacterial membrane --- Certainty=0.4885(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 9941> which encodes amino acid sequence <SEQ ID 9942> 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 5899> which encodes the amino acid sequence <SEQ ID 5900>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.81   Transmembrane   20 -  36 (  19 -  36)

----- Final Results -----
               bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 16/33 (48%), Positives = 26/33 (78%)

Query: 1  MNKMIWWILGGIYLISIIILIVEIIRAPEMDDH 33
          ++KM WW+L G++ +  I LI+E+I APEM+D+
Sbjct: 12 VSKMFWWLLLGVWGLRTIWLIIEVITAPEMEDY 44
```

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

Example 1903

A DNA sequence (GBSx2011) was identified in *S.agalactiae* <SEQ ID 5901> which encodes the amino acid sequence <SEQ ID 5902>. This protein is predicted to be ifn-response binding factor 1 (irebf-1). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4771(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:AAD41248 GB:AF106927 unknown [Streptococcus suis]
 Identities = 258/272 (94%), Positives = 266/272 (96%)

Query: 1   MKRITANQYQTSERYYKLPKILFESERYKDMKLEVKVAYAVLKDRLELSLSKGWIDEDGA 60
           MKRITANQYQTSERYYKLPKILFESERYKDMKLEVKVAYAVLKDRLELSLSKGWIDEDGA
```

```
Sbjct: 1   MKRITANQYQTSERYYKLPKILFESERYKDMKLEVKVAYAVLKDRLELSLSKQWIDEDGA 60

Query: 61  IYLIYSNSNLMALLGCSKSKLLSIKKTLREYGLIDEVQQSSSERGRMANKIYLGELEHEP 120
           IYLIYSNSNLMALLGCSKSKLLSIKKTLREYGLIDEVQQSSSE+GRMANKIYLGELEHE
Sbjct: 61  IYLIYSNSNLMALLGCSKSKLLSIKKTLREYGLIDEVQQSSSEKGRMANKIYLGELEHET 120

Query: 121 TPVLHTDGASVKKTLGESQRKTGPVLYSAPSETEGSETKYSETEGSDLVMKDEEERQLVD 180
           TPVLHTDGASVKKTLG SQRKTGPVL SAPSETEGSETKYSET+GSD +++DEEERQ VD
Sbjct: 121 TPVLHTDGASVKKTLGGSQRKTGPVLNSAPSETEGSETKYSETKGSDFLIEDEEERQQVD 180

Query: 181 EKKEENFTSKVDGVTKYDRDYIWGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALEQMRF 240
           EK+EENFTSKVDGVT+YDRDYIWGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALE MRF
Sbjct: 181 EKQEENFTSKVDGVTRYDRDYIWGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALEHMRF 240

Query: 241 ARSAEVIAEYVFNGVLSEWTKQLRRQEVKGGE 272
           ARSAEVIAEYVFNGVLSEWTKQLRRQEVKGG+
Sbjct: 241 ARSAEVIAEYVFNGVLSEWTKQLRRQEVKGGD 272
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5903> which encodes the amino acid sequence <SEQ ID 5904>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5248(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 = 84/122 (68%), Positives = 99/122 (80%), Gaps = 2/122 (1%)

Query: 145 VLYSAPSETEGSETKYSETEGSDLVMKDEEERQLVD--EKKEENFTSKVDGVTKYDRDYI 202
           VL SAPSETE SET+ SET+ S+LV++DEEER+      +K E +FT +VD VTKYD+DYI
Sbjct: 1   VLNSAPSETEKSETEGSETKESNLVIEDEEERKECTSVKKTEGHFTRQVDQVTKYDKDYI 60

Query: 203 WGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALEQMRFARSAEVIAEYVFNGVLSEWTKQ 262
           W LVH QLR+ GLSQ+ASD  M YF +RY YALE +RFAR+AE IAEYVFNGVLSEWTKQ
Sbjct: 61  WSLVHSQLREGGLSQAASDLVMSYFEERYAYALEHIRFARTAEAIAEYVFNGVLSEWTKQ 120

Query: 263 LR 264
           LR
Sbjct: 121 LR 122
```

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

Example 1904

A DNA sequence (GBSx2012) was identified in *S.agalactiae* <SEQ ID 5905> which encodes the amino acid sequence <SEQ ID 5906>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4191(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 9375> which encodes amino acid sequence <SEQ ID 9376> was also identified.

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

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

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

Example 1905

A DNA sequence (GBSx2013) was identified in *S.agalactiae* <SEQ ID 5907> which encodes the amino acid sequence <SEQ ID 5908>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3723(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 1906

A DNA sequence (GBSx2014) was identified in *S.agalactiae* <SEQ ID 5909> which encodes the amino acid sequence <SEQ ID 5910>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3053(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 1907

A DNA sequence (GBSx2015) was identified in *S.agalactiae* <SEQ ID 5911> which encodes the amino acid sequence <SEQ ID 5912>. This protein is predicted to be 50S ribosomal protein L7/l12 (rplL). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1034(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 9943> which encodes amino acid sequence <SEQ ID 9944> was also identified.

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

```
>GP:CAB11881 GB:Z99104 ribosomal protein L12 (BL9) [Bacillus subtilis]
 Identities = 83/123 (67%), Positives = 95/123 (76%), Gaps = 2/123 (1%)

Query: 6   MALNIENIIAEIKEATILELNDLVKAIEEEFGVTAAAPVAAA--AAGGEAAAAKDSFDVE 63
           MALNIE IIA +KEAT+LELNDLVKAIEEEFGVTAAAPVA A  AA G AA  +  FD+
Sbjct: 1   MALNIEEIIASVKEATVLELNDLVKAIEEEFGVTAAAPVAVAGGAAAGGAAEEQSEFDLI 60

Query: 64  LTAAGDKKVGVIKVVREITGEGLKEAKAIVDNAPSVIKEGASEAEANEIKEKLEAAGASV 123
           L  AG +K+ VIKVVREITG GLKEAK +VDN P  +KEG ++ EA E+K KLE  GASV
Sbjct: 61  LAGAGSQKIKVIKVVREITGLGLKEAKELVDNTPKPLKEGIAKEEAEELKAKLEEVGASV 120

Query: 124 TLK 126
           +K
Sbjct: 121 EVK 123
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5913> which encodes the amino acid sequence <SEQ ID 5914>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1164(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 = 104/126 (82%), Positives = 113/126 (89%)

Query: 1   MEEITMALNIENIIAEIKEATILELNDLVKAIEEEFGVTAAAPVAAAAAGGEAAAAKDSF 60
           +EEITMALNIENIIAEIKEA+ILELNDLVKAIEEEFGVTAAAPVAAAAAGG   AAKDSF
Sbjct: 1   LEEITMALNIENIIAEIKEASILELNDLVKAIEEEFGVTAAAPVAAAAAGGAEEAAKDSF 60

Query: 61  DVELTAAGDKKVGVIKVVREITGEGLKEAKAIVDNAPSVIKEGASEAEANEIKEKLEAAG 120
           DVELT+AGDKKVGVIK VREITG GLKEAK +VD AP+ +KEG + AEA EIK KLE AG
Sbjct: 61  DVELTSAGDKKVGVIKAVREITGLGLKEAKGLVDGAPANVKEGVAAAEAEEIKAKLEEAG 120

Query: 121 ASVTLK 126
           A++TLK
Sbjct: 121 ATITLK 126
```

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

Example 1908

A DNA sequence (GBSx2017) was identified in *S.agalactiae* <SEQ ID 5915> which encodes the amino acid sequence <SEQ ID 5916>. This protein is predicted to be ribosomal protein L10 (rplJ). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1251(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:CAB11880 GB:Z99104 ribosomal protein L10 (BL5) [Bacillus subtilis]
 Identities = 96/164 (58%), Positives = 125/164 (75%), Gaps = 1/164 (0%)

Query: 14  MSEAIIAKKAEQVELIAEKMKAAASIVVVDSRGLTVEQDTNLRRSLRESDVEFKVIKNSI 73
           MS AI  KK   VE IA K+K + S ++VD RGL V + T LR+ LRE++VE KV KN++
Sbjct: 1   MSSAIETKKVV-VEEIASKLKESKSTIIVDYRGLNVSEVTELRKQLREANVESKVYKNTM 59

Query: 74  LTRAAEKAGLEDLKELFVGPSAVAFSNEDVIAPAKVISDFAKDAEALEIKGGSVDGKFTS 133
             RA E+A L  L +   GP+A+AFS EDV+APAKV++DFAK+ EALEIK G ++GK ++
Sbjct: 60  TRRAVEQAELNGLNDFLTGPNAIAFSTEDVVAPAKVLNDFAKNHEALEIKAGVIEGKVST 119

Query: 134 VEEINALAKLPNKEGMLSMLLSVLQAPVRNVAYAVKAVAEKDEE 177
           VEE+ ALA+LP +EG+LSMLLSVL+APVRN+A A KAVAE+ EE
Sbjct: 120 VEEVKALAELPPREGLLSMLLSVLKAPVRNLALAAKAVAEQKEE 163
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5917> which encodes the amino acid sequence <SEQ ID 5918>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood = -5.47   Transmembrane     7 -  23 (   5 -  24)

----- 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>
```

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

```
 Identities = 149/176 (84%), Positives = 162/176 (91%)

Query: 4   SQKIKTEVKLMSEAIIAKKAEQVELIAEKMKAAASIVVVDSRGLTVEQDTNLRRSLRESD 63
           S KIKTEVKLMSEAIIAKKAEQVELIAEKMKAAASIV+VDSRGLTV+QDT LRRSLRES
Sbjct: 23  SPKIKTEVKLMSEAIIAKKAEQVELIAEKMKAAASIVIVDSRGLTVDQDTVLRRSLRESG 82

Query: 64  VEFKVIKNSILTRAAEKAGLEDLKELFVGPSAVAFSNEDVIAPAKVISDFAKDAEALEIK 123
           VEFKVIKNSILTRAAEKAGL++LK++FVGPSAVAFSNEDVIAPAKVI+DF K A+ALEIK
Sbjct: 83  VEFKVIKNSILTRAAEKAGLDELKDVFVGPSAVAFSNEDVIAPAKVINDFTKTADALEIK 142

Query: 124 GGSVDGKFTSVEEINALAKLPNKEGMLSMLLSVLQAPVRNVAYAVKAVAEKDEEVA 179
           GG+++G  +S EEI ALA LPN+EGMLSMLLSVLQAPVRNVAYAVKAVAE  E  A
Sbjct: 143 GGAIEGAVSSKEEIQALATLPNREGMLSMLLSVLQAPVRNVAYAVKAVAENKEGAA 198
```

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

Example 1909

A DNA sequence (GBSx2018) was identified in *S.agalactiae* <SEQ ID 5919> which encodes the amino acid sequence <SEQ ID 5920>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood = -7.22   Transmembrane  125 - 141 ( 106 - 143)
   INTEGRAL     Likelihood = -1.91   Transmembrane  108 - 124 ( 106 - 124)

----- Final Results -----
               bacterial membrane --- Certainty=0.3888(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 10931> which encodes amino acid sequence <SEQ ID 10932> was also identified.

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

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

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

Example 1910

A DNA sequence (GBSx2019) was identified in *S.agalactiae* <SEQ ID 5921> which encodes the amino acid sequence <SEQ ID 5922>. This protein is predicted to be Clp-like ATP-dependent protease binding subunit (clpC). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3483(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:AAA68910 GB:L34677 Clp-like ATP-dependent protease binding
          subunit [Bos taurus]
 Identities = 437/589 (74%), Positives = 514/589 (87%), Gaps = 5/589 (0%)

Query: 10  DPFGN-MDDIFNSLMGNMGGYNSENKRYLINGREVTPEEFSQYRQTGKLPGQELNNQNTP 68
           DPF N MDD+FN LMG M G NSEN+RYLINGREVTPEE++ +RQTGKLPG      Q
Sbjct: 2   DPFNNDMDDLFNQLMGGMNGVNSENRRYLINGREVTPEEYAAFRQTGKLPGVTDPTQ-AK 60

Query: 69  TNQVSADSVLTKLGTNLTDQARQHLLDPVIGRNKEIQETAEILARRTKNNPVLVGDAGVG 128
           T Q   DS+L KLG NLT +A++  LDPVIGRNKEIQETAEIL+RRTKNNPVLVGDAGVG
Sbjct: 61  TKQPQPDSMLAKLGRNLTQEAKEGKLDPVIGRNKEIQETAEILSRRTKNNPVLVGDAGVG 120

Query: 129 KTAVIEGLAQAIINGDVPAAIKNKEIISIDISSLEAGTQYRGSFEENIQNIIKEVKETGN 188
           KTAV+EGLAQAI+ GDVPAAIKNK+IISIDISSLEAGTQYRGSFEEN+Q +I EVK+ GN
Sbjct: 121 KTAVVEGLAQAIVAGDVPAAIKNKQIISIDISSLEAGTQYRGSFEENMQKLIDEVKKDGN 180

Query: 189 IILFFDEIHQILGAGSTGGDSGSKGLADILKPALSRGELTVIGATTQDEYRNTILKNAAL 248
           +ILFFDEIHQI+GAG+ G  SGSKG+ADILKPALSRGE+T+IGATTQDEYRNTILK+AAL
Sbjct: 181 VILFFDEIHQIIGAGNAGDASGSKGMADILKPALSRGEVTLIGATTQDEYRNTILKDAAL 240

Query: 249 ARRFNEVKVNAPSAQDTFNILMGIRNLYEQHHNVVLPDSVLKAAVDLSIQYIPQRSLPDK 308
           +RRFN+V VNAPS +DTF IL G+R LYE+HHNV LPD VLKAA+D S+QYIPQRSLPDK
Sbjct: 241 SRRFNQVTVNAPSKEDTFKILQGLRKLYEKHHNVSLPDEVLKAAIDYSVQYIPQRSLPDK 300

Query: 309 AIDLIDMTAAHLAAQHPVTDLKSLEKEIAEQRDKQEKAVNTEDFEEALKVKTRIEELQNQ 368
           AIDLID+TAAHLA++HPV D K++E+EI +   KQ++AV  ED++ A + K ++ +LQ+Q
Sbjct: 301 AIDLIDVTAAHLASKHPVKDAKTIEEEIKKTEAKQQEAVEKEDYQAAQEAKDQVAKLQDQ 360

Query: 369 IDNHTEGQKVTATINDIAMSIERLTGVPVSNMGASDIERLKELGNRLKGKVIGQNDAVEA 428
           + +H+E ++V AT +D+A. ++ER+TG+PVS MGASDIERLK L  RL+GKVIGQ +AVEA
Sbjct: 361 LKDHSESERVVATPSDVAAAVERMTGIPVSKMGASDIERLKGLATRLEGKVIGQQEAVEA 420

Query: 429 VARAIRRNRAGFDDGNRPIGSFLFVGPTGVGKTELAKQLAFDMFGSKDAIVRLDMSEYND 488
           V+RAIRRNRAGFD+GNRPIGSFLFVGPTGVGKTELAKQLA DMFGS + I+RLDMSEY D
Sbjct: 421 VSRAIRRNRAGFDEGNRPIGSFLFVGPTGVGKTELAKQLALDMFGSTNDIIRLDMSEYTD 480

Query: 489 RTAVSKLIGATAGYVGYDDNSNTLTERIRRNPYSIVLLDEIEKADPQVITLLLQVLDDGR 548
           RTAVSKLIG TAGYVGYDDNSNTLTE++RR+PYSIVLLDEIEKA+PQVITLLLQVLDDGR
```

```
Sbjct: 481 RTAVSKLIGTTAGYVGYDDNSNTLTEKVRRHPYSIVLLDEIEKANPQVITLLLQVLDDGR 540

Query: 549 LTDGQGNTINFKNTVIIATSNAGFGNEAFTGDSDKDLKIMERISPYFRP 597
           LTDGQGNT++FKNT+IIATSNAGF ++A  G+   D K+M+++ PYFRP
Sbjct: 541 LTDGQGNTVDFKNTIIIATSNAGFSSDAVAGE---DAKLMDKLQPYFRP 586
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5923> which encodes the amino acid sequence <SEQ ID 5924>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2718(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 = 551/697 (79%), Positives = 616/697 (88%), Gaps = 3/697 (0%)

Query: 5   NFYNRDPFGNMDDIFNSLMGNMGGYNSENKRYLINGREVTPEEFSQYRQTGKLPGQELNN 64
           +F  +DPF NMDDIFN LM NMGGY SEN RYL+NGRE+TPEEF  YRQTG+LP
Sbjct: 3   HFSGKDPFVNMDDIFNQLMANMGGYRSENPRYLVNGREITPEEFQHYRQTGQLPVATTKA 62

Query: 65  QNTPTNQVSADSVLTKLGTNLTDQARQHLLDPVIGRNKEIQETAEILARRTKNNPVLVGD 124
            N+      ADSVLT+LGTNLT +ARQ  LDPVIGRNKEIQ+TAEILARRTKNNPVLVGD
Sbjct: 63  TNSQMLTPKADSVLTQLGTNLTQEARQGHLDPVIGRNKEIQDTAEILARRTKNNPVLVGD 122

Query: 125 AGVGKTAVIEGLAQAIINGDVPAAIKNKEIISIDISSLEAGTQYRGSFEENIQNIIKEVK 184
           AGVGKTAVIEGLAQAI+NGDVPAAIKNKEI+SIDISSLEAGTQYRGSFEE IQN+I+EVK
Sbjct: 123 AGVGKTAVIEGLAQAIVNGDVPAAIKNKEIVSIDISSLEAGTQYRGSFEETIQNLIQEVK 182

Query: 185 ETGNIILFFDEIHQILGAGSTGGDSGSKGLADILKPALSRGELTVIGATTQDEYRNTILK 244
           E GNIILFFDEIHQI+GAG+T  DSGSKGLADILKPALSRGELT+IGATTQDEYRNTILK
Sbjct: 183 EAGNIILFFDEIHQIVGAGATSSDSGSKGLADILKPALSRGELTLIGATTQDEYRNTILK 242

Query: 245 NAALARRFNEVKVNAPSAQDTFNILMGIRNLYEQHHNVVLPDSVLKAAVDLSIQYIPQRS 304
           NAALARRFNEVKVNAPSA+DTF+ILMGIRNLYEQHH++ LPD+VLKAAVD SIQYIPQRS
Sbjct: 243 NAALARRFNEVKVNAPSAEDTFHILMGIRNLYEQHHHITLPDNVLKAAVDYSIQYIPQRS 302

Query: 305 LPDKAIDLIDMTAAHLAAQHPVTDLKSLEKEIAEQRDKQEKAVNTEDFEEALKVKTRIEE 364
           LPDKAIDL+DMTAAHLAAQHPVTDLK+LE EIA+Q++ QEKAV  EDFE+AL  KTRIE
Sbjct: 303 LPDKAIDLLDMTAAHLAAQHPVTDLKTLETEIAKQKESQEKAVAKEDFEKALAAKTRIET 362

Query: 365 LQNQIDNHTEGQKVTATINDIAMSIERLTGVPVSNMGASDIERLKELGNRLKGKVIGQND 424
           LQ QI+ H + Q VTAT+NDIA S+ERLTG+PVSNMG +D+ERLK + +RLK  VIGQ++
Sbjct: 363 LQKQIEQHNQSQNVTATVNDIAESVERLTGIPVSNMGTNDLERLKGISSRLKSHVIGQDE 422

Query: 425 AVEAVARAIRRNRAGFDDGNRPIGSFLFVGPTGVGKTELAKQLAFDMFGSKDAIVRLDMS 484
           AV AVARAIRRNRAGFDDG RPIGSFLFVGPTGVGKTELAKQLA D+FGSKDAI+RLDMS
Sbjct: 423 AVAAVARAIRRNRAGFDDGKRPIGSFLFVGPTGVGKTELAKQLALDLFGSKDAIIRLDMS 482

Query: 485 EYNDRTAVSKLIGATAGYVGYDDNSNTLTERIRRNPYSIVLLDEIEKADPQVITLLLQVL 544
           EYNDRTAVSKLIG TAGYVGYDDN+NTLTER+RRNPY+IVLLDEIEKADPQ+ITLLLQVL
Sbjct: 483 EYNDRTAVSKLIGTTAGYVGYDDNNNTLTERVRRNPYAIVLLDEIEKADPQIITLLLQVL 542

Query: 545 DDGRLTDGQGNTINFKNTVIIATSNAGFGNEAFTGDSDKDLKIMERISPYFRPEFLNRFN 604
           DDGRLTDGQGNTINFKNTVIIATSNAGFG +         +  IM+RI+PYFRPEFLNRFN
Sbjct: 543 DDGRLTDGQGNTINFKNTVIIATSNAGFGQQ---DTETSESNIMDRIAPYFRPEFLNRFN 599

Query: 605 GVIEFSHLSKDDLSEIVDLMLDEVNQTIGKKGIDLVVDENVKSHLIELGYDEAMGVRPLR 664
            +I+F+HL K+ L EIVDLML EVNQT  KKGI L + ++ K+HLI+LGY+ AMG RPLR
Sbjct: 600 SIIKFNHLQKESLEEIVDLMLAEVNQTTAKKGISLTITDDAKAHLIDLGYNHAMGARPLR 659

Query: 665 RVIEQEIRDRITDYYLDHTDVKHLKANLQDGQIVISE 701
           R+IEQEIRDRITDYYLDH +VK L+A L++GQ+VI +
```

```
Sbjct: 660 RIIEQEIRDRITDYYLDHPEVKKLQAILKEGQLVIRQ 696
```

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

Example 1911

A DNA sequence (GBSx2020) was identified in *S.agalactiae* <SEQ ID 5925> which encodes the amino acid sequence <SEQ ID 5926>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -4.78   Transmembrane    8 -  24 (   7 -  25)

----- Final Results -----
               bacterial membrane --- Certainty=0.2911(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 9945> which encodes amino acid sequence <SEQ ID 9946> was also identified.

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

```
>GP:AAC73364 GB:AE000134 putative enzyme [Escherichia coli K12]
 Identities = 142/307 (46%), Positives = 195/307 (63%), Gaps = 6/307 (1%)

Query: 39  KELLESKKTLILHGALGTELESRGCDVSGKLWSAKYLIEDPAAIQTIHEDYIRAGADIVT 98
           + LL+ +  L+L GA+ TELE+RGC+++   LWSAK L+E+P  I+ +H DY RAGA
Sbjct: 8   RALLDKQDILLLDGAMATELEARGCNLADSLWSAKVLVENPELIREVHLDYYRAGAQCAI 67

Query: 99  TSTYQATLQGLAQVGVSESQTEDLIRLTVQLAKAAREQVWKSLTKEEKSERIYPLISGDV 158
           T++YQAT  G A  G+ E+Q++ LI  +V+LA+ ARE      L +  ++  +  L++G V
Sbjct: 68  TASYQATPAGFAARGLDEAQSKALIGKSVELARKAREAY---LAENPQAGTL--LVAGSV 122

Query: 159 GPYAAFLADGSEYTGLYDIDKQGLKNFHRHRIELLLDEGVDILALETIPNAQEAEALIEL 218
           GPY A+LADGSEY G Y    + + FHR R+E LLD G D+LA ET+PN  E EAL EL
Sbjct: 123 GPYGAYLADGSEYRGDYHCSVEAFQAFHRPRVEALLDAGADLLACETLPNFSEIEALAEL 182

Query: 219 LAEDFPQVEAYMSFTSQDGKTISDGSAVADLAKAIDVSPQVVALGINCSSPSLVADFLQA 278
           L   +P+  A+ SFT +D + +SDG+ + D+    +   PQVVALGINC +      LQ
Sbjct: 183 LTA-YPRARAWFSFTLRDSEHLSDGTPLRDVVALLAGYPQVVALGINCIALENTTAALQH 241

Query: 279 IAEQTNKPLVTYPNSGEVYDGASQSWQSSPDHSHTLLENTSDWQKLGAQVVGGCCRTRPA 338
           +   T  PLV YPNSGE YD  S++W    +H   L +    WQ  GA+++GGCCRT PA
Sbjct: 242 LHGLTVLPLVVYPNSGEHYDAVSKTWHHHGEHCAQLADYLPQWQAAGARLIGGCCRTTPA 301

Query: 339 DIADLSA 345
           DIA L A
Sbjct: 302 DIAALKA 308
```

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

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

A related GBS gene <SEQ ID 8933> and protein <SEQ ID 8934> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:      5.48
GvH: Signal Score (-7.5): -2.64
     Possible site: 20
```

```
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value;  -4.78 threshold:  0.0
   INTEGRAL    Likelihood = -4.78   Transmembrane    8 -  24 (   7 -  25)
   PERIPHERAL  Likelihood =  2.49     259
 modified ALOM score:   1.46

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2911(Affirmative) < succ>
                bacterial outside --- 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:

```
ORF01312(412 - 1338 of 1644)
OMNI|NT01EC0303(55 - 357 of 358) conserved hypothetical protein
%Match = 23.8
%Identity = 46.6  %Similarity = 64.3
Matches = 142  Mismatches = 107  Conservative Sub.s = 54

288       318       348       378       408       438       468       498
LISQSFCS*FRL*GLLGIAHNVLGFTSVFHLLFSAIFITNYVTRNGDLMGRFKELLESKKTLILHGALGTELESRGCDVS
                                                  :: ||: :  |:| ||: ||||:|||:::
AWWPVLGWHSIQRRELRCGAGYRLLRCAMVLISLLNPETQNRSQNMSQNNPLRALLDKQDILLLDGAMATELEARGCNLA
                 20        30        40        50        60        70        80

528       558       588       618       648       678       708       738
GKLWSAKYLIEDPAAIQTIHEDYIRAGADIVTTSTYQATLQGLAQVGVSESQTEDLIRLTVQLAKAAREQVWKSLTKEEK
  ||||| |:|:|  |: :| || ||||    |::|||| |:|  |: |:|:: ||  :|:||: |||         |
DSLWSAKVLVENPELIREVHLDYYRAGAQCAITASYQATPAGFAARGLDEAQSKALIGKSVELARKARE-----AYLAEN
               100       110       120       130       140       150

768       798       828       858       888       918       948       978
SERIYPLISGDVGPYAAFLADGSEYTGLYDIDKQGLKNFHRHRIELLLDEGVDILALETIPNAQEAEALIELLAEDFPQV
 :    |::| |||| |:|||||| | |    : :: ||| |:| ||| | |:|| ||:||  | ||| |||   :|:
PQAGTLLVAGSVGPYGAYLADGSEYRGDYHCSVEAFQAFHRPRVEALLDAGADLLACETLPNFSEIEALAELLT-AYPRA
         170       180       190       200       210       220       230

1008      1038      1068      1098      1128      1158      1188      1218
EAYMSFTSQDGKTISDGSAVADLAKAIDVSPQVVALGINCSSPSLVADFLQAIAEQTNKPLVTYPNSGEVYDGASQSWQS
 |: ||| :| : :|||: : |:   :  |||||||||| :      || :   |  ||| ||||| ||  |::|:
RAWFSFTLRDSEHLSDGTPLRDVVALLAGYPQVVALGINCIALENTTAALQHLHGLTVLPLVVYPNSGEHYDAVSKTWHH
         250       260       270       280       290       300       310

1248      1278      1308      1338      1368      1398      1428      1458
SPDHSHTLLENTSDWQKLGAQVVGGCCRTRPADIADLSAHLK*VKYLEEG*GKFDFLFQSTRKPAWILPNGFCFYLSEMT
 :|   | :    ||  ||:::|||||| ||||| | |
HGEHCAQLADYLPQWQAAGARLIGGCCRTTPADIAALKARS
         330       340       350
```

SEQ ID 8934 (GBS381) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 6; MW 42kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 4; MW 66.9kDa).

Example 1912

A DNA sequence (GBSx2021) was identified in *S.agalactiae* <SEQ ID 5927> which encodes the amino acid sequence <SEQ ID 5928>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2996(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 1913

A DNA sequence (GBSx2022) was identified in *S.agalactiae* <SEQ ID 5929> which encodes the amino acid sequence <SEQ ID 5930>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.62   Transmembrane  176 - 192 ( 168 - 198)
   INTEGRAL    Likelihood =-11.57   Transmembrane   89 - 105 (  80 - 111)
   INTEGRAL    Likelihood =-10.03   Transmembrane  337 - 353 ( 332 - 359)
   INTEGRAL    Likelihood = -9.87   Transmembrane  292 - 308 ( 285 - 316)
   INTEGRAL    Likelihood = -4.51   Transmembrane   58 -  74 (  52 -  78)
   INTEGRAL    Likelihood = -3.88   Transmembrane  267 - 283 ( 267 - 286)
   INTEGRAL    Likelihood = -3.08   Transmembrane  125 - 141 ( 125 - 142)
   INTEGRAL    Likelihood = -2.13   Transmembrane  212 - 228 ( 212 - 228)

----- Final Results -----
             bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9377> which encodes amino acid sequence <SEQ ID 9378> was also identified.

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

```
>GP:CAB12034 GB:Z99105 similar to histidine permease [Bacillus subtilis]
 Identities = 221/384 (57%), Positives = 291/384 (75%), Gaps = 2/384 (0%)

Query: 2   PVTGSFHTYATKFISPGTGFTVAWLYWICWTVALGTEFLGAAMLMQRWFPNVPAWAFASF 61
           PVTG+FHTYA K+I PGTGFTVAWLYW+ WTVALG+EF  A +LMQRWFP+   W +++
Sbjct: 76  PVTGAFHTYAAKYIGPGTGFTVAWLYWLIWTVALGSEFTAAGLLMQRWFPHTSVWMWSAV 135

Query: 62  FALVIFGLNALSVRFFAEAESFFSSIKVIAIIIFIILGLGAMFGLVSFEGQHKAILFTHL 121
           FAL IF LNA SV+FFAE+E +FSSIKV+AI++FI+LG  AMFG++  +G   A + ++
Sbjct: 136 FALFIFLLNAFSVKFFAESEFWFSSIKVLAIVLFILLGGSAMFGIIPIKGGEAAPMLSNF 195

Query: 122 TANGA-FPNGIVAVVSVMLAVNYAFSGTELIGIAAGETDNPKEAVPRAIKTTIGRLVVFF 180
           TA G  FPNG V ++  ML+VN+AFSGTELIGIAAGE+ +P + +P+AIKTT+ RL +FF
Sbjct: 196 TAEGGLFPNGFVPILMTMLSVNFAFSGTELIGIAAGESVDPDKTIPKAIKTTVWRLSLFF 255

Query: 181 VLTIVVLASLLPMKEAGVSTAPFVDVFDKMGIPFTADIMNFVILTAILSAGNSGLYASSR 240
           V TI VL+ L+P+++AGV  +PFV VFD++G+P+ ADIMNFVILTAILSA NSGLYASSR
Sbjct: 256 VGTIFVLSGLIPIQDAGVIKSPFVAVFDRVGVPYAADIMNFVILTAILSAANSGLYASSR 315

Query: 241 MLWSLANEGMLSKSVVKINKHGVPMRALLLSMAGAVLSLFSSIYAADTVYLALVSIAGFA 300
           MLWSL+ E  L  +  K+   G P  AL+ SM G +LSL SS++A DTVY+ LVSI+GFA
Sbjct: 316 MLWSLSKEKTLHPTFAKLTSKGTPFNALVFSMIGGILSLLSSVFAPDTVYVVLVSISGFA 375

Query: 301 VVVVWLAIPVAQINFRKEFLKE-NQLEDLSYKTPFTPVLPYITIILLLISIVGIAWDSSQ 359
           VVVVW+ I  +Q  FRK +++  N++ DL Y+TP  P +P     +L L S+VGIA+D +Q
Sbjct: 376 VVVVWMGIAASQFMFRKRYIEAGNKVTDLKYRTPLYPFVPIAAFLLCLASVVGIAFDPNQ 435

Query: 360 RAGLYFGVPFIIFCYIYHKLRYKK 383
           R  LY GVPF+  CY  + ++ +K
Sbjct: 436 RIALYCGVPFMAICYAIYYVKNRK 459
```

There is also homology to SEQ ID 4070.

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

Example 1914

A DNA sequence (GBSx2023) was identified in *S.agalactiae* <SEQ ID 5931> which encodes the amino acid sequence <SEQ ID 5932>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2378(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

There is also homology to SEQ ID 5642.

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

Example 1915

A DNA sequence (GBSx2024) was identified in *S.agalactiae* <SEQ ID 5933> which encodes the amino acid sequence <SEQ ID 5934>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4935(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 1916

A DNA sequence (GBSx2025) was identified in *S.agalactiae* <SEQ ID 5935> which encodes the amino acid sequence <SEQ ID 5936>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0530(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 1917

A DNA sequence (GBSx2026) was identified in *S.agalactiae* <SEQ ID 5937> which encodes the amino acid sequence <SEQ ID 5938>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0175(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:AAF63739 GB:AF236863 hypothetical GTP-binding protein
          [Lactococcus lactis]
 Identities = 142/193 (73%), Positives = 165/193 (84%)

Query: 6   LNTHNASILLSAANKSHYPQDDLPEVALAGRSNVGKSSFINTLLGRKNLARTSSKPGKTQ 65
           +NT+N +I +SAA+K  YP++D PE+ALAGRSNVGKSSFINTLL RKN ARTS +PGKTQ
Sbjct: 3   INTNNLTITISAASKKQYPENDWPEIALAGRSNVGKSSFINTLLNRKNFARTSGQPGKTQ 62

Query: 66  LLNFYNIDDKLRFVDVPGYGYAKVSKTERAKWGKMIEEYLVTRDNLRVVVSLVDFRHDPS 125
           LLNFYNIDD+L FVDVPGYGYA+VSK ER KWGKMIEEYL TR+NL+ VVSLVD RH+PS
Sbjct: 63  LLNFYNIDDQLHFVDVPGYGYARVSKKEREKWGKMIEEYLTTRENLKAVVSLVDIRHEPS 122

Query: 126 ADDIQMYEFLKYYEIPVIIVATKADKIPRGKWNKHESSIKKKLNFDKKDHFIVFSSVDRT 185
            DD+ MYEFLKYY IPVI+VATKADK+PRGKWNKHES IKK + FD  D FI+FSS D+T
Sbjct: 123 EDDLMMYEFLKYYHIPVILVATKADKVPRGKWNKHESIIKKAMKFDSTDDFIIFSSTDKT 182

Query: 186 GLDESWDTILSEL 198
           G++E+W  IL  L
Sbjct: 183 GIEEAWTAILKYL 195
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5939> which encodes the amino acid sequence <SEQ ID 5940>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0123(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 167/196 (85%), Positives = 183/196 (93%)

Query: 3   EEFLNTHNASILLSAANKSHYPQDDLPEVALAGRSNVGKSSFINTLLGRKNLARTSSKPG 62
           E+ LNTHNASILLSAANKSHYPQDDLPE+ALAGRSNVGKSSFINT+LGRKNLARTSSKPG
Sbjct: 4   EQVLNTHNASILLSAANKSHYPQDDLPEIALAGRSNVGKSSFINTILGRKNLARTSSKPG 63

Query: 63  KTQLLNFYNIDDKLRFVDVPGYGYAKVSKTERAKWGKMIEEYLVTRDNLRVVVSLVDFRH 122
           KTQLLNF+NIDDKLRFVDVPGYGYAKVSK+ERAKWGKMIEEYL +RDNLR VVSLVD RH
Sbjct: 64  KTQLLNFFNIDDKLRFVDVPGYGYAKVSKSERAKWGKMIEEYLTSRDNLRAVVSLVDLRH 123

Query: 123 DPSADDIQMYEFLKYYEIPVIIVATKADKIPRGKWNKHESSIKKKLNFDKKDHFIVFSSV 182
            PS +DIQMY+FLKYY+IPVI+VATKADKIPRGKWNKHES +KK LNFDK D FIVFSSV
Sbjct: 124 APSKEDIQMYDFLKYYDIPVIVVATKADKIPRGKWNKHESVVKKALNFDKSDTFIVFSSV 183
```

```
Query: 183 DRTGLDESWDTILSEL 198
           +R G+D+SWD IL ++
Sbjct: 184 ERIGIDDSWDAILEQV 199
```

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

Example 1918

A DNA sequence (GBSx2027) was identified in *S.agalactiae* <SEQ ID 5941> which encodes the amino acid sequence <SEQ ID 5942>. This protein is predicted to be protease ClpX (clpX). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2389(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 9947> which encodes amino acid sequence <SEQ ID 9948> was also identified.

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

```
>GP:AAF63738 GB:AF236863 protease ClpX [Lactococcus lactis]
 Identities = 305/395 (77%), Positives = 357/395 (90%), Gaps = 1/395 (0%)

Query: 18  NVYCSFCGKSQDEVKKIIAGNGVFICNECVALSQEIIKEELAEEVLADLAEVPKPKELLE 77
           N+ CSFCGKSQD+VKK+IAG+ V+ICNEC+ LS  I++EEL EE  +++ EV  PKE+ +
Sbjct: 8   NIQCSFCGKSQDDVKKMIAGSDVYICNECIELSTRILEEELKEEQDSEMLEVKTPKEMFD 67

Query: 78  ILNQYVVGQDRAKRALAVAVYNHYKRVSYTESS-DDDVDLQKSNILMIGPTGSGKTFLAQ 136
            LN+YV+GQ++AKRALAVAVYNHYKR+++T S    +D++LQKSNIL+IGPTGSGKTFLAQ
Sbjct: 68  HLNEYVIGQEKAKRALAVAVYNHYKRINFTASKIAEDIELQKSNILLIGPTGSGKTFLAQ 127

Query: 137 TLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGIIYVDEIDKIA 196
           TLAKSLNVPFAIADATSLTEAGYVGEDVENILLKL+QA+D+N+ERAERGIIY+DEIDKIA
Sbjct: 128 TLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLLQASDFNIERAERGIIYIDEIDKIA 187

Query: 197 KKGENVSITRDVSGEGVQQALLKIIEGTVASVPPQGGRKHPNQEMIQINTKNILFIVGGA 256
           KK ENVSITRDVSGEGVQQALLKIIEGTVASVPPQGGRKHPNQEMIQI+TKNILFIVGGA
Sbjct: 188 KKSENVSITRDVSGEGVQQALLKIIEGTVASVPPQGGRKHPNQEMIQIDTKNILFIVGGA 247

Query: 257 FDGIEDLVKQRLGEKVIGFGQTSRKIDDNASYMQEIISEDIQKFGLIPEFIGRLPVVAAL 316
           FDGIE++VKQRLGEK+IGFG  ++K+ D  SYMQEII+EDIQKFGLIPEFIGRLP+VAAL
Sbjct: 248 FDGIEEIVKQRLGEKIIGFGANNKKLSDEDSYMQEIIAEDIQKFGLIPEFIGRLPIVAAL 307

Query: 317 ELLTAEDLVRILTEPRNALVKQYQTLLSYDGVELEFDQDALLAIADKAIERKTGARGLRS 376
           E LT EDL++ILTEP+NAL+KQY+ LL +D VELEF   AL+AIA KAIERKTGARGLRS
Sbjct: 308 ERLTEEDLIQILTEPKNALIKQYKQLLLFDNVELEFKDGALMAIAKKAIERKTGARGLRS 367

Query: 377 IIEETMLDIMFEIPSQEDVTKVRITKAAVEGTDKP 411
           IIEE M+DIMFE+PS E++TKV IT+A V+G  +P
Sbjct: 368 IIEEVMMDIMFEVPSHEEITKVIITEAVVDGKAEP 402
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5943> which encodes the amino acid sequence <SEQ ID 5944>. Analysis of this protein sequence reveals the following:

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

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

```
bacterial cytoplasm --- Certainty=0.2711(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 = 378/409 (92%), Positives = 393/409 (95%), Gaps = 1/409 (0%)

Query: 9   MAGNRNNDMNVYCSFCGKSQDEVKKIIAGNGVFICNECVALSQEIIKEELAEEVLADLAE 68
            MAG+R ND+ VYCSFCGKSQD+VKKIIAGN VFICNECVALSQEIIKEELAEEVLADL E
Sbjct: 1   MAGSRTNDIKVYCSFCGKSQDDVKKIIAGNNVFICNECVALSQEIIKEELAEEVLADLTE 60

Query: 69  VPKPKELLEILNQYVVGQDRAKRALAVAVYNHYKRVSYTES-SDDDVDLQKSNILMIGPT 127
            VPKPKELL++LNQYVVGQDRAKRAL+VAVYNHYKRVS+TES  DDDVDLQKSNILMIGPT
Sbjct: 61  VPKPKELLDVLNQYVVGQDRAKRALSVAVYNHYKRVSFTESRDDDDVDLQKSNILMIGPT 120

Query: 128 GSGKTFLAQTLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGII 187
            GSGKTFLAQTLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGII
Sbjct: 121 GSGKTFLAQTLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGII 180

Query: 188 YVDEIDKIAKKGENVSITRDVSGEGVQQALLKIIEGTVASVPPQGGRKHPNQEMIQINTK 247
            YVDEIDKIAKKGENVSITRDVSGEGVQQALLKIIEGTVASVPPQGGRKHPNQEMIQI+TK
Sbjct: 181 YVDEIDKIAKKGENVSITRDVSGEGVQQALLKIIEGTVASVPPQGGRKHPNQEMIQIDTK 240

Query: 248 NILFIVGGAFDGIEDLVKQRLGEKVIGFGQTSRKIDDNASYMQEIISEDIQKFGLIPEFI 307
            NILFIVGGAFDGIE++VKQRLGEKVIGFGQ SRKIDDNASYMQEIISEDIQKFGLIPEFI
Sbjct: 241 NILFIVGGAFDGIEEIVKQRLGEKVIGFGQNSRKIDDNASYMQEIISEDIQKFGLIPEFI 300

Query: 308 GRLPVVAALELLTAEDLVRILTEPRNALVKQYQTLLSYDGVELEFDQDALLAIADKAIER 367
            GRLPVVAALE L   DL++ILTEPRNALVKQYQ LLSYDGVEL FD++AL AIA+KAIER
Sbjct: 301 GRLPVVAALEQLNTSDLIQILTEPRNALVKQYQALLSYDGVELAFDKEALEAIANKAIER 360

Query: 368 KTGARGLRSIIEETMLDIMFEIPSQEDVTKVRITKAAVEGTDKPVLETA 416
            KTGARGLRSIIEETMLDIMFEIPSQEDVTKVRITKAAVEG  KPVLETA
Sbjct: 361 KTGARGLRSIIEETMLDIMFEIPSQEDVTKVRITKAAVEGKSKPVLETA 409
```

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

Example 1919

A DNA sequence (GBSx2028) was identified in *S.agalactiae* <SEQ ID 5945> which encodes the amino acid sequence <SEQ ID 5946>. 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 1920

A DNA sequence (GBSx2029) was identified in *S.agalactiae* <SEQ ID 5947> which encodes the amino acid sequence <SEQ ID 5948>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4029(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 9949> which encodes amino acid sequence <SEQ ID 9950> was also identified.

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

```
>GP:AAC33872 GB:AF055727 dihydrofolate reductase [Streptococcus pneumoniae]
 Identities = 83/162 (51%), Positives = 118/162 (72%), Gaps = 1/162 (0%)

Query: 25  MTKQIIAIWAEDEDHLIGVNGGLPWRLPKELHHFKETTMGQALLMGRKTFDGMNRRVLPG 84
           MTK+I+AIWA+DE+ LIG     LPW LP EL HFKETT+  A+LMGR TFDGM RR+LP
Sbjct: 1   MTKKIVAIWAQDEEGLIGKENRLPWHLPAELQHFKETTLNHAILMGRVTFDGMGRRLLPK 60

Query: 85  RETIILTKDEQFQADGVTVLNSVEQVIKWFQEHNKTLFIVGGASIYKAFLPYCEAIIKTK 144
           RET+ILT++ + + DGV     V+ V+ W+Q+  K L+I+GG  I++AF PY + +I T
Sbjct: 61  RETLILTRNPEEKIDGVATFQDVQSVLDWYQDQEKNLYIIGGKQIFQAFEPYLDEVIVTH 120

Query: 145 VHGKFKGDTYFP-DVNLSEFKVISRDYFEKDEQNAHAFTVTY 185
           +H + +GDTYFP +++LS F+ +S  ++ KDE+N + FT+ Y
Sbjct: 121 IHARVEGDTYFPEELDLSLFETVSSKFYAKDEKNPYDFTIQY 162
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5949> which encodes the amino acid sequence <SEQ ID 5950>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1214(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 = 82/160 (51%), Positives = 119/160 (74%)

Query: 25  MTKQIIAIWAEDEDHLIGVNGGLPWRLPKELHHFKETTMGQALLMGRKTFDGMNRRVLPG 84
           MTK+IIAIWAEDE  LIG+ G LPW LPKEL HFK+TT+ QA+LMGR TF+GMN + LP
Sbjct: 1   MTKEIIAIWAEDEAGLIGIAGKLPWYLPKELEHFKKTTLHQAILMGRVTFEGMNCKRLPQ 60

Query: 85  RETIILTKDEQFQADGVTVLNSVEQVIKWFQEHNKTLFIVGGASIYKAFLPYCEAIIKTK 144
           R+T+++T++  +Q D V  + S+E+V++W+    +KTL+I+GG  + +AF  Y + IIKT
Sbjct: 61  RQTLVMTRNRDYQVDEVLTMTSIEKVLEWYHAQDKTLYIIGGNKVLEAFNGYFDRIIKTV 120

Query: 145 VHGKFKGDTYFPDVNLSEFKVISRDYFEKDEQNAHAFTVT 184
           +H +FKGDTY P+++ S F   S+ ++ +D +N + FTVT
Sbjct: 121 IHHRFKGDTYRPNLDFSHFTQESQTFYARDAKNPYDFTVT 160
```

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

Example 1921

A DNA sequence (GBSx2030) was identified in *S.agalactiae* <SEQ ID 5951> which encodes the amino acid sequence <SEQ ID 5952>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
```

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1577(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:AAA25221 GB:M33770 thymidylate synthase (EC 2.1.1.45)
           [Lactococcus lactis]
 Identities = 215/280 (76%), Positives = 245/280 (86%), Gaps = 2/280 (0%)

Query: 1   MTKADLLFKDNITKIMSEGVFSEQARPRYKNGEMANSKYITGAFAEYDLSKGEFPITTLR 60
           MT AD +FK NI  I+  GVFSE ARP+YK+G+MANSKY+TG+F  YDL KGEFPITTLR
Sbjct: 1   MTYADQVFKQNIQNILDNGVFSENARPKYKDGQMANSKYVTGSFVTYDLQKGEFPITTLR 60

Query: 61  PIPIKSAIKEIFWIYQDQTNDLAVLNDKYGVTYWNDWEVGHTGTIGQRYGAVVKKHNIIS 120
           PIPIKSAIKE+ WIYQDQT++L+VL +KYGV YW +W +G  GTIGQRYGA VKK+NII
Sbjct: 61  PIPIKSAIKELMWIYQDQTSELSVLEEKYGVKYWGEWGIGD-GTIGQRYGATVKKYNIIG 119

Query: 121 KLLKQLEDNPWNRRNVISLWDYEAFEETEGLLPCAFQTMFDVRRV-NGELYLDATLTQRS 179
           KLL+ L  NPWNRRN+I+LW YE FEETEGLLPCAFQTMFDVRR  +G++YLDATL QRS
Sbjct: 120 KLLEGLAKNPWNRRNIINLWQYEDFEETEGLLPCAFQTMFDVRREKDGQIYLDATLIQRS 179

Query: 180 NDMLVAHHINAMQYVALQMMIAKHFGWRVGKFFYFINNLHIYDNQFEQAQELLKRQPSEC 239
           NDMLVAHHINAMQYVALQMMIAKHF W+VGKFFYF+NNLHIYDNQFEQA EL+KR  SE
Sbjct: 180 NDMLVAHHINAMQYVALQMMIAKHFSWKVGKFFYFVNNLHIYDNQFEQANELMKRTASEK 239

Query: 240 NPKLVLNVPDGTDFFDIKPDDFALVDYDPIKPQLRFDLAI 279
            P+LVLNVPDGT+FFDIKP+DF LVDY+P+KPQL+FDLAI
Sbjct: 240 EPRLVLNVPDGTNFFDIKPEDFELVDYEPVKPQLKFDLAI 279
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5953> which encodes the amino acid sequence <SEQ ID 5954>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3131(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 = 227/279 (81%), Positives = 251/279 (89%)

Query: 1   MTKADLLFKDNITKIMSEGVFSEQARPRYKNGEMANSKYITGAFAEYDLSKGEFPITTLR 60
           MTKAD +FK NI KI++EG  SEQARP+YK+G  A+SKYITGAFAEYDL+KGEFPITTLR
Sbjct: 9   MTKADQIFKANIQKIINEGSLSEQARPKYKDGRTAHSKYITGAFAEYDLAKGEFPITTLR 68

Query: 61  PIPIKSAIKEIFWIYQDQTNDLAVLNDKYGVTYWNDWEVGHTGTIGQRYGAVVKKHNIIS 120
           PIPIKSAIKE+FWIYQDQ+N L VL  KY V YWN+WEV  T TIGQRYGAVVKKH+IIS
Sbjct: 69  PIPIKSAIKELFWIYQDQSNSLDVLEAKYNVHYWNEWEVDQTRTIGQRYGAVVKKHDIIS 128

Query: 121 KLLKQLEDNPWNRRNVISLWDYEAFEETEGLLPCAFQTMFDVRRVNGELYLDATLTQRSN 180
           K+LKQL +NPWNRRNVISLWDYEAFEET+GLLPCAFQ MFDVRRV  +LYLDA+LTQRSN
Sbjct: 129 KILKQLAENPWNRRNVISLWDYEAFEETKGLLPCAFQIMFDVRRVGEDLYLDASLTQRSN 188

Query: 181 DMLVAHHINAMQYVALQMMIAKHFGWRVGKFFYFINNLHIYDNQFEQAQELLKRQPSECN 240
           D+LVAHHINAMQYVALQMMIAKHFGW++GKFFYF+NNLHIYDNQF+QAQELLKRQP
Sbjct: 189 DILVAHHINAMQYVALQMMIAKHFGWKIGKFFYFVNNLHIYDNQFDQAQELLKRQPVASQ 248

Query: 241 PKLVLNVPDGTDFFDIKPDDFALVDYDPIKPQLRFDLAI 279
           PKLVLNVPD T+FFDIKPDDF L +YDP+KPQL FDLAI
Sbjct: 249 PKLVLNVPDRTNFFDIKPDDFELQNYDPVKPQLHFDLAI 287
```

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

Example 1922

A DNA sequence (GBSx2031) was identified in *S.agalactiae* <SEQ ID 5955> which encodes the amino acid sequence <SEQ ID 5956>. This protein is predicted to be HMG-CoA synthase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0816(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 5957> which encodes the amino acid sequence <SEQ ID 5958>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1670(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 = 260/385 (67%), Positives = 325/385 (83%)

Query: 36  MKIGIDKIGFATSQYVLEMTDLAIARQVDPEKFSKGLLLDSLSITPVTEDIVTLAASAAN 95
           M IGIDKIGFATSQYVL++ DLA+ARQVDP KFS+GLL++S S+ P+TEDI+TLAASAA+
Sbjct: 14  MTIGIDKIGFATSQYVLKLEDLALARQVDPAKFSQGLLIESFSVAPITEDIITLAASAAD 73

Query: 96  DILSDEDKETIDMVIVATESSIDQSKAASVYVHQLLEIQPFARSFEMKEACYSATAALDY 155
            IL+DED+  IDMVI+ATESS DQSKA+++YVH L+ IQPFARSFE+K+ACYSATAALDY
Sbjct: 74  QILTDEDRAKIDMVILATESSTDQSKASAIYVHHLVGIQPFARSFEVKQACYSATAALDY 133

Query: 156 AKLHVEKHPDSKVLVIASDIAKYGIKSTGESTQGAGSIAMLISQNPSILELKEDHLAQTR 215
           AKLHV   PDS+VLVIASDIA+YG+ S GESTQG+GSIA+L++ NP IL L ED++AQTR
Sbjct: 134 AKLHVASKPDSRVLVIASDIARYGVGSPGESTQGSGSIALLVTANPRILALNEDNVAQTR 193

Query: 216 DIMDFWRPNYSDVPYVNGMFSTKQYLDMLKTTWKVYQKRFNTSLSDYAAFCFHIPFPKLA 275
           DIMDFWRPNYS  PYV+G++STKQYL+ L+TTW+ YQKR N  LSD AA CFHIPFPKLA
Sbjct: 194 DIMDFWRPNYSFTPYVDGIYSTKQYLNCLETTWQAYQKRENLQLSDLAAVCFHIPFPKLA 253

Query: 276 LKGFNKILDNNLDEQKKAELQENFEHSITYSKKIGNCYTGSLYLGLLSLLENSQNLKAGD 335
           LKG N I+DN +  + + +L E F+ SI+YSK+IGN YTGSLYLGLLSLLENS+ L++GD
Sbjct: 254 LKGLNNIMDNTVPPEHREKLIEAFQASISYSKQIGNIYTGSLYLGLLSLLENSKVLQSGD 313

Query: 336 QIAFFSYGSGAVAEIFTGQLVDGYQNKLQSDRMDQLNKRQKITVTEYEKLFFEKTILDEN 395
           +I FFSYGSGAV+E ++GQLV GY   L ++R   L++R +++V++YE LF+E+  LD+N
Sbjct: 314 KIGFFSYGSGAVSEFYSGQLVAGYDKMLMTNRQALLDQRTRLSVSKYEDLFYEQVQLDDN 373

Query: 396 GNANFNTYRTGTFSLDSICEHQRIY 420
           GNANF+ Y TG F+L +I EH+RIY
Sbjct: 374 GNANFDIYLTGKFALTAIKEHRRIY 398
```

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

Example 1923

A DNA sequence (GBSx2032) was identified in *S.agalactiae* <SEQ ID 5959> which encodes the amino acid sequence <SEQ ID 5960>. This protein is predicted to be HMG-CoA reductase (mvaA). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.49   Transmembrane  348 - 364 ( 348 - 364)
   INTEGRAL    Likelihood = -1.33   Transmembrane   53 -  69 (  53 -  69)

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

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

```
>GP:AAG02454 GB:AF290098 HMG-CoA reductase [Streptococcus pneumoniae]
 Identities = 266/421 (63%), Positives = 343/421 (81%), Gaps = 3/421 (0%)

Query: 3   KISWTGFSKKSPEERIHYLEEQDFLADSSLEIVTNQDLLSLSLANQMAENVIGRIALPFS 62
           KISW GFSKKS +ER+  L+ Q  L+      +   + +S+++A+Q++ENV+G  +LP+S
Sbjct: 2   KISWNGFSKKSYQERLELLKAQALLSPERQASLEKDEQMSVTVADQLSENVVGTFSLPYS 61

Query: 63  LVPDVLVNGKVYQVPYVTEEPSVVAAASFAAKIIKRSGGFLTTVHNRKMIGQVALYDVQD 122
           LVP+VLVNG+ Y VPYVTEEPSVVAAAS+A+KIIKR+GGF   VH R+MIGQVALY V +
Sbjct: 62  LVPEVLVNGQGYTVPYVTEEPSVVAAASYASKIIKRAGGFTAQVHQRQMIGQVALYQVAN 121

Query: 123 SQHTKESILNQKQQLLEIANAAHPSIVKRGGGACDLTIEI---KEDFLIVYLMVDTKEAM 179
            +  +E I ++K +LLE+AN A+PSIVKRGGGA DL +E    + DFL+VY+ VDT+EAM
Sbjct: 122 PKLAQEKIASKKAELLELANQAYPSIVKRGGGARDLHVEQIKGEPDFLVVYIHVDTQEAM 181

Query: 180 GANMVNTMMEALSSPLEDISKGKSLMSILSNYATESLVTATCRVDLRFLSRQKEEAIKLA 239
           GANM+NTM+EAL   LE++S+G+SLM ILSNYAT+SLVTA+CR+  R+LSRQK++  ++A
Sbjct: 182 GANMLNTMLEALKPVLEELSQGQSLMGILSNYATDSLVTASCRIAFRYLSRQKDQGREIA 241

Query: 240 QKMTMASQLAQVDPYRASTHNKGIFNGIDAIVLATGNDWRAIEAGAHTYAVKDGQYRGLS 299
           +K+ +ASQ AQ DPYRA+THNKGIFNGIDAI++ATGNDWRAIEAGAH +A +DG+Y+GLS
Sbjct: 242 EKIALASQFAQADPYRAATHNKGIFNGIDAILIATGNDWRAIEAGAHAFASRDGRYQGLS 301

Query: 300 RWSYKVDDNCLEGTLTLPMPVATKGGSIGINPSVHLAHDLLGRPNAKELASIILSIGLAQ 359
            W+  ++   L G +TLPMPVATKGGSIG+NP V L+HDLLG P+A+ELA II+SIGLAQ
Sbjct: 302 CWTLDLEREELVGEMTLPMPVATKGGSIGLNPRVALSHDLLGNPSARELAQIIVSIGLAQ 361

Query: 360 NFAALKALVSTGIQAGHMKLQAKSLALLAGAKEEQISEVVKQLLDSKHMNLETAQKIVNKL 420
           NFAALKALVSTGIQ GHMKLQAKSLALLAGA E +++ +V++L+  K  NLETAQ+ +  L
Sbjct: 362 NFAALKALVSTGIQQGHMKLQAKSLALLAGASESEVAPLVERLISDKTFNLETAQRYLENL 422
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5961> which encodes the amino acid sequence <SEQ ID 5962>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3929(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 = 257/422 (60%), Positives = 330/422 (77%)

Query: 2   TKISWTGFSKKSPEERIHYLEEQDFLADSSLEIVTNQDLLSLSLANQMAENVIGRIALPF 61
           T ++W+GFSKK+ EER+  +E+   L   +L +    LL +  ANQM ENV+GR+ALPF
```

```
Sbjct: 4   TNLNWSGFSKKTFEERLQLIEKFKLLNAENLNQLKTDVLLPIQTANQMTENVLGRLALPF 63

Query: 62  SLVPDVLVNGKVYQVPYVTEEPSVVAAASFAAKIIKRSGGFLTTVHNRKMIGQVALYDVQ 121
           S+ PD LVNG  YQ+P+VTEEPSVVAAASFAAK+IKRSGGF     NR+MIGQ+ LYD+
Sbjct: 64  SIAPDFLVNGSTYQMPFVTEEPSVVAAASFAAKLIKRSGGFKAQTLNRQMIGQIVLYDID 123

Query: 122 DSQHTKESILNQKQQLLEIANAAHPSIVKRGGGACDLTIEIKEDFLIVYLMVDTKEAMGA 181
              + K +IL++ ++L+ +AN A+PSIVKRGGGA  + +E K +FLI YL VDT+EAMGA
Sbjct: 124 QIDNAKAAILHKTKKLIALANKAYPSIVKRGGGARTIHLEEKGEFLIFYLTVDTQEAMGA 183

Query: 182 NMVNTMMEALSSPLEDISKGKSLMSILSNYATESLVTATCRVDLRFLSRQKEEAIKLAQK 241
           NMVNTMMEAL   L  +SKG  LM+ILSNYATESLVT +C + +R L   K ++++LAQK
Sbjct: 184 NMVNTMMEALVPDLTRLSKGHCLMAILSNYATESLVTTSCEIPVRLLDHDKTKSLQLAQK 243

Query: 242 MTMASQLAQVDPYRASTHNKGIFNGIDAIVLATGNDWRAIEAGAHTYAVKDGQYRGLSRW 301
           + +AS+LAQVDPYRA+THNKGIFNGIDA+V+ATGNDWRAIEAGAH YA ++G Y+GLS+W
Sbjct: 244 IELASRLAQVDPYRATTHNKGIFNGIDAVVIATGNDWRAIEAGAHAYASRNGSYQGLSQW 303

Query: 302 SYKVDDNCLEGTLTLPMPVATKGGSIGINPSVHLAHDLLGRPNAKELASIILSIGLAQNF 361
             +  D   L G +TLPMP+A+KGGSIG+NP+V +AHDLL +P+AK LA +I S+GLAQNF
Sbjct: 304 HFDQDKQVLLGQMTLPMPIASKGGSIGLNPTVSIAHDLLNQPDAKTLAQLIASVGLAQNF 363

Query: 362 AALKALVSTGIQAGHMKLQAKSLALLAGAKEEQISEVVKQLLDSKHMNLETAQKIVNKLT 421
           AALKAL S+GIQAGHMKL AKSLALLAGA +++I+ +V  LL  K +NLE A   +++L
Sbjct: 364 AALKALTSSGIQAGHMKLHAKSLALLAGATQDEIAPLVNALLADKPINLEKAHFYLSQLR 423

Query: 422 KS 423
           +S
Sbjct: 424 QS 425
```

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

Example 1924

A DNA sequence (GBSx2033) was identified in *S.agalactiae* <SEQ ID 5963> which encodes the amino acid sequence <SEQ ID 5964>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2355(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 5965> which encodes the amino acid sequence <SEQ ID 5966>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2687(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 = 76/138 (55%), Positives = 100/138 (72%), Gaps = 2/138 (1%)

Query: 7   PKWEELPELDLYLDQVLLYVNQLINPKTITNDKLLTASMINNYVKHNYISKPIKKKYNRR 66
           P W++LP+LDLYLDQVLLYVNQ  +   ++++K LTASMINNYVKH Y++KPIKKKY ++
```

```
Sbjct: 7   PYWKDLPDLDLYLDQVLLYVNQCTDFSEVSDNKSLTASMINNYVKHGYVTKPIKKKYQKQ 66

Query: 67  QVARLIVITAPKQVFAIQEISQTLELLTADNHSEEAYNGFAACMNKEE--VHDLPPVVIS 124
           Q+ARLI I+ PK VF IQ+IS+ LE L A   SE  YN F  C N++      D+PP+V
Sbjct: 67  QLARLIAISLPKTVFPIQDISRVLEELQAQADSESLYNTFVTCWNQKAPIEEDIPPIVQV 126

Query: 125 ACQTLNLYQETQKLVLEL 142
           ACQT+  Y +T  L+ E+
Sbjct: 127 ACQTVKDYHKTIYLLQEV 144
```

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

Example 1925

A DNA sequence (GBSx2034) was identified in *S.agalactiae* <SEQ ID 5967> which encodes the amino acid sequence <SEQ ID 5968>. This protein is predicted to be hemolysin iii. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.08   Transmembrane  142 - 158 ( 140 - 165)
   INTEGRAL    Likelihood = -6.79   Transmembrane   26 -  42 (  19 -  44)
   INTEGRAL    Likelihood = -5.63   Transmembrane  200 - 216 ( 196 - 217)
   INTEGRAL    Likelihood = -5.41   Transmembrane  104 - 120 ( 102 - 121)
   INTEGRAL    Likelihood = -3.98   Transmembrane   51 -  67 (  49 -  69)
   INTEGRAL    Likelihood = -1.86   Transmembrane  172 - 188 ( 169 - 188)

----- Final Results -----
              bacterial membrane --- Certainty=0.4630(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 9951> which encodes amino acid sequence <SEQ ID 9952> was also identified.

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

```
>GP:CAA58877 GB:X84058 novel hemolytic factor [Bacillus cereus]
 Identities = 79/204 (38%), Positives = 132/204 (63%), Gaps = 4/204 (1%)

Query: 17  EELANSITHAVGALLMLILLPITAVYSHNHFGLQAALGTSIFVTSLFLMFLSSSIYHSMT 76
           EE+AN+ITH +GA+L +  L I  +++  H    A +  +++  S+FL++L S++ HS+
Sbjct: 14  EEIANAITHGIGAILSIPALIILIIHASKHGTASAVVAFTVYGVSMFLLYLFSTLLHSIH 73

Query: 77  YNSLQKYVLRMIDHSMIYIAIAGSYTPVALSLIGGWLGYLIIFLQWGITLFGILYKIFAP 136
           +  ++K +  ++DHS IY+ IAG+YTP  L  + G LG+ ++ + W + + GI++KIF
Sbjct: 74  HPKVEK-LFTILDHSAIYLLIAGTYTPFLLITLRGPLGWTLLAIIWTLAIGGIIFKIFFV 132

Query: 137 KINDKFSLVLYLIMGWLVIF-IFPAIITKTGPAFWGLLLAGGICYTIGALFYA-RKRPYD 194
           +   K S + Y+IMGWL+I  I P     TG  F  LLLAGGI Y++GA+F+   K P++
Sbjct: 133 RRFIKASTLCYIIMGWLIIVAIKPLYENLTGHGF-SLLLAGGILYSVGAIFFLWEKLPFN 191

Query: 195 HMIWHLFILLASILQYIGIVYFML 218
           H IWHLF+L  S + +  +++++L
Sbjct: 192 HAIWHLFVLGGSAMMFFCVLFYVL 215
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5969> which encodes the amino acid sequence <SEQ ID 5970>. Analysis of this protein sequence reveals the following:

```
     Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.51   Transmembrane  144 - 160 ( 138 - 163)
   INTEGRAL    Likelihood = -9.87   Transmembrane   49 -  65 (  45 -  71)
```

```
    INTEGRAL    Likelihood = -7.11   Transmembrane  198 - 214 ( 193 - 215)
    INTEGRAL    Likelihood = -6.16   Transmembrane  102 - 118 ( 100 - 120)
    INTEGRAL    Likelihood = -2.97   Transmembrane   20 -  36 (  20 -  41)
    INTEGRAL    Likelihood = -1.01   Transmembrane  167 - 183 ( 167 - 185)

----- 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:CAA58877 GB:X84058 novel hemolytic factor [Bacillus cereus]
 Identities = 82/204 (40%), Positives = 128/204 (62%), Gaps = 4/204 (1%)

Query: 15  EEVANSVTHAIGAFAMLILLPISASYAYQTYDLKAAIGISIFVISLFLMFLSSTIYHSMA 74
           EE+AN++TH IGA   +  L I   +A +     A +  +++ +S+FL++L ST+ HS+
Sbjct: 14  EEIANAITHGIGAILSIPALIILIIHASKHGTASAVVAFTVYGVSMFLLYLFSTLLHSIH 73

Query: 75  YGSVHKYILRIIDHSMIYIAIAGSYTPVALSLVSGWLGYIIIVLQWGITLFGILYKIFAK 134
           +  V K +  I+DHS IY+ IAG+YTP  L  + G LG+ ++ + W + + GI++KIF
Sbjct: 74  HPKVEK-LFTILDHSAIYLLIAGTYTPFLLITLRGPLGWTLLAIIWTLAIGGIIFKIFFV 132

Query: 135 RINEKFSLMLYIVMGWL-VVFILPVIIQKTSLAFGLLMLFGGLSYTIGAVFYA-KKRPYF 192
           R   K S + YI+MGWL +V I P+    T   F LL L GG+ Y++GA+F+  +K P+
Sbjct: 133 RRFIKASTLCYIIMGWLIIVAIKPLYENLTGHGFSLL-LAGGILYSVGAIFFLWEKLPFN 191

Query: 193 HMIWHLFILLASALQFIAITFFML 216
           H IWHLF+L  SA+ F  + F++L
Sbjct: 192 HAIWHLFVLGGSAMMFFCVLFYVL 215
```

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

```
 Identities = 153/213 (71%), Positives = 181/213 (84%)

Query: 6   SIKLSPQLSFGEELANSITHAVGALLMLILLPITAVYSHNHFGLQAALGTSIFVTSLFLM 65
           + K S  LSF EE+ANS+THA+GA  MLILLPI+A Y++  + L+AA+G SIFV SLFLM
Sbjct: 4   TFKQSLPLSFSEEVANSVTHAIGAFAMLILLPISASYAYQTYDLKAAIGISIFVISLFLM 63

Query: 66  FLSSSIYHSMTYNSLQKYVLRMIDHSMIYIAIAGSYTPVALSLIGGWLGYLIIFLQWGIT 125
           FLSS+IYHSM Y S+ KY+LR+IDHSMIYIAIAGSYTPVALSL+ GWLGY+II LQWGIT
Sbjct: 64  FLSSTIYHSMAYGSVHKYILRIIDHSMIYIAIAGSYTPVALSLVSGWLGYIIIVLQWGIT 123

Query: 126 LFGILYKIFAPKINDKFSLVLYLIMGWLVIFIFPAIITKTGPAFWGLLLAGGICYTIGAL 185
           LFGILYKIFA +IN+KFSL+LY++MGWLV+FI P II KT  AF  L+L GG+ YTIGA+
Sbjct: 124 LFGILYKIFAKRINEKFSLMLYIVMGWLVVFILPVIIQKTSLAFGLLMLFGGLSYTIGAV 183

Query: 186 FYARKRPYDHMIWHLFILLASILQYIGIVYFML 218
           FYA+KRPY HMIWHLFILLAS LQ+I I +FML
Sbjct: 184 FYAKKRPYFHMIWHLFILLASALQFIAITFFML 216
```

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

Example 1926

A DNA sequence (GBSx2035) was identified in *S.agalactiae* <SEQ ID 5971> which encodes the amino acid sequence <SEQ ID 5972>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3641(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:CAB12492 GB:Z99107 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 81/302 (26%), Positives = 157/302 (51%), Gaps = 10/302 (3%)

Query: 1   MKSAYIFFNPKSGKDEQALAKEVKSYLIEHDFQDDY-VRIITPSSVEEAVALAKKASEDH 59
           MK A I +NP SG++   + K+  + +++   Q  Y       +   +A   AK+A+
Sbjct: 1   MKRARIIYNPTSGRE---IFKKHLAQVLQKFEQAGYETSTHATTCAGDATHAAKEAALRE 57

Query: 60  IDLVIPLGGDGTINKICGGVYAGGAYPTIGLVPAGTVNNFSKALNIPQERNL-ALENLLN 118
            DL+I  GGDGTIN++  G+       PT+G++P GT N+F++AL IP+E  L A + ++N
Sbjct: 58  FDLIIAAGGDGTINEVVNGLAPLDNRPTLGVIPVGTTNDFARALGIPREDILKAADTVIN 117

Query: 119 GHVKSVDICKVNDDYMISSLTLGLLADIAANVTSEMKRKLGPFAFLGDAYRILKRNRSYS 178
           G  + +DI +VN  Y I+    G L ++  +V S++K  LG  A+       +L   R
Sbjct: 118 GVARPIDIGQVNGQYFINIAGGGRLTELTYDVPSKLKTMLGQLAYYLKGMEMLPSLRPTE 177

Query: 179 ITLAYDNNVRSLRTRLLLITMTNSIAGMPAFSPEATIDDGLFRVYTMEHIHFFKLLLHLR 238
           + + YD  +      L L+T+TNS+ G    +P+++++DG+F +  ++  +  + + +
Sbjct: 178 VEIEYDGKLFQGEIMLFLVTLTNSVGGFEKLAPDSSLNDGMFDLMILKKANLAEFIRVAT 237

Query: 239 QFRKGDFSQAKEIKHFHTNNLTISTFKRKKSAIPKVRIDGDPGDQLPVKVEVIPKALKFI 298
              +G+    + I +   N + ++  ++       ++ +DG+ G  LP +   + + +
Sbjct: 238 MALRGEHINDQHIIYTKANRVKVNVSEKM-----QLNLDGEYGGMLPGEFVNLYRHIHVV 292

Query: 299 IP 300
           +P
Sbjct: 293 MP 294
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5119> which encodes the amino acid sequence <SEQ ID 5120>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4258(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 = 172/300 (57%), Positives = 229/300 (76%)

Query: 1   MKSAYIFFNPKSGKDEQALAKEVKSYLIEHDFQDDYVRIITPSSVEEAVALAKKASEDHI 60
           MK+  IF+NP SGK E  LA++VK Y  +H F +D V++ITP   ++A  LAK+A++D I
Sbjct: 1   MKTVRIFYNPNSGKKESQLARQVKDYFCQHGFSEDSVKVITPKDADQAFQLAKQAAKDKI 60

Query: 61  DLVIPLGGDGTINKICGGVYAGGAYPTIGLVPAGTVNNFSKALNIPQERNLALENLLNGH 120
           DLVIPLGGDGT+NKI GG+Y GGA+  IGLVP+GTVNNF+KA++IP +   AL+ +L G
Sbjct: 61  DLVIPLGGDGTLNKIIGGIYEGGAHCLIGLVPSGTVNNFAKAMHIPLQITEALDTILTGQ 120

Query: 121 VKSVDICKVNDDYMISSLTLGLLADIAANVTSEMKRKLGPFAFLGDAYRILKRNRSYSIT 180
           +K VDICK N  YMISSLTLGLLADIAA+VT+E KR+ GP AFL D+ RILKRNRSY+I+
Sbjct: 121 IKQVDICKANQQYMISSLTLGLLADIAADVTAEEKRRFGPLAFLKDSIRILKRNRSYAIS 180

Query: 181 LAYDNNVRSLRTRLLLITMTNSIAGMPAFSPEATIDDGLFRVYTMEHIHFFKLLLHLRQF 240
           L   N+   L+T+ LLITMTN+IAG P+FSP A  DDG F+VYTM+ + FFK L H+  F
Sbjct: 181 LISHNNHRIHLKTKFLLITMTNTIAGFPSFSPGAQADDGYFQVYTMKKVSFFKFLWHINDF 240

Query: 241 RKGDFSQAKEIKHFHTNNLTISTFKRKKSAIPKVRIDGDPGDQLPVKVEVIPKALKFIIP 300
           ++GDFS+A+EI HF  N L++    +K++ +P+ RIDGD  D LP+++++IPKA+  I+P
Sbjct: 241 KQGDFSKAEEISHFQANTLSLLPQAKKQAILPRTRIDGDKSDYLPIQLDIIPKAVSIIVP 300
```

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

Example 1927

A DNA sequence (GBSx2036) was identified in *S.agalactiae* <SEQ ID 5973> which encodes the amino acid sequence <SEQ ID 5974>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3628(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:BAB10885 GB:AB010693 gene_id:K21C13.21~pir||T04769~strong
          similarity to unknown protein [Arabidopsis thaliana]
 Identities = 85/291 (29%), Positives = 150/291 (51%), Gaps = 28/291 (9%)

Query: 10  DQEWEVPVESGRYHMIVGEFCPYAQRPQIARQLLGLDKHISISFVDDV------------ 57
           D + + P ESGRYH+ +    CP+A R      ++ GLD+ I+ S V  +
Sbjct: 29  DPDSQFPAESGRYHLYISYACPWACRCLSYLKIKGLDEAITFSSVHAIWGRTKETDDHRG 88

Query: 58  ----PSDIGLIFSQPEQVTGAKSLRDIYHLTDPTYQGPYTIPILIDKTDNRIVCKESADL 113
                SD  L  ++P+ + GAKS+R++Y +  P Y+G YT+P+L DK     +V  ES+++
Sbjct: 89  WVFPDSDTELPGAEPDYLNGAKSVRELYEIASPNYEGKYTVPVLWDKKLKTVVNNESSEI 148

Query: 114 LRLFTTDFSDLHQEDAPVLFSQETASLIDNDIKDINKNFQSLMYKLAFLDKQADYDTYSK 173
           +R+F T+F+ + +  +  L+      +I+     +     + +YK  F  KQ  Y+
Sbjct: 149 IRMFNTEFNGIAKTPSLDLYPSHLRDVINETNGWVFNGINNGVYKCGFARKQEPYNEAVN 208

Query: 174 EFFTFLDQKEHLLGQRPFLLGDNLSEVDIHFFTPLVRWDIAGRDLLLLNQKALEDYPNIF 233
           + +  +D+ E +LG++ ++ G+  +E DI  F  L+R+D       N++ L +YPNIF
Sbjct: 209 QLYEAVDRCEEVLGKQRYICGNTFTEADIRLFVTLIRFDEVYAVHFKCNKRLLREYPNIF 268

Query: 234 SWAKTLYNDFNLKTLTNPQSIKNNYY-----LGKFGRAVRHHTIVPTGPNM 279
           ++ K +Y    + +  N + IK +YY     +  FG       I+P GPN+
Sbjct: 269 NYIKDIYQIHGMSSTVNMEHIKQHYYGSHPTINPFG-------IIPHGPNI 312
```

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 1928

A DNA sequence (GBSx2037) was identified in *S.agalactiae* <SEQ ID 5975> which encodes the amino acid sequence <SEQ ID 5976>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2647(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:BAB07793 GB:AB037666 hypothetical protein [Streptomyces sp.
          CL190]
```

```
 Identities = 127/331 (38%), Positives = 194/331 (58%), Gaps = 9/331 (2%)

Query: 4   RKDDHIKYALKYQSHY---NSFDDIELIHSSLPKYNVNDIDLSTHFAGQSFEFPFYINAM 60
           RKDDH++ A++  + +   N FDD+  +H +L   +  D+ L+T FAG S++ P YINAM
Sbjct: 6   RKDDHVRLAIEQHNAHSGRNQFDDVSFVHHALAGIDRPDVSLATSFAGISWQVPIYINAM 65

Query: 61  TGGSEKGKAVNHKLAQVAQATGIVMVTGSYSAALKNDE--DDSYPTTDLYPDLKLATNIG 118
           TGGSEK   +N  LA  A+ TG+ + +GS +A +K+     D      D  P+  +  NI
Sbjct: 66  TGGSEKTGLINRDLATAARETGVPIASGSMNAYIKDPSCADTFRVLRDENPNGFVIANIN 125

Query: 119 LDKPVPAAESTVKAMNPIFLQVHVNVMQELLMPEGEREFHMWRSHLKEYVDNIQCPLILK 178
               V  A+  +  +    LQ+H+N  QE  MPEG+R F  W   +++     +  P+I+K
Sbjct: 126 ATTTVDNAQRAIDLIEANALQIHINTAQETPMPEGDRSFASWVPQIEKIAAAVDIPVIVK 185

Query: 179 EVGFGMDLQSIKDAYDIGITTVDISGRGGTSFAYIENQRGR--DRSYLNTWGQTTAQSLI 236
           EVG G+  Q+I     D+G+   D+SGRGGT FA IEN R    D ++L+ WGQ+TA  L+
Sbjct: 186 EVGNGLSRQTILLLADLGVQAADVSGRGGTDFARIENGRRELGDYAFLHGWGQSTAACLL 245

Query: 237 NAQSMMDKMDILASGGIRHPLDMVKCLVLGAKAVGLSRTVLELVERYPVDDVIAILNSWK 296
           +AQ +   + +LASGG+RHPLD+V+ L LGA+AVG S   L  +    VD +I  L +W
Sbjct: 246 DAQDI--SLPVLASGGVRHPLDVVRALALGARAVGSSAGFLRTLMDDGVDALITKLTTWL 303

Query: 297 EDLRMIMCALNCKKITDLRQVNYILYGQLKE 327
           + L  +   L  +   DL + + +L+G+L++
Sbjct: 304 DQLAALQTMLGARTPADLTRCDVLLHGELRD 334
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5977> which encodes the amino acid sequence <SEQ ID 5978>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2823(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 = 244/329 (74%), Positives = 284/329 (86%)

Query: 1   MTNRKDDHIKYALKYQSHYNSFDDIELIHSSLPKYNVNDIDLSTHFAGQSFEFPFYINAM 60
           MTNRKDDHIKYALKYQS YN+FDDIELIH SLP Y+++DIDLSTHFAGQ F+FPFYINAM
Sbjct: 31  MTNRKDDHIKYALKYQSPYNAFDDIELIHHSLPSYDLSDIDLSTHFAGQDFDFPFYINAM 90

Query: 61  TGGSEKGKAVNHKLAQVAQATGIVMVTGSYSAALKNDEDDSYPTTDLYPDLKLATNIGLD 120
           TGGS+KGKAVN KLA+VA ATGIVMVTGSYSAALKN  DDSY   ++  +LKLATNIGLD
Sbjct: 91  TGGSQKGKAVNEKLAKVAAATGIVMVTGSYSAALKNPNDDSYRLHEVADNLKLATNIGLD 150

Query: 121 KPVPAAESTVKAMNPIFLQVHVNVMQELLMPEGEREFHMWRSHLKEYVDNIQCPLILKEV 180
           KPV   + TV+ M P+FLQVHVNVMQELLMPEGER FH W+ HL EY   I  P+ILKEV
Sbjct: 151 KPVALGQQTVQEMQPLFLQVHVNVMQELLMPEGERVFHTWKKHLAEYASQIPVPVILKEV 210

Query: 181 GFGMDLQSIKDAYDIGITTVDISGRGGTSFAYIENQRGRDRSYLNTWGQTTAQSLINAQS 240
           GFGMD+ SIK A+D+GI T DISGRGGTSFAYIENQRG DRSYLN WGQTT Q L+NAQ
Sbjct: 211 GFGMDVNSIKLAHDLGIQTFDISGRGGTSFAYIENQRGGDRSYLNDWGQTTVQCLLNAQG 270

Query: 241 MMDKMDILASGGIRHPLDMVKCLVLGAKAVGLSRTVLELVERYPVDDVIAILNSWKEDLR 300
           +MD+++ILASGG+RHPLDM+KC VLGA+AVGLSRTVLELVE+YP + VIAI+N WKE+L+
Sbjct: 271 LMDQVEILASGGVRHPLDMIKCFVLGARAVGLSRTVLELVEKYPTERVIAIVNGWKEELK 330

Query: 301 MIMCALNCKKITDLRQVNYILYGQLKEAN 329
           +IMCAL+CK I +L+ V+Y+LYG+L++ N
Sbjct: 331 IIMCALDCKTIKELKGVDYLLYGRLQQVN 359
```

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

Example 1929

A DNA sequence (GBSx2038) was identified in *S.agalactiae* <SEQ ID 5979> which encodes the amino acid sequence <SEQ ID 5980>. This protein is predicted to be phosphomevalonate kinase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0785(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:AAG02457 GB:AF290099 phosphomevalonate kinase [Streptococcus pneumoniae]
 Identities = 170/330 (51%), Positives = 233/330 (70%), Gaps = 1/330 (0%)

Query: 1   MVKVQTGGKLYIAGEYAILYPGQVAILKNVPIYMTALATFADNYSLYSDMFNYTASLQPD 60
           M+ V+T GKLY AGEYAIL PGQ+A++K++PIYM A   F+D+Y +YSDMF++   L+P+
Sbjct: 1   MIAVKTCGKLYWAGEYAILEPGQLALIKDIPIYMRAEIAFSDSYRIYSDMFDFAVDLRPN 60

Query: 61  KQYSLIQETILLMEEWLINFGKNIKPIHLEITGKLERYGLKFGIGSSGSVVVLTIKAMAA 120
             YSLIQETI LM ++L   G+N++P  L+I GK+ER G KFG+GSSGSVVVL +KA+ A
Sbjct: 61  PDYSLIQETIALMGDFLAVRGQNLRPFSLKICGKMEREGKKFGLGSSGSVVVLVVKALLA 120

Query: 121 LYEIEMPSDLLFKLSAYVLLKRGDNGSMGDIACIAYEHLISYSAFDRRAVSKMIETKPLE 180
           LY + +  +LLFKL++ VLLKRGDNGSMGD+ACI  E L+ Y +FDR+  +  +E + L
Sbjct: 121 LYNLSVDQNLLFKLTSAVLLKRGDNGSMGDLACIVAEDLVLYQSFDRQKAAAWLEEENLA 180

Query: 181 QVLEAEWGYRITKIQALLEMDFLVGWTMQPSISKEMINIVKSTITQRFLDDTKYQVVQLL 240
            VLE +WG+ I++++  LE DFLVGWT + ++S  M+  +K  I Q FL  +K  VV L+
Sbjct: 181 TVLERDWGFFISQVKPTLECDFLVGWTKEVAVSSHMVQQIKQNINQNFLSSSKETVVSLV 240

Query: 241 SAFKEGDKEAIKRCLEEISLLLFNLHPSIYTDKLQKLKEASKGLDIVTKSSGSGGGDCGI 300
            A ++G  E +   +E  S LL  L   IYT  L++LKEAS+ L  V KSSG+GGGDCGI
Sbjct: 241 EALEQGKAEKVIEQVEVASKLLEGLSTDIYTPLLRQLKEASQDLQAVAKSSGAGGGDCGI 300

Query: 301 AISFN-KNDNQTLIKRWESAGIELLSKETL 329
           A+SF+ ++   TL  RW   GIELL +E +
Sbjct: 301 ALSFDAQSSRNTLKNRWADLGIELLYQERI 330
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5981> which encodes the amino acid sequence <SEQ ID 5982>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2669(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 = 171/325 (52%), Positives = 227/325 (69%), Gaps = 2/325 (0%)

Query: 4   VQTGGKLYIAGEYAILYPGQVAILKNVPIYMTALATFADNYSLYSDMFNYTASLQPDKQY 63
           VQTGGKLY+ GEYAIL PGQ A++  +P+ MTA  + A +  L SDMF++ A + PD  Y
Sbjct: 22  VQTGGKLYLTGEYAILTPGQKALIHFIPLMMTAEISPAAHIQLASDMFSHKAGMTPDASY 81
```

```
Query: 64  SLIQETILLMEEWLINFGKNIKPIHLEITGKLERYGLKFGIGSSGSVVVLTIKAMAALYE 123
           +LIQ T+    ++L      ++P  L ITGK+ER G KFGIGSSGSV +LT+KA++A Y+
Sbjct: 82  ALIQATVKTFADYLGQSIDQLEPFSLIITGKMERDGKKFGIGSSGSVTLLTLKALSAYYQ 141

Query: 124 IEMPSDLLFKLSAYVLLKRGDNGSMGDIACIAYEHLISYSAFDRRAVSKMIETKPLEQVL 183
           I +  +LLFKL+AY LLK+GDNGSMGDIACIAY+ L++Y++FDR  VS  ++T PL+++L
Sbjct: 142 ITLTPELLFKLAAYTLLKQGDNGSMGDIACIAYQTLVAYTSFDREQVSNWLQTMPLKKLL 201

Query: 184 EAEWGYRITKIQALLEMDFLVGWTMQPSISKEMINIVKSTITQRFLDDTKYQVVQ-LLSA 242
             +WGY I  IQ  L  DFLVGWT  P+IS++MI  V ++IT  FL  T YQ+ Q  + A
Sbjct: 202 VKDWGYHIQVIQPALPCDFLVGWTKIPAISRQMIQQVTASITPAFL-RTSYQLTQSAMVA 260

Query: 243 FKEGDKEAIKRCLEEISLLLFNLHPSIYTDKLQKLKEASKGLDIVTKSSGSGGGDCGIAI 302
            +EG KE +K+ L   S LL  LHP+IY  KL  L  A +  D V KSSGSGGGDCGIA+
Sbjct: 261 LQEGHKEELKKSLAGASHLLKELHPAIYHPKLVTLVAACQKQDAVAKSSGSGGGDCGIAL 320

Query: 303 SFNKNDNQTLIKRWESAGIELLSKE 327
           +FN++   TLI +W+ A I LL +E
Sbjct: 321 AFNQDARDTLISKWQEADIALLYQE 345
```

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

Example 1930

A DNA sequence (GBSx2039) was identified in *S.agalactiae* <SEQ ID 5983> which encodes the amino acid sequence <SEQ ID 5984>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.75   Transmembrane   20 -  36 (  18 -  36)

----- Final Results -----
               bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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 1931

A DNA sequence (GBSx2040) was identified in *S.agalactiae* <SEQ ID 5985> which encodes the amino acid sequence <SEQ ID 5986>. This protein is predicted to be mevalonate diphosphate decarboxylase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1557(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:AAG02456 GB:AF290099 mevalonate diphosphate decarboxylase
          [Streptococcus pneumoniae]
```

```
 Identities = 219/312 (70%); Positives = 264/312 (84%)

Query: 1   MDGKSISVKSYANIAIIKYWGKADAEKMIPATSSISLTLENMYTETRLTALGKDAKKDEF 60
           MD + ++V+SYANIAIIKYWGK   ++M+PATSSISLTLENMYTET L+ L  +   DEF
Sbjct: 1   MDREPVTVRSYANIAIIKYWGKKKEKEMVPATSSISLTLENMYTETTLSPLPANVTADEF 60

Query: 61  YISGVLQNDHEHDKMSAILDRFRQNRSGFVKIETTNNMPTAAGLSSSSSGLSALVKACND 120
           YI+G LQN+ EH KMS I+DR+R    GFV+I+T NNMPTAAGLSSSSSGLSALVKACN
Sbjct: 61  YINGQLQNEVEHAKMSKIIDRYRPAGEGFVRIDTQNNMPTAAGLSSSSSGLSALVKACNA 120

Query: 121 FFGTNLSQSQLAQEAKFASGSSSRSFFGPVAAWDKDSGDIYKVHTNLDLAMIMLVLNDKR 180
           +F   L +SQLAQEAKFASGSSSRSF+GP+ AWDKDSG+IY V T+L LAMIMLVL DK+
Sbjct: 121 YFKLGLDRSQLAQEAKFASGSSSRSFYGPLGAWDKDSGEIYPVETDLKLAMIMLVLEDKK 180

Query: 181 KPISSREGMKICTETSTTFNEWVRQSEQDYQDMLVYLKNNDFQKVGQLTERNALAMHSTT 240
           KPISSR+GMK+C ETSTTF++WVRQSE+DYQDML+YLK NDF K+G+LTE+NALAMH+TT
Sbjct: 181 KPISSRDGMKLCVETSTTFDDWVRQSEKDYQDMLIYLKENDFAKIGELTEKNALAMHATT 240

Query: 241 KTATPAFSYLTEETYKAMDVVKKLREKGHECYYTMDAGPNVKVLCLRQDLEALAAILEKD 300
           KTA+PAFSYLT+ +Y+AM  V++LREKG  CY+TMDAGPNVKV C  +DLE L+ I  +
Sbjct: 241 KTASPAFSYLTDASYEAMAFVRQLREKGEACYFTMDAGPNVKVFCQEKDLEHLSEIFGQR 300

Query: 301 YRIIVSTTKELA 312
           YR+IVS TK+L+
Sbjct: 301 YRLIVSKTKDLS 312
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5987> which encodes the amino acid sequence <SEQ ID 5988>. Analysis of this protein sequence reveals the following:

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

------ Final Results ------
              bacterial cytoplasm --- Certainty=0.1271(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 = 221/313 (70%), Positives = 258/313 (81%)

Query: 1   MDGKSISVKSYANIAIIKYWGKADAEKMIPATSSISLTLENMYTETRLTALGKDAKKDEF 60
           +D   I+V SYANIAIIKYWGK +  KMIP+TSSISLTLENM+T T ++ L  A  D+F
Sbjct: 1   VDPNVITVTSYANIAIIKYWGKENQAKMIPSTSSISLTLENMFTTTSVSFLPDTATSDQF 60

Query: 61  YISGVLQNDHEHDKMSAILDRFRQNRSGFVKIETTNNMPTAAGLSSSSSGLSALVKACND 120
           YI+G+LQND EH K+SAI+D+FRQ     FVK+ET NNMPTAAGLSSSSSGLSALVKAC+
Sbjct: 61  YINGILQNDEEHTKISAIIDQFRQPGQAFVKMETQNNMPTAAGLSSSSSGLSALVKACDQ 120

Query: 121 FFGTNLSQSQLAQEAKFASGSSSRSFFGPVAAWDKDSGDIYKVHTNLDLAMIMLVLNDKR 180
            F T L Q  LAQ+AKFASGSSSRSFFGPVAAWDKDSG IYKV T+L +AMIMLVLN  +
Sbjct: 121 LFDTQLDQKALAQKAKFASGSSSRSFFGPVAAWDKDSGAIYKVETDLKMAMIMLVLNAAK 180

Query: 181 KPISSREGMKICTETSTTFNEWVRQSEQDYQDMLVYLKNNDFQKVGQLTERNALAMHSTT 240
           KPISSREGMK+C +TSTTF++WV QS  DYQ ML YLK N+F+KVGQLTE NALAMH+TT
Sbjct: 181 KPISSREGMKLCRDTSTTFDQWVEQSAIDYQHMLTYLKTNNFEKVGQLTEANALAMHATT 240

Query: 241 KTATPAFSYLTEETYKAMDVVKKLREKGHECYYTMDAGPNVKVLCLRQDLEALAAILEKD 300
           KTA P FSYLT+E+Y+AM+ VK+LR++G  CY+TMDAGPNVKVLCL +DL  LA  L K+
Sbjct: 241 KTANPPFSYLTKESYQAMEAVKELRQEGFACYFTMDAGPNVKVLCLEKDLAQLAERLGKN 300

Query: 301 YRIIVSTTKELAD 313
           YRIIVS TK+L D
Sbjct: 301 YRIIVSKTKDLPD 313
```

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

Example 1932

A DNA sequence (GBSx2041) was identified in *S.agalactiae* <SEQ ID 5989> which encodes the amino acid sequence <SEQ ID 5990>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1512(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 5991> which encodes the amino acid sequence <SEQ ID 5992>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1117(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
Identities = 182/290 (62%), Positives = 223/290 (76%)

Query: 1   MKEKFGIGKAHSKIILMGEHSVVYGYPAIAIPLKNIEVTCLIEEAPQLIALDMTDPLSTA 60
           M E  G GKAHSKIIL+GEH+VVYGYPAIA+PL +IEV C I  A + +  D  D LSTA
Sbjct: 6   MNENIGYGKAHSKIILIGEHAVVYGYPAIALPLTDIEVVCHIFPADKPLVFDFYDTLSTA 65

Query: 61  IFAALDYLGKTSSKIAYHIESQVPERRGMGSSAAVAIAAIRAVFDYFDEDLEADLLECLV 120
           I+A+LDYL +    IAY I SQVP++RGMGSSAAV+IAAIRAVF Y  E L  DLLE LV
Sbjct: 66  IYASLDYLQRLQEPIAYEIVSQVPQKRGMGSSAAVSIAAIRAVFSYCQEPLSDDLLEILV 125

Query: 121 NRAEMIAHSNPSGLDAKTCLSENTIKFIRNIGFSTVPMHLNAYLVIADTGIHGHTKEAVD 180
           N+AE+IAH+NPSGLDAKTCLS++ IKFIRNIGF T+ + LN YL+IADTGIHGHT+EAV+
Sbjct: 126 NKAEIIAHTNPSGLDAKTCLSDHAIKFIRNIGFETIEIALNGYLIIADTGIHGHTREAVN 185

Query: 181 KVKSSGEAVLPFLKELGYLAEASEDAIHKSDSKQLGSLMTKAHQSLKQLGVSSLEADHLV 240
           KV    E  LP+L +LG L +A E AI++ +   +G LMT+AH +LK +GVS  +AD LV
Sbjct: 186 KVAQFEETNLPYLAKLGALTQALERAINQKNKVAIGQLMTQAHSALKAIGVSISKADQLV 245

Query: 241 EVAISCGALGAKMSGGGLGGCIIALVKEKREAERLSQQLEREGAVNTWTE 290
           E A+  GALGAKM+GGGLGGC+IAL   K  AE++S +L+ EGAVNTW +
Sbjct: 246 EAALRAGALGAKMTGGGLGGCMIALADTKDMAEKISHRLKEEGAVNTWIQ 295
```

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

Example 1933

A DNA sequence (GBSx2042) was identified in *S.agalactiae* <SEQ ID 5993> which encodes the amino acid sequence <SEQ ID 5994>. This protein is predicted to be a histidine protein kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.43   Transmembrane   12 -  28 (   4 -  33)
```

```
   INTEGRAL    Likelihood = -9.29   Transmembrane  163 - 179 ( 157 - 191)

----- Final Results -----
               bacterial membrane --- Certainty=0.6371(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAF79919 GB:AF039082 putative histidine protein kinase
          [Lactococcus lactis]
 Identities = 78/315 (24%), Positives = 154/315 (48%), Gaps = 33/315 (10%)

Query: 101 SDRQIKNYAKRIVSQNSHSGHITYNFSTYSYLLKKVGKNDYLVVFLDTTNQYLDNQRLLQ 160
           +++QI N  + +  +N +  +  Y + T S  +           V++   +     Q  +
Sbjct: 84  NEKQI-NTIQTVSVKNPYGDNWHYRYLTTSQFIITNSDGTVTPVYVQIFSNVDQIQDAMS 142

Query: 161 LSIWM---SLVSFIVFMVIVSV-LSGRVILPFVANYEKQRRFITNAGHELKTPLAIISAN 216
            ++W+   ++++F +  VI+S+ L+   + P +A YEKQ+ F+ NA HEL+TPLAI+
Sbjct: 143 RAMWVIVTTMITFWILSVIISLYLANWTLKPILAAYEKQKEFVENASHELRTPLAILQNR 202

Query: 217 NELV-----EMMSGESEWTKSTNDQIQRLTGLINGMVSLAR------FEEQPDISM---- 261
            EL+       +  +SE    +  +++ +  L + +++LAR       E +P  +
Sbjct: 203 LELLFQKPTATIIDQSENISESLSEVRNMRLLTSNLLNLARRDSGIKIEPEPTTATYFEN 262

Query: 262 VDLDFSHITKDAAEDFKGPIIKDGKDFIMSIQPGIHVKAEEKSLFELVTLLVDNANKYCD 321
           +   +  +T++A + F G +  +G            V  ++  + +L+T+L DNA KY D
Sbjct: 263 IFNSYEMLTENAGKKFSGNLKLEGT-----------VNLDQALIKQLLTILFDNALKYTD 311

Query: 322 PMGTVTVKLSRSSRLRRAKLEVSNTYKNGKDIDYSKFFERFYREDESHNNKKSGYGIGLS 381
             G ++V + ++         V++  +   D D  K F+RF+R D++   +K G G+GLS
Sbjct: 312 SEGEISVDVIKNGGF--LTFAVADNGEGISDEDKKKIFDRFFRVDKARTRQKGGLGLGLS 369

Query: 382 IVTSLVHLFKGSIDV 396
           +   +V  + G I V
Sbjct: 370 LAKQIVEAYNGKITV 384
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5751> which encodes the amino acid sequence <SEQ ID 5752>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.30   Transmembrane   18 -  34 (  13 -  42)
   INTEGRAL    Likelihood =-10.35   Transmembrane  170 - 186 ( 163 - 199)

----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 233/410 (56%), Positives = 303/410 (73%), Gaps = 1/410 (0%)

Query: 1   MFRNLRLRFIGIAALAILVVLFSVVGVLNSANHYQTKNEIYRVLTILADNNGRIPNKLEF 60
           MF  +R+RFI IA++AI ++L S+VG++N+A  YQ++ EI R+L +++ N G++P   E
Sbjct: 10  MFNRIRIRFIMIASIAIFIILSSIVGIINTARCYQSQQEINRILHLISSNKGKLPGTTES 69

Query: 61  SKELGDDLSTDAIFQFRYFSARTDAKGNVTSFDSRNIFEVSDRQIKNYAKRIVSQNSHSG 120
           SK LG  LS D++ QFRY+S   +A G++ S ++ NI  +   + + +A+         G
Sbjct: 70  SKRLGTKLSEDSLSQFRYYSVIFNANGHLLSSNTANISALDREEAQYFARLFAKSGEEKG 129

Query: 121 HITYNFSTYSYLLKKVGKNDYLVVFLDTTNQYLDNQRLLQLSIWMSLVSFIVFMVIVSVL 180
              + S YSYL+ ++   + LVV LDTT  +     LL +S+ ++   FI F+V+VS+
Sbjct: 130 SYRHQDSVYSYLITQLPNEEKLVVILDTTFYFRSVGDLLAVSVMLAFGGFIFFVVLVSLF 189

Query: 181 SGRVILPFVANYEKQRRFITNAGHELKTPLAIISANNELVEMMSGESEWTKSTNDQIQRL 240
```

```
            SG VI PFV NYEKQRRFITNAGHELKTPLAIISANNELVE+M+GESEWTKST+DQ++RL
Sbjct: 190 SGMVIKPFVQNYEKQRRFITNAGHELKTPLAIISANNELVELMTGESEWTKSTSDQVKRL 249

Query: 241 TGLINGMVSLARFEEQPDISMVDLDFSHITKDAAEDFKGPIIKDGKDFIMSIQPGIHVKA 300
           TGLIN M++LAR EEQPD+ +  +DFS I +DAAEDFK  ++KDGK F ++IQP I +KA
Sbjct: 250 TGLINQMITLARLEEQPDVVLHMVDFSAIAQDAAEDFKSLVLKDGKRFDLTIQPNIMIKA 309

Query: 301 EEKSLFELVTLLVDNANKYCDPMGTVTVKLSRSSRLR-RAKLEVSNTYKNGKDIDYSKFF 359
           EEKSLFELVT+LVDNANKYCDP G V V L+   R R RAKLEVSNTY  GK IDYS+FF
Sbjct: 310 EEKSLFELVTILVDNANKYCDPKGLVKVSLTTIGRRRKRAKLEVSNTYLEGKSIDYSRFF 369

Query: 360 ERFYREDESHNNKKSGYGIGLSIVTSLVHLFKGSIDVNYKHDTITFVIYI 409
           ERFYREDESHN+K+ GYGIGLS+  S+V LFKG+I VNYK+D I F + I
Sbjct: 370 ERFYREDESHNSKEKGYGIGLSMAESMVKLFKGTITVNYKNDAIVFTVVI 419
```

SEQ ID 5994 (GBS273) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 14; MW 46kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 5; MW 71kDa).

GBS273-GST was purified as shown in Figure 208, lane 4.

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

Example 1934

A DNA sequence (GBSx2043) was identified in *S.agalactiae* <SEQ ID 5995> which encodes the amino acid sequence <SEQ ID 5996>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2181(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 1935

A DNA sequence (GBSx2044) was identified in *S.agalactiae* <SEQ ID 5997> which encodes the amino acid sequence <SEQ ID 5998>. This protein is predicted to be two-component response regulator (trcR). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2503(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 9379> which encodes amino acid sequence <SEQ ID 9380> was also identified.

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

```
>GP:BAB04091 GB:AP001508 two-component response regulator [Bacillus halodurans]
 Identities = 71/183 (38%), Positives = 120/183 (64%), Gaps = 3/183 (1%)

Query: 9   RVLIAEDEEQMSRVLSTAISHQGYVVDVAYDGQTAIDLANQNAYDVMVMDVMMPVKTGIE 68
           R+LI EDE++++RVL   + H+GY  D A+ G   ++    +A+D++++DVM+P  +G+E
Sbjct: 3   RILIIEDEKKIARVLQLELEHEGYETDAAFSGSDGLETFQAHAWDLVLLDVMLPELSGLE 62

Query: 69  AVKEIRQSGNKSHIIMLTAMAEIDDRVTGLDAGADDYLTKPFSLKELLARLRSMSRRLE- 127
            ++ IR +   + II+LTA   I D+V+GLD GA+DY+TKPF ++ELLAR+R+  R ++
Sbjct: 63  VLRRIRMTDPVTPIILLTARNSIPDKVSGLDLGANDYITKPFEIEELLARVRACLRTVQT 122

Query: 128 -DFTPNVLSLGRVTLSVGEQELQCEN-TIRLAGKEAKMLAFFMLNHDKELSTQQLFEHVW 185
            +   + L    +T++   +++Q  N TI L  KE ++L FF+ N  + LS +Q+  +VW
Sbjct: 123 RERVEDTLMFQELTINEKTRDVQRGNETIELTPKEFELLVFFIKNKGQVLSREQILTNVW 182

Query: 186 GAD 188
           G D
Sbjct: 183 GFD 185
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5999> which encodes the amino acid sequence <SEQ ID 6000>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2391(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 = 125/185 (67%), Positives = 151/185 (81%)

Query: 8   MRVLIAEDEEQMSRVLSTAISHQGYVVDVAYDGQTAIDLANQNAYDVMVMDVMMPVKTGI 67
           M++L+AEDE QMS VL+TA++HQGY VDV ++GQ AID A  NAYD+M++D+MMP+K+GI
Sbjct: 1   MKILLAEDEWQMSNVLTTAMTHQGYDVDVVFNGQEAIDKAKDNAYDIMILDIMMPIKSGI 60

Query: 68  EAVKEIRQSGNKSHIIMLTAMAEIDDRVTGLDAGADDYLTKPFSLKELLARLRSMSRRLE 127
           EA+KEIR SGN SHIIMLTAMAEI+DRVTGLDAGADDYLTKPFSLKELLARLRSM RR+E
Sbjct: 61  EALKEIRASGNCSHIIMLTAMAEINDRVTGLDAGADDYLTKPFSLKELLARLRSMERRVE 120

Query: 128 DFTPNVLSLGRVTLSVGEQELQCENTIRLAGKEAKMLAFFMLNHDKELSTQQLFEHVWGA 187
            FTP VL    VTL++ EQEL   N TRLA KE K+LAF M+N  K L  +++ E+VWG
```

```
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:BAB05604 GB:AP001513 unknown conserved protein [Bacillus halodurans]
 Identities = 67/182 (36%), Positives = 111/182 (60%), Gaps = 4/182 (2%)

Query: 17  LEDFSQRIQLENDKAKVETGYKLYEHIIGRIKTSDSMIEKCRRKQLPVTVDSALKTIRDS 76
           L++ + +I +   + +    Y   EH+  R+K+ +S++ K +R+    T++S  + +RD
Sbjct: 29  LQELNTKIDILKQEFQYIHDYNPIEHVSSRVKSPESIVNKIQRRGNDFTLESIRENVRDI 88

Query: 77  IGVRIICGFVNDIYQIIERIKAFDDCRIVVEKDYIQHVKPNGYRSYHVILEIDTPYPDCL 136
            G+RI C F +DIY + E++    D  +V  KDYI++ KPNGYRS H+IL I    P  +
Sbjct: 89  AGIRITCSFESDIYTLSEQLMQQHDISVVETKDYIKNPKPNGYRSLHLILSI----PIFM 144

Query: 137 GNSDGKYYIEIQLRTIAQDSWASLEHQMKYKHDIENPERIVRELKRCADEMASVDLTMQT 196
            +    Y+E+Q+RTIA D WASLEH++ YK++   PE +++ELK  A+  A +D  M+
Sbjct: 145 SDRVQDVYVEVQIRTIAMDFWASLEHKIYYKYNKNVPEHLLKELKDAAESAALLDQKMEK 204

Query: 197 IR 198
           I+
Sbjct: 205 IQ 206
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6003> which encodes the amino acid sequence <SEQ ID 6004>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1057(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 = 127/206 (61%), Positives = 162/206 (77%)

Query: 3   TNIYGDYGRYLPLILEDFSQRIQLENDKAKVETGYKLYEHIIGRIKTSDSMIEKCRRKQL 62
           ++IY  +  YLPL+L+  +  I  EN K+K ETG+KLYEH   RIK+  SMIEKC+RKQL
Sbjct: 11  SSIYSGFEVYLPLVLQTITDVIIAENIKSKKETGFKLYEHFTSRIKSEASMIEKCQRKQL 70

Query: 63  PVTVDSALKTIRDSIGVRIICGFVNDIYQIIERIKAFDDCRIVVEKDYIQHVKPNGYRSY 122
           P+T  SALK I+DSIG+RIICGF++DIY++++ +K+     +  EKDYI + KPNGYRSY
Sbjct: 71  PLTSKSALKIIKDSIGIRIICGFIDDIYRMVDLLKSIPGMSVNTEKDYILNAKPNGYRSY 130

Query: 123 HVILEIDTPYPDCLGNSDGKYYIEIQLRTIAQDSWASLEHQMKYKHDIENPERIVRELKR 182
           H+ILE++T +PD LG   G Y+IE+QLRTIAQDSWASLEHQMKYKH + N E I RELKR
Sbjct: 131 HLILELETHFPDILGEKKGCYFIEVQLRTIAQDSWASLEHQMKYKHQVANAEMITRELKR 190

Query: 183 CADEMASVDLTMQTIRQLIESGTKKE 208
           CADE+AS D+TMQTIRQLI+  T++E
Sbjct: 191 CADELASCDVTMQTIRQLIQETTEEE 216
```

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

Example 1937

A DNA sequence (GBSx2046) was identified in *S.agalactiae* <SEQ ID 6005> which encodes the amino acid sequence <SEQ ID 6006>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3250(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:CAA37193 GB:X53013 ORF1 (AA 1 - 384) [Lactococcus lactis]
 Identities = 30/55 (54%), Positives = 37/55 (66%)

Query: 1   MEFYYKTLKRKFINDADTIFIEQSQFEIFIYIETDHNSSSSHVVLDYQSQKEFEK 55
           ME +YKTLKR+ INDA      ++  EIF YIET +N+    H  LDYQS K+FEK
Sbjct: 327 MESFYKTLKRELINDAHFETRAEATQEIFKYIETYYNTKWMHSGLDYQSPKDFEK 381
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6007> which encodes the amino acid sequence <SEQ ID 6008>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3065(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 31/59 (52%), Positives = 39/59 (65%)

Query: 1  MEFYYKTLKRKFINDADTIFIEQSQFEIFIYIETDHNSSSSHVVLDYQSQKEFEKIITN 59
          ME +YKTLKR+ +NDA     I+Q+Q EIF Y ET +N     H  L Y S  EFEKI+T+
Sbjct: 13 MEAFYKTLKRELVNDAHFATIKQAQLEIFKYSETYYNPKRLHSALGYLSPVEFEKIVTH 71
```

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

Example 1938

A DNA sequence (GBSx2047) was identified in *S.agalactiae* <SEQ ID 6009> which encodes the amino acid sequence <SEQ ID 6010>. This protein is predicted to be R5 protein. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.98   Transmembrane   30 -  46 (  29 -  51)
   INTEGRAL    Likelihood = -2.76   Transmembrane  967 - 983 ( 966 - 985)

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

A related GBS nucleic acid sequence <SEQ ID 8935> which encodes amino acid sequence <SEQ ID 8936> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
SRCFLG: 0
McG: Length of UR:   2
     Peak Value of UR:   2.44
     Net Charge of CR: 2
McG: Discrim Score:      0.78
GvH: Signal Score (-7.5): -0.0599995
```

```
     Possible site: 39
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 40
ALOM program   count: 0 value:   7.37 threshold:  0.0
   PERIPHERAL  Likelihood =  7.37.    194
 modified ALOM score:  -1.97

*** 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>

LPXTG motif: 944-948
```

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

SEQ ID 8936 (GBS200) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 3; MW 107.4kDa), in Figure 169 (lane 4; MW 122kDa) and in Figure 238 (lane 11; MW 122kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 3; MW 132kDa).

Purified Thio-GBS200-His is shown in Figure 244, lane 9.

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

Example 1939

A DNA sequence (GBSx2048) was identified in *S.agalactiae* <SEQ ID 6011> which encodes the amino acid sequence <SEQ ID 6012>. This protein is predicted to be a 16.1 kDa transcriptional regulator. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3919(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 9953> which encodes amino acid sequence <SEQ ID 9954> was also identified.

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

```
>GP:CAB16108 GB:Z99124 similar to transcriptional regulator (MarR
          family) [Bacillus subtilis]
 Identities = 30/114 (26%), Positives = 59/114 (51%), Gaps = 3/114 (2%)

Query: 29  DVEHLAGPQGHLVMYLYKHPDKDMSIKAVEEILHISKSVASNLVKRMEKNGFIAIVPSKT 88
           D++   G   +LV  +Y++P   +  + + E++ + ++ A+  +K++E  GFI  +P +
Sbjct: 25  DLDLTRGQYLYLVR-IYENPG--IIQEKLAEMIKVDRTTAARAIKKLEMQGFIQKLPDEQ 81

Query: 89  DKRVKYLYLTHLGKKKATQFEIFLEKLHSTMLAGITKEEIRTTKKVIRTLAKNM 142
           +K++K L+ T  GKK          E       L+G T EE  T   ++  + KN+
Sbjct: 82  NKKIKKLFPTEKGKKVYPLLRREGEHSTEVALSGFTSEEKETISALLHRVRKNI 135
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6013> which encodes the amino acid sequence <SEQ ID 6014>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4175(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 = 27/64 (42%), Positives = 46/64 (71%)

Query: 3  MENPLQKARILVNQLEKYLDHYAKEYDVEHLAGPQGHLVMYLYKHPDKDMSIKAVEEILH 62
          M   +   R L++Q+E+  D  AK+YDVEHLAGPQG+++++L KH ++++ +K +E+ L
Sbjct: 1  MSQVIGDLRELIHQIEQISDEIAKKYDVEHLAGPQGYVLVFLAKHQNQEIFVKDIEKQLR 60

Query: 63 ISKS 66
          I +S
Sbjct: 61 IFQS 64
```

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

Example 1940

A DNA sequence (GBSx2049) was identified in *S.agalactiae* <SEQ ID 6015> which encodes the amino acid sequence <SEQ ID 6016>. This protein is predicted to be 5`-nucleotidase family protein. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -2.66   Transmembrane  668 - 684 ( 665 - 684)

----- 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:CAB12747 GB:Z99108 similar to 5'-nucleotidase [Bacillus subtilis]
 Identities = 178/535 (33%), Positives = 270/535 (50%), Gaps = 55/535 (10%)

Query: 28   DQVGVQVIGVNDFHGALDNTGTANMPDGKVANAGTAAQLD---AYMDDAQKDFKQTNPNG 84
            + V ++++ +ND HG +D     ++ DG     GT  ++D   AY+ + + + K
Sbjct: 586  EHVPLRILSMNDLHGKIDQQYELDL-DGNGTVDGTFGRMDYAAAYLKEKKAEKKN----- 639

Query: 85   ESIRVQAGDMVGASPANSGLLQDEPTVKNFNAMNVEYGTLGNHEFDEGLAEYNRIVTGKA 144
             S+ V AGDM+G S   S LLQDEPTV+    +  + GT+GNHEFDEG  E  RI+ G
Sbjct: 640  -SLIVHAGDMIGGSSPVSSLLQDEPTVELMEDIGFDVGTVGNHEFDEGTDELLRILNG-G 697

Query: 145  PAPDSNINNITKSYPHEAAKQEIVVANVIDKVNKQIPYNWKPYAIKNIPVNNKSVNVGFI 204
              P        +++P       +V AN       ++    +P+      +N + V V FI
Sbjct: 698  DHPKGTSGYDGQNFP-------LVCANC------KMKSTGEPFLPAYDIINVEGVPVAFI 744

Query: 205  GIVTKDIPNLVLRKNYEQYEFLDEAETIVKYAKELQAKNVKAIVVLAHVPATSKNDIAEG 264
            G+VT+    +V+ +  +  EF DEA  + K A+EL+ K VKAI VLAH+ A    +  G
Sbjct: 745  GVVTQSAAGMVMPEGIKNIEFTDEATAVNKAAEELKKKGVKAIAVLAHMSAEQNGNAITG 804

Query: 265  EAAEMMKKVNQLFPENSVDIVFAGHNHQYTNGLVGKTRIVQALSQGKAYADVRGVLDTDT 324
            E+A++  K      ++ +D++FA HNHQ  NG V    IVQA   GKA   V   +D  T
```

```
Sbjct: 805  ESADLANKT-----DSEIDVIFAAHNHQVVNGEVNGKLIVQAFEYGKAIGVVDVEIDKTT 859

Query: 325  QDFIETPSAKVIAVAPGKKTGSADIQAIVDQANTIVKQVTEAKIGTAEVSVMITRSVDQD 384
             +D ++  SA+++.V   K        AI+ +  TI + +    +G A V +    S D D
Sbjct: 860  KDIVK-KSAEIVYVDQSKIEPDVSASAILKKYETIAEPIISEVVGEAAVDMEGGYSNDGD 918

Query: 385  NVSPVGSLITEAQLAIARKSWPDIDFAMTNNGGIRADLLIKPDGTITWGAAQAVQPFGNI 444
               +P+G+LI +   A  +      DFA+ N GGIR  L     G ITWG    +QPFGN+
Sbjct: 919  --TPLGNLIADGMRAAMK-----TDFALMNGGGIREAL---KKGPITWGDLYNIQPFGNV 968

Query: 445  LQVVEITGRDLYKALNEQYDQKQNFFLQIAGLRYTYTDNKEGGEETPFKVVKAYKSNGEE 504
             L  +EI G+DL + +N Q          I+G  +TYT +KE G+    K+     ++G E
Sbjct: 969  LTKLEIKGKDLREIINAQISPVFGPDYSISG--FTYTWDKETGKAVDMKM-----ADGTE 1021

Query: 505  INPDAKYKLVINDFLFGGGDGFASFRNAKLLGAINP-----DTEVFMAYITDLEK 554
             I PDA Y L +N+F+        A ++   LLG  NP     D E  + Y+   ++
Sbjct: 1022 IQPDATYTLTVNNFMATATG--AKYQPIGLLGK-NPVTGPEDLEATVEYVKSFDE 1073
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1607> which encodes the amino acid sequence <SEQ ID 1608>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.67   Transmembrane  662 - 678 ( 661 - 679)
   INTEGRAL    Likelihood = -2.02   Transmembrane   19 -  35 (  18 -  35)

----- Final Results -----
              bacterial membrane --- Certainty=0.2869(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 415/688 (60%), Positives = 517/688 (74%), Gaps = 21/688 (3%)

Query: 1    MKKKIILKSSVLGLVAGTSIMFSSVFADQVGVQVIGVNDFHGALDNTGTANMPDGKVANA 60
             MKK  ILKSSVL ++    +++ + V ADQV VQ +GVNDFHGALDNTGTA  P GK+ NA
Sbjct: 14   MKKYFILKSSVLSILTSFTLLVTDVQADQVDVQFLGVNDFHGALDNTGTAYTPSGKIPNA 73

Query: 61   GTAAQLDAYMDDAQKDFKQTNPNGESIRVQAGDMVGASPANSGLLQDEPTVKNFNAMNVE 120
             GTAAQL AYMDDA+ DFKQ N +G SIRVQAGDMVGASPANS LLQDEPTVK FN M  E
Sbjct: 74   GTAAQLGAYMDDAEIDFKQANQDGTSIRVQAGDMVGASPANSALLQDEPTVKVFNKMKFE 133

Query: 121  YGTLGNHEFDEGLAEYNRIVTGKAPAPDSNINNITKSYPHEAAKQEIVVANVIDKVNKQI 180
             YGTLGNHEFDEGL E+NRI+TG+AP P+S IN+ITK Y HEA+ Q IV+ANVIDK  K I
Sbjct: 134  YGTLGNHEFDEGLDEFNRIMTGQAPDPESTINDITKQYEHEASHQTIVIANVIDKKTKDI 193

Query: 181  PYNWKPYAIKNIPVNNKSVNVGFIGIVTKDIPNLVLRKNYEQYEFLDEAETIVKYAKELQ 240
             PY WKPYAIK+I +N+K V +GFIG+VT +IPNLVL++NYE Y+FLD AETI KYAKELQ
Sbjct: 194  PYGWKPYAIKDIAINDKIVKIGFIGVVTTEIPNLVLKQNYEHYQFLDVAETIAKYAKELQ 253

Query: 241  AKNVKAIVVLAHVPATSKNDIAEGEAAEMMKKVNQLFPENSVDIVFAGHNHQYTNGLVGK 300
              ++V AIVVLAHVPATSK+ + + E A +M+KVNQ++PE+S+DI+FAGHNHQYTNG +GK
Sbjct: 254  EQHVHAIVVLAHVPATSKDGVVDHEMATVMEKVNQIYPEHSIDIIFAGHNHQYTNGTIGK 313

Query: 301  TRIVQALSQGKAYADVRGVLDTDTQDFIETPSAKVIAVAPGKKTGSADIQAIVDQANTIV 360
             TRIVQALSQGKAYADVRG LDTDT DFI+TPSA V+AVAPG KT ++DI+AI++ AN IV
Sbjct: 314  TRIVQALSQGKAYADVRGTLDTDTNDFIKTPSANVVAVAPGIKTENSDIKAIINHANDIV 373

Query: 361  KQVTEAKIGTAEVSVMITRSVDQDNVSPVGSLITEAQLAIARKSWPDIDFAMTNNGGIRA 420
             K VTE KIGTA  S  I+++ + D  SPVG+L T AQL IA+K++P +DFAMTNNGGIR+
Sbjct: 374  KTVTERKIGTATNSSTISKTENIDKESPVGNLATTAQLTIAKKTFPTVDFAMTNNGGIRS 433

Query: 421  DLLIKPDGTITWGAAQAVQPFGNILQVVEITGRDLYKALNEQYDQKQNFFLQIAGLRYTY 480
             DL++K D TITWGAAQAVQPFGNILQV+++TG+ +Y  LN+QYD+ Q +FLQ++GL YTY
Sbjct: 434  DLVVKNDRTITWGAAQAVQPFGNILQVIQMTGQHIYDVLNQQYDENQTYFLQMSGLTYTY 493
```

```
Query: 481 TDNKEGGEETPFKVVKAYKSNGEEINPDAKYKLVINDFLFGGGDGFASFRNAKLLGAINP 540
           TDN      +TPFK+VK YK NGEEIN     Y +V+NDFL+GGGDGF++F+ AKL+GAIN
Sbjct: 494 TDNDPKNSDTPFKIVKVYKDNGEEINLTTTYTVVVNDFLYGGGDGFSAFKKAKLIGAINT 553

Query: 541 DTEVFMAYITDLEKAGKKVSVPNNKPKIYVTMKMVNETITQNDGTHSIIKKLYLDRQGNI 600
           DTE F+ YIT+LE +GK V+       K YVT  + + T    + G HSII K++ +R GN
Sbjct: 554 DTEAFITYITNLEASGKTVNATIKGVKNYVTSNLESSTKVNSAGKHSIISKVFRNRDGNT 613

Query: 601 VAQEIVSDTLNQTKSKSTKINPVTTIHKKQLHQFTAINPMRNYGKPSNSTTVKSKQLPKT 660
           V+ E++SD L  T++ +  +    T                      +N  T+ S  LP T
Sbjct: 614 VSSEVISDLLTSTENTNNSLGKKET--------------------TTNKNTISSSTLPIT 653

Query: 661 NSEYGQSFLMSVFG-VGLIGIALNTKKK 687
              Y  S +M++   + L G+    KK+
Sbjct: 654 GDNYKMSPIMTILALISLGGLNAFIKKR 681
```

SEQ ID 6016 (GBS328) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 4; MW 73kDa). The GBS328-His fusion product was purified (Figure 213, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 268), 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 1941

A DNA sequence (GBSx2050) was identified in *S.agalactiae* <SEQ ID 6017> which encodes the amino acid sequence <SEQ ID 6018>. This protein is predicted to be peptide deformylase (def-2). Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.70   Transmembrane   55 -  71 (  55 -  74)

----- Final Results -----
               bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB09662 GB:Z96934 peptide deformylase [Clostridium
           beijerinckii]
 Identities = 71/136 (52%), Positives = 96/136 (70%)

Query: 1   MIKPIVRDTFFLQQKSQMASRADVSLAKDLQETLHANQNYCVGMAANMIGSLKRVIIINV 60
           MIKPIV+D  FL QKS+ A++ D+ +  DL +TL AN  +CVG+AANMIG  KR+++  V
Sbjct: 1   MIKPIVKDILFLGQKSEEATKNDMVVIDDLIDTLRANLEHCVGLAANMIGVKKRILVFTV 60

Query: 61  GITNLVMFNPVVVAKSDPYETEESCLSLVGCRSTQRYCHITISYRDINWKEQQIKLTDFP 120
           G   + M NPV++ K  PYETEESCLSL+G R T+RY  I ++Y D N+ +++     F
Sbjct: 61  GNLIVPMINPVILKKEKPYETEESCLSLIGFRKTKRYETIEVTYLDRNFNKKKQVFNGFT 120

Query: 121 AQICQHELDHLEGILI 136
           AQI QHE+DH EGI+I
Sbjct: 121 AQIIQHEMDHFEGIII 136
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6019> which encodes the amino acid sequence <SEQ ID 6020>. Analysis of this protein sequence reveals the following:

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

```
   INTEGRAL    Likelihood = -3.61   Transmembrane   55 -  71 (  55 -  73)

----- Final Results -----
               bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 77/136 (56%), Positives = 103/136 (75%)

Query: 1   MIKPIVRDTFFLQQKSQMASRADVSLAKDLQETLHANQNYCVGMAANMIGSLKRVIIINV 60
           MI+ I+ D F LQQK+Q+A + D+ + +DLQ+TL    +  C+GMAANMIG  KR++I+++
Sbjct: 1   MIREIITDHFLLQQKAQVAKKEDLWIGQDLQDTLAFYRQECLGMAANMIGEQKRIVIVSM 60

Query: 61  GITNLVMFNPVVVAKSDPYETEESCLSLVGCRSTQRYCHITISYRDINWKEQQIKLTDFP 120
           G  +LVMFNPV+V+K   Y+T+ESCLSL G R TQRY  IT+ Y D NW+ +++ LT
Sbjct: 61  GFIDLVMFNPVMVSKKGIYQTKESCLSLSGYRKTQRYDKITVEYLDHNWRPKRLSLTGLT 120

Query: 121 AQICQHELDHLEGILI 136
           AQICQHELDHLEGILI
Sbjct: 121 AQICQHELDHLEGILI 136
```

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

Example 1942

A DNA sequence (GBSx2051) was identified in *S.agalactiae* <SEQ ID 6021> which encodes the amino acid sequence <SEQ ID 6022>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2880(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:BAB05820 GB:AP001514 NADP-specific glutamate dehydrogenase
          [Bacillus halodurans]
 Identities = 298/444 (67%), Positives = 362/444 (81%), Gaps = 2/444 (0%)

Query: 7   YVASVLEKVKKQNEHEEEFLQAVEEVFESLVPVFDKYPQYIEENLLERLVEPERVISFRV 66
           YV  V E VK++N +E EF QAV+EVF+SL+PV  K+PQY+++ +LER+VEPERVISFRV
Sbjct: 16  YVQHVYETVKRRNPNEHEFHQAVKEVFDSLLPVLVKHPQYVKQAILERIVEPERVISFRV 75

Query: 67  PWVDDKGQVQVNRGYRVQFSSAIGPYKGGLRFHPTVTQSIVKFLGFEQIFKNSLTGLPIG 126
           PWVDD+G VQVNRG+RVQF+SA+GPYKGGLRFHP+V  SI+KFLGFEQIFKN+LTG PIG
Sbjct: 76  PWVDDQGNVQVNRGFRVQFNSALGPYKGGLRFHPSVNASIIKFLGFEQIFKNALTGQPIG 135

Query: 127 GGKGGSNFDPKGKSDNEVMRFTQSFMTELQKYIGPDLDVPAGDIGVGGREIGYLYGQYKR 186
           GGKGGS+FDPKGKSD E+MRF+QSFM+EL  YIGPD+DVPAGDIGVG +EIGY++GQYK+
Sbjct: 136 GGKGGSDFDPKGKSDGEIMRFSQSFMSELSNYIGPDIDVPAGDIGVGAKEIGYMFGQYKK 195

Query: 187 L-NGYQNGVLTGKGLTYGGSLARTEATGYGAVYFAKEMLAARGQDLTGKVALVSGSGNVA 245
           +  G++ GVLTGKG+ YGGSLAR EATGYG VYF +EM+   G    G   +VSGSGNV+
Sbjct: 196 MRGGFEAGVLTGKGIGYGGSLARKEATGYGTVYFVEEMIKDHGFSFAGSTVVVSGSGNVS 255

Query: 246 IYATEKLQELGATVVAVSDSSGYVYDPDGIDLETLKQIKEVERARIVKYTEKHPKANFTP 305
           IYA EK  +LGA VVA SDS GYVYD +GIDL+T+K++KEVER RI +Y  +HP A++
Sbjct: 256 IYAMEKAMQLGAKVVACSDSGGYVYDKNGIDLQTVKRLKEVERKRISEYVNEHPHAHYVQ 315

Query: 306 ADQGSIWSIKADLAFPCATQNELDEEDAKLLVENGVLAVTEGANMPSTLGAIKVFQKAGV 365
```

```
              G IWS+  D+A PCATQNELDE  A +L+ NGV AV EGANMPSTL A+  FQ+ GV
Sbjct: 316 GCSG-IWSVPCDIALPCATQNELDEAAAATMLIANGVKAVGEGANMPSTLQAVHTFQEHGV 374

Query: 366 AFGPAKAANAGGVAVSALEMAQNSSRRAWTFEEVDQELQRIMKTIFVNASEAADEFGDSG 425
            F PAKAANAGGV+VSALEMAQNS+R AWTFEEVD +L  IMK I+  + +AA+ +  SG
Sbjct: 375 LFAPAKAANAGGVSVSALEMAQNSTRLAWTFEEVDAKLYEIMKNIYRESIKAAELYEASG 434

Query: 426 NLVLGANIAGFLKVAQAMSAQGIV 449
           NLV+GANIAGF+KVA AM + G+V
Sbjct: 435 NLVVGANIAGFVKVADAMISHGVV 458
```

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 1943

A DNA sequence (GBSx2052) was identified in *S.agalactiae* <SEQ ID 6023> which encodes the amino acid sequence <SEQ ID 6024>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.55   Transmembrane   61 -  77 (  55 -  87)
   INTEGRAL    Likelihood = -7.70   Transmembrane  177 - 193 ( 175 - 202)
   INTEGRAL    Likelihood = -7.06   Transmembrane   99 - 115 (  95 - 122)
   INTEGRAL    Likelihood = -5.89   Transmembrane   42 -  58 (  40 -  60)
   INTEGRAL    Likelihood = -3.08   Transmembrane  160 - 176 ( 159 - 176)
   INTEGRAL    Likelihood = -2.44   Transmembrane  124 - 140 ( 122 - 144)

----- 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 9955> which encodes amino acid sequence <SEQ ID 9956> was also identified.

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

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

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

Example 1944

A DNA sequence (GBSx2053) was identified in *S.agalactiae* <SEQ ID 6025> which encodes the amino acid sequence <SEQ ID 6026>. This protein is predicted to be ABC transporter, ATP-binding protein (msbA). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.72   Transmembrane  152 - 168 ( 147 - 192)
   INTEGRAL    Likelihood = -5.47   Transmembrane  267 - 283 ( 264 - 288)
   INTEGRAL    Likelihood = -4.30   Transmembrane  171 - 187 ( 169 - 192)
   INTEGRAL    Likelihood = -2.13   Transmembrane   67 -  83 (  67 -  83)
   INTEGRAL    Likelihood = -0.32   Transmembrane  493 - 509 ( 493 - 509)

----- Final Results -----
              bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
               bacterial outside --- 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:CAB69752 GB:AL137187 putative ABC transporter [Streptomyces coelicolor A3(2)]
 Identities = 269/611 (44%), Positives = 392/611 (64%), Gaps = 31/611 (5%)

Query: 9   RLWSYLTRYKATLFLAIFLKVLSSFMSILEPFILGLAITELTANLV--DMAKG------- 59
           RL S      +ATLF  +   V+S  ++++ P ILG A   + A +V  DM  G
Sbjct: 27  RLVSQFRPERATLFTLLACVVVSVGLNVVGPKILGRATDLVFAGIVGRDMPSGATKEQVL 86

Query: 60  --------------------VSGAELNVPYIAGILIIYFFRGVFYELGSYGSNYFMTTVV 99
                              V G  ++   +  +L++          L    +   +  V
Sbjct: 87  ATMREHGDGNVADMLRSTDFVPGQGIDFGAVGEVLLLALATFAVAGLLMAVATRLVNRAV 146

Query: 100 QKSIRDIRHDLNRKINKVPVSYFDKHQFGDMLGRFTSDVETVSNALQQSFLQIINAFLSI 159
            +++   +R D+  K++++P+SYFDK Q G++L R T+D++ +   LQQS  Q+IN+ L+I
Sbjct: 147 NRTMFRLREDVQTKLSRLPLSYFDKRQRGEVLSRATNDIDNIGQTLQQSMGQLINSLLTI 206

Query: 160 ILVVVMVLYLNVPLAMIIIACIPVTYFSAQAILKRSQPYFKEQAKILGELNGFVQEKLTG 219
           I V+ M+ Y++  LA++ +  +P+++  A  + KRSQP F +Q +  G+LN  ++E  TG
Sbjct: 207 IGVLAMMFYVSWILALVALVTVPLSFVVATRVGKRSQPQFVQQWRSTGQLNAHIEEMYTG 266

Query: 220 FNIIKLYGREEASSQEFRDITDNLRHVGFKASFISGIMMPVLNSISDFIYLIIAFVGGLQ 279
             ++K++GR+E S+++F +  D L   GFKA F SGIM P++  +S+  Y+++A VGGL+
Sbjct: 267 HALVKVFGRQEESAKQFAEQNDALYEAGFKAQFNSGIMQPLMMCVSNLNYVLVAVVGGLR 326

Query: 280 VIAGTLTIGNMQAFVQYVWQISQPVQTITQLAGVLQSAKSSLERIFEVLD-EEEEANQVT 338
           V +G L+IG++QAF+QY  Q S P+  +  +A ++QS  +S ER+FE+LD EE+ A+ +
Sbjct: 327 VASGQLSIGDVQAFIQYSRQFSMPLTQVASMANLVQSGVASAERVFELLDAEEQSADPIP 386

Query: 339 EKLSHDLTGQVSFHGVDFHYSPDKPLIRDFNLDVEPGQMIAIVGPTGAGKTTLINLLMRF 398
                DL G+V    V F Y P+KPLI D +L VEPG  +AIVGPTGAGKTTL+NLLMRF
Sbjct: 387 GARPEDLRGRVELEHVSFRYDPEKPLIEDLSLKVEPGHTVAIVGPTGAGKTTLVNLLMRF 446

Query: 399 YDVSEGAITVDGHDIRHLSRQDFRQQFGMVLQDAWLYEGTIKENLRFG-NLEASDEDIVA 457
           Y+VS G IT+DG DI  +SR + R   GMVLQD WL+ GTI EN+ +G + E +  +I
Sbjct: 447 YEVSGGRITLDGVDIAKMSRDELRAGIGMVLQDTWLFGGTIAENIAYGASREVTRGEIEE 506

Query: 458 AAKAANVDHFIRTLPGGYNMVMNQESSNISLGQKQLLTIARALLADPKILILDEATSSVD 517
           AA+AA+ D F+RTLP GY+ V++ E + +S G+KQL+TIARA L+DP IL+LDEATSSVD
Sbjct: 507 AARAAHADRFVRTLPDGYDTVIDDEGTGVSAGEKQLITIARAFLSDPVILVLDEATSSVD 566

Query: 518 TRLELLIQKAMKKLMEGRTSFVIAHRLSTIQEADNILVLKDGQIIEQGNHQKLLADKGFY 577
           TR E+LIQKAM KL  GRTSFVIAHRLSTI++AD ILV++DG I+EQG H +LL   G Y
Sbjct: 567 TRTEVLIQKAMAKLAHGRTSFVIAHRLSTIRDADTILVMEDGAIVEQGAHTELLTADGAY 626

Query: 578 YELYNSQFSNS 588
             LY +QF+ +
Sbjct: 627 ARLYKAQFAEA 637
```

There is also homology to SEQ IDs 160 and 6546.

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

Example 1945

A DNA sequence (GBSx2054) was identified in *S.agalactiae* <SEQ ID 6027> which encodes the amino acid sequence <SEQ ID 6028>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.88   Transmembrane  242 - 258 ( 235 - 263)
   INTEGRAL    Likelihood = -9.82   Transmembrane  159 - 175 ( 129 - 177)
   INTEGRAL    Likelihood = -9.71   Transmembrane   52 -  68 (  49 -  77)
```

```
    INTEGRAL    Likelihood = -8.49   Transmembrane  134 - 150 ( 129 - 158)
    INTEGRAL    Likelihood = -1.17   Transmembrane  272 - 288 ( 272 - 289)

----- Final Results -----
               bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB69751 GB:AL137187 putative ABC transporter [Streptomyces
          coelicolor A3(2)]
 Identities = 226/565 (40%), Positives = 342/565 (60%), Gaps = 1/565 (0%)

Query: 6   SYLKRYPNWLWLDLLGAMLFVTVILGMPTALAGMIDNGVTKGDRTGVYLWTFIMFIFVVL 65
           +YL+ Y   + L +    L     L +PT  A +ID GV KGD   +  +  +M    +
Sbjct: 8   TYLRPYKKPIALLVALQFLQTCASLYLPTLNAHIIDEGVVKGDSGYILSYGALMIGISLA 67

Query: 66  GIIGRITMAYASSRLTTTMIRDMRNDMYAKLQEYSHHEYEQIGVSSLVTRMTSDTFVLMQ 125
            ++  I   +  +R    + RD+R  ++ ++Q +S  E    G  SL+TR T+D   +
Sbjct: 68  QVVCNIGAVFYGARTAAALGRDVRGAVFDRVQSFSAREVGHFGAPSLITRTTNDVQQVQM 127

Query: 126 FAEMSLRLGLVTPMVMIFSVVMILITSPSLAWLVAVAMPLLVGVILYVAIKTKPLSERQQ 185
            A M+  L +  P++ +  +VM L     L+ ++   +P+L   +  K +PL  + Q
Sbjct: 128 LALMTFTLMVSAPIMCVGGIVMALGLDVPLSGVLLGVVPVLAICVTLIVRKLRPLFRKMQ 187

Query: 186 TMLDKINQYVRENLTGLRVVRAFARENFQSQKFQVANQRYTDTSTGLFKLTGLTEPLFVQ 245
             LD +N+ +RE +TG RV+RAF R+ ++ Q+F+ AN   T+ + G   L  L  P+ +
Sbjct: 188 VRLDTVNRVLREQITGNRVIRAFVRDEYEQQRFRKANTELTEVALGTGNLLALMFPVVMT 247

Query: 246 IIIAMIVAIVWFALDPLQRGAIKIGDLVAFIEYSFHALFSFLLFANLFTMYPRMVVSSHR 305
           ++    +A+VWF    +  G ++IGDL AF+ Y    + S ++   +F M PR  V + R
Sbjct: 248 VVNLSSIAVVWFGAHRIDSGGMQIGDLTAFLAYLMQIVMSVMMATFMFMMVPRAEVCAER 307

Query: 306 IREVMDMPISINPNTEGVTDTKLKGHLEFDNVTFAYPGETESPVLHDISFKAKPGETIAF 365
           I+EV++   S+ P    VT+ +  GHLE     F YPG  E PVL  I   A+PGET A
Sbjct: 308 IQEVLETESSVVPPVAPVTELRRHGHLEIREAGFRYPG-AEEPVLRHIDLVARPGETTAV 366

Query: 366 IGSTGSGKSSLVNLIPRFYDVTLGKILVDGVDVRDYNLKSLRQKIGFIPQKALLFTGTIG 425
           IGSTGSGKS+L+ L+PR +D T G++LV+GVDVR  + K+L + +  +PQK  LF GT+
Sbjct: 367 IGSTGSGKSTLLGLVPRLFDATDGEVLVNGVDVRTVDPKTLAKVVSLVPQKPYLFAGTVA 426

Query: 426 ENLKYGKADATIDDLRQAVDISQAKEFIESHQEAFETHLAEGGSNLSGGQKQRLSIARAV 485
            NL+YG  DAT ++L  A+ ++QAKEF+   +   +  +A+GG+N+SGGQ+QRL+IAR +
Sbjct: 427 TNLRYGNPDATDEELWHALAVAQAKEFVSELEGGLDAPIAQGGTNVSGGQRQRLAIARTL 486

Query: 486 VKDPDLYIFDDSFSALDYKTDATLRARLKEVTGDSTVLIVAQRVGTIMDADQIIVLDEGE 545
           V+ P++Y+FDDSFSALDY TDA LRA L + T ++TV+IVAQRV TI DAD+I+VLDEG
Sbjct: 487 VQRPEIYLFDDSFSALDYATDAALRAELAQETAEATVVIVAQRVATIRDADRIVVLDEGR 546

Query: 546 IVGRGTHAQLIENNAIYREIAESQL 570
           +VG G H +L+ +N  YREI  SQL
Sbjct: 547 VVGVGRHHELMADNETYREIVLSQL 571
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4985> which encodes the amino acid sequence <SEQ ID 4986>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-16.24   Transmembrane  155 - 171 ( 145 - 176)
   INTEGRAL    Likelihood = -7.48   Transmembrane  130 - 146 ( 122 - 150)
   INTEGRAL    Likelihood = -5.04   Transmembrane   13 -  29 (  12 -  30)
   INTEGRAL    Likelihood = -5.04   Transmembrane   56 -  72 (  52 -  75)
   INTEGRAL    Likelihood = -4.14   Transmembrane  239 - 255 ( 238 - 259)
   INTEGRAL    Likelihood = -1.70   Transmembrane  269 - 285 ( 269 - 288)

----- 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>
```

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

```
 Identities = 175/511 (34%), Positives = 296/511 (57%), Gaps = 3/511 (0%)

Query: 59  MFIFVVLGIIGRITMAYASSRLTTTMIRDMRNDMYAKLQEYSHHEYEQIGVSSLVTRMTS 118
           + I  +LG++         ++++   +  DMR   + K+Q++S+   E      +LV R+T+
Sbjct: 56  LLIIALLGLMSGAINTVLAAKIAQGVSADMREKTFRKIQDFSYANIEAFNAGNLVVRLTN 115

Query: 119 DTFVLMQFAEMSLRLGLVTPMVMIFSVVMILITSPSLAWLVAVAMPLLVGVILYVAIKTK 178
           D   +     M  ++     P++ I + +M + T P L W++ V + L+   ++  V  +
Sbjct: 116 DINQIQSLVMMMFQILFRLPILFIGAFIMAVQTFPQLWWVIVVMVILIALIMGLVMRQMG 175

Query: 179 PLSERQQTMLDKINQYVRENLTGLRVVRAFARENFQSQKFQVANQRYTDTSTGLFKLTGL 238
           P   + Q ++DKIN+  +ENL G+RVV++F +E  Q  KF+  +      +  +     L
Sbjct: 176 PRFGKFQRLMDKINRIAKENLRGVRVVKSFVQEQQQYTKFKETSNDLLALNLSIGYGFSL 235

Query: 239 TEPLFVQIIIAMIVAIVWFALDPLQRGAIKIGDLVAFIEYSFHALFSFLLFANLFTMYPR 298
            +P  + +    +   +      ++     IG++ +F+ Y    +FS ++  ++     R
Sbjct: 236 MQPALMLVSYLAVYVSINVVSTMVETDPTVIGNIASFMTYMMQIMFSIIVVGSMGMQVSR 295

Query: 299 MVVSSHRIREVMDMPISINPNTEGVTDTKLKGHLEFDNVTFAYPGETESPVLHDISFKAK 358
             VS  RIR+++    ++    E   +  + G + FD+V+F YP + E P L  ISF  +
Sbjct: 296 AFVSMARIRQILSTEPAMTFENE--KEETISGSIVFDDVSFTYPNDDE-PTLKHISFAIE 352

Query: 359 PGETIAFIGSTGSGKSSLVNLIPRFYDVTLGKILVDGVDVRDYNLKSLRQKIGFIPQKAL 418
           PG+ +  +G+TGSGKS+L  LIPR +D   G+IL+ G  ++  +  +LRQ +  + QKA+
Sbjct: 353 PGQMVGIVGATGSGKSTLAQLIPRLFDPQDGQILLGGKPIKTLSQTTLRQSVSIVLQKAI 412

Query: 419 LFTGTIGENLKYGKADATIDDLRQAVDISQAKEFIESHQEAFETHLAEGGSNLSGGQKQR 478
           LF+GTI +NL+ G A A ID +++A  I+QAKEFI+      +E+ + E GSNLSGGQKQR
Sbjct: 413 LFSGTIADNLRQGSAKADIDAMQKAAQIAQAKEFIDRMDSRYESQVEERGSNLSGGQKQR 472

Query: 479 LSIARAVVKDPDLYIFDDSFSALDYKTDATLRARLKEVTGDSTVLIVAQRVGTIMDADQI 538
           LSIAR V+  P + I DDS SALD K++  ++  L      +T +IVAQ++ +++ AD+I
Sbjct: 473 LSIARGVINHPKILILDDSTSALDAKSEKRVQEALSHKLEGTTTVIVAQKISSVVKADKI 532

Query: 539 IVLDEGEIVGRGTHAQLIENNAIYREIAESQ 569
           +VLD+G+++G GTHA+L+ NNAIYREI E+Q
Sbjct: 533 LVLDQGQLIGEGTHAELVANNAIYREIYETQ 563
```

There is also homology to SEQ IDs 72 and 6552.

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

Example 1946

A DNA sequence (GBSx2055) was identified in *S.agalactiae* <SEQ ID 6029> which encodes the amino acid sequence <SEQ ID 6030>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2391(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:CAA51784 GB:X73368 ORF 18.3 [Salmonella typhimurium]
```

```
 Identities = 58/162 (35%), Positives = 92/162 (55%), Gaps = 8/162 (4%)

Query: 1   MIIRPIIKNDDQAVAQLIRQSLRAYDL--DKPDTAYSDPHLDHLTSYYEKIEKSGFFVIE 58
           + +R I  D+ A+A++IRQ    Y L  DK  T  +DP+LD L  Y +   + ++V+E
Sbjct: 9   LTVRRITTADNAAIARVIRQVSAEYGLTADKGYTV-ADPNLDELYQVYSQ-PGAAYWVVE 66

Query: 59  ERDEIIGCGGFGPLKNL---IAEMQKVYIAERFRGKGLATDLVKMIEVEARKIGYRQLYL 115
           +   ++G GG  PL      I E+QK+Y     RG+GLA  L  M    AR+ G+++ YL
Sbjct: 67  QNGCVVGGGGVAPLSCSEPDICELQKMYFLPVIRGQGLAKKLALMALDHAREQGFKRCYL 126

Query: 116 ETASTLSRATAVYKHMGYCALSQPIANDQGHTAMDIWMIKDL 157
           ET + L  A A+Y+ +G+  +S+P+    GH   ++ M+KDL
Sbjct: 127 ETTAFLREAIALYERLGFEHISEPL-GCTGHVDCEVRMLKDL 167
```

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 1947

A DNA sequence (GBSx2056) was identified in *S.agalactiae* <SEQ ID 6031> which encodes the amino acid sequence <SEQ ID 6032>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1738(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:CAB12566 GB:Z99108 similar to ABC transporter (ATP-binding
          protein) [Bacillus subtilis]
 Identities = 269/625 (43%), Positives = 397/625 (63%), Gaps = 11/625 (1%)

Query: 1   MSDFLVDGLTKSVGDKTVFSNVSFIIHSLDRIGIIGVNGTGKTTLLDVISGELGFDGDRS 60
           MS    + L K+ GDKT+F ++SF I   +RIG+IG NGTGK+TLL VI+G    +
Sbjct: 1   MSILKAENLYKTYGDKTLFDHISFHIEENERIGLIGPNGTGKSTLLKVIAGLESIE--EG 58

Query: 61  PFSSANDYKIAYLKQEPDFDDSQTILDTVLSSDLREMALIKEYELLLNHY-----EESKQ 115
             + +   ++ +L Q+P+    QT+L+ + S +   M  ++EYE  L         E +Q
Sbjct: 59  EITKSGSVQVEFLHQDPELPAGQTVLEHIYSGESAVMKTLREYEKALYELGKDPENEQRQ 118

Query: 116 SRLEKVMAEMDSLDAWSIESEVKTVLSKLGITDLQLSVGELSGGLRRRVQLAQVLLNDAD 175
             L    A+MD+ +AW   +  KTVLSKLG+ D+   V ELSGG ++RV +A+ L+  AD
Sbjct: 119 KHLLAAQAKMDANNAWDANTLAKTVLSKLGVNDVTKPVNELSGGQKKRVAIAKNLIQPAD 178

Query: 176 LLLLDEPTNHLDIDTIAWLTNFLKNSKKTVLFITHDRYFLDNVATRIFELDKAQITEYQG 235
           LL+LDEPTNHLD +TI WL  +L       V+ +THDRYFL+ V  RI+EL++  + Y+G
Sbjct: 179 LLILDEPTNHLDNETIEWLEGYLSQYPGAVMLVTHDRYFLNRVTNRIYELERGSLYTYKG 238

Query: 236 NYQDYVRLRAEQDERDAASLHKKKQLYKQELAWMRTQPQARATKQQARINRFQNLKNDLH 295
           NY+ ++  RAE++ +       K++ L ++ELAW+R   +AR+TKQ+ARI+R + LK
Sbjct: 239 NYEVFLEKRAEREAQAEQKETKRQNLLRRELAWLRRGAKARSTKQKARIDRVETLKEQTG 298

Query: 296 QTSDTSDLEMTFETSRIGKKVINFENVSFSYPDKSILKDFNLLIQNKDRIGIVGDNGVGK 355
             S  S L+     + R+GK+VI  ENV  +Y  + ++  FN L+   +RIGI+G NG+GK
Sbjct: 299 PQSSGS-LDFAIGSHRLGKQVIEAENVMIAYDGRMLVDRFNELVIPGERIGIIGPNGIGK 357

Query: 356 STLLNLIVQDLQPDSGNVSIGETIRVGYFSQQLHNMDGSKRVINYLQEVADEVKTSVGTT 415
           +TLLN +      PD G+++IG+T+R+GY++Q    M+G  +VI+Y++E A+ VKT+ G
```

```
Sbjct: 358 TTLLNALAGRHTPDGGDITIGQTVRIGYYTQDHSEMNGELKVIDYIKETAEVVKTADGDM 417

Query: 416 SVTE-LLEQFLFPRSTHGTQIAKLSGGEKKRLYLLKILIEKPNVLLLDEPTNDLDIATLT 474
             E +LE+FLFPRS   T I KLSGGEK+RLYLL++L+++PNVL LDEPTNDLD  TL+
Sbjct: 418 ITAEQMLERFLFPRSMQQTYIRKLSGGEKRRLYLLQVLMQEPNVLFLDEPTNDLDTETLS 477

Query: 475 VLENFLQGFGGPVITVSHDRYFLDKVANKIIAFEDND-IREFFGNYTDYLDEKAFNEQNN 533
           VLE+++  F G VITVSHDRYFLD+V ++++I FE N  I  F G+Y+DY++E    +
Sbjct: 478 VLEDYIDQFPGVVITVSHDRYFLDRVVDRLIVFEGNGVISRFQGSYSDYMEESKAKKAAP 537

Query: 534 EVISKKESTKTSREKQSRKRMSYFEKQEWATIEDDIMILENTITRIENDMQTCGSDFTRL 593
           +  + +E T  +  K+ RK++SY ++ EW  IED I  LE    ++E D+   GSDF ++
Sbjct: 538 KP-AAEEKTAEAEPKKKRKKLSYKDQLEWDGIEDKIAQLEEKHEQLEADIAAAGSDFGKI 596

Query: 594 SDLQKELDAKNEALLEKYDRYEYLS 618
            +L  E     E L    DR+  LS
Sbjct: 597 QELMAEQAKTAEELEAAMDRWTELS 621
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6033> which encodes the amino acid sequence <SEQ ID 6034>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2591(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 = 467/624 (74%), Positives = 535/624 (84%), Gaps = 3/624 (0%)

Query: 1   MSDFLVDGLTKSVGDKTVFSNVSFIIHSLDRIGIIGVNGTGKTTLLDVISGELGFDGDRS 60
           MS FLV+ LTK+VGDKTVF ++SFIIH  DRIGIIGVNGTGKTTLLDV+SG LGFDGD S
Sbjct: 1   MSHFLVEKLTKTVGDKTVFQDISFIIHDFDRIGIIGVNGTGKTTLLDVLSGRLGFDGDHS 60

Query: 61  PFSSANDYKIAYLKQEPDFDDSQTILDTVLSSDLREMALIKEYELLLNHYEESKQSRLEK 120
           PFS ANDYKIAYL Q+P+F+D+ ++LDTVLS+D++ + LI++YELL+ +Y E KQ  LE
Sbjct: 61  PFSKANDYKIAYLTQDPEFNDAASVLDTVLSADVKAIQLIRQYELLMANYTEDKQESLES 120

Query: 121 VMAEMDSLDAWSIESEVKTVLSKLGITDLQLSVGELSGGLRRRVQLAQVLLNDADLLLLD 180
           +M+EMD LDAWSIES+VKTVLSKLGITDL+  VG+LSGG+RRRVQLAQVLL  ADLLLLD
Sbjct: 121 LMSEMDRLDAWSIESDVKTVLSKLGITDLEQKVGDLSGGMRRRVQLAQVLLGAADLLLLD 180

Query: 181 EPTNHLDIDTIAWLTNFLKNSKKTVLFITHDRYFLDNVATRIFELDKAQITEYQGNYQDY 240
           EPTNHLDIDTIAWLT +LK +KKTVLFITHDRYFLD+VATRIFELDKA +TEYQGNYQDY
Sbjct: 181 EPTNHLDIDTIAWLTTYLKTAKKTVLFITHDRYFLDHVATRIFELDKAGLTEYQGNYQDY 240

Query: 241 VRLRAEQDERDAASLHKKKQLYKQELAWMRTQPQARATKQQARINRFQNLKNDLHQTSDT 300
           VRL+AEQDERDAA+LHKKKQLYKQELAWMRTQPQARATKQQARINRF +LK ++HQ S
Sbjct: 241 VRLKAEQDERDAANLHKKKQLYKQELAWMRTQPQARATKQQARINRFSDLKKEVHQDSSA 300

Query: 301 SDLEMTFETSRIGKKVINFENVSFSYPDKSILKDFNLLIQNKDRIGIVGDNGVGKSTLLN 360
             LEMTFETSRIGKKVI+FE++SF+Y D+ ++KDFNL+IQNKDRIGIVGDNGVGKSTLLN
Sbjct: 301 DKLEMTFETSRIGKKVIHFEDLSFAYGDRQLIKDFNLIIQNKDRIGIVGDNGVGKSTLLN 360

Query: 361 LIVQDLQPDSGNVSIGETIRVGYFSQQLHNMDGSKRVINYLQEVADEVKTSVGTTSVTEL 420
           +I  DL+P SG + IG+TIRVGYFSQQL +MD +KRVINYLQEVADEVKTSVGTTS++EL
Sbjct: 361 IINGDLKPTSGKLDIGDTIRVGYFSQQLKDMDETKRVINYLQEVADEVKTSVGTTSISEL 420

Query: 421 LEQFLFPRSTHGTQIAKLSGGEKKRLYLLKILIEKPNVLLLDEPTNDLDIATLTVLENFL 480
           LEQFLFPRS+HGT IAKLSGGEKKRLYLLK+LIEKPNVLLLDEPTNDLDIATL VLENFL
Sbjct: 421 LEQFLFPRSSHGTLIAKLSGGEKKRLYLLKLLIEKPNVLLLDEPTNDLDIATLKVLENFL 480

Query: 481 QGFGGPVITVSHDRYFLDKVANKIIAFEDNDIREFFGNYTDYLDEKAFNEQNNEVISKKE 540
             F GPVITVSHDRYFLDKVA KI+AFE+ DIR F+GNY+DYLDEK F ++  E    K
```

```
Sbjct: 481 ANFAGPVITVSHDRYFLDKVATKILAFEEGDIRVFYGNYSDYLDEKVFEKETVEADLAKT 540

Query: 541 STKTS---REKQSRKRMSYFEKQEWATIEDDIMILENTITRIENDMQTCGSDFTRLSDLQ 597
           +        +K+ RKRMSY EKQEWA IED I  +E  I  IEN M T  SD+ +L+ LQ
Sbjct: 541 TVTEEVPLPQKEERKRMSYLEKQEWAQIEDKIATIEANIEEIENQMLTVVSDYGQLAQLQ 600

Query: 598 KELDAKNEALLEKYDRYEYLSELD 621
           KELD +N  LL  Y+R+EYLS LD
Sbjct: 601 KELDQRNNDLLLAYERFEYLSGLD 624
```

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

Example 1948

A DNA sequence (GBSx2057) was identified in *S.agalactiae* <SEQ ID 6035> which encodes the amino acid sequence <SEQ ID 6036>. This protein is predicted to be poly(a) polymerase (papS). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2658(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 9957> which encodes amino acid sequence <SEQ ID 9958> was also identified.

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

```
>GP:AAB38446 GB:L47709 poly(A) polymerase [Bacillus subtilis]
 Identities = 157/395 (39%), Positives = 235/395 (58%), Gaps = 14/395 (3%)

Query: 11  FQKALPILKKIKKAGYEAYFVGGSVRDVLLDRPIHDVDIATSSYPEETKQIFKRTVDVGI 70
           F KALP+L+ + +AG++AYFVGG+VRD  + R I DVDIAT + P++ +++F+RTVDVG
Sbjct: 5   FIKALPVLRILIEAGHQAYFVGGAVRDSYMKRTIGDVDIATDAAPDQVERLFQRTVDVGK 64

Query: 71  EHGTVLVLEKGGEYEITTFRTEEVYVDYRRPSQVNFVRSLEEDLKRRDFTVNAFALNEDG 130
           EHGT++VL +   YE+TTFRTE  YVD+RRPS+V F+ SLEEDLKRRD T+NA A+  DG
Sbjct: 65  EHGTIIVLWEDETYEVTTFRTESDYVDFRRPSEVQFISSLEEDLKRRDLTINAMAMTADG 124

Query: 131 EVIDLFHGLDDLDNHLLRAVGLASERFNEDALRIMRGLRFSASLNFDIETTTFEAMKKHA 190
           +V+D F G  D+D  ++R VG   +RF EDALR++R +RF + L F +   T EA+ K
Sbjct: 125 KVLDYFGGKKDIDQKVIRTVGKPEDRFQEDALRMLRAVRFMSQLGFTLSPETEEAIAKEK 184

Query: 191 SLLEKISVERSFIEFDKLLLAPYWRKGMLALIDSHAFNYLPCLKNRELQLSAFLSQLDKD 250
           SLL  +SVER  IEF+KLL     R+ +  LI +  + LP   ++   L   +S  +
Sbjct: 185 SLLSHVSVERKTIEFEKLLQGRASRQALQTLIQTRLYEELPGFYHKRENL---ISTSEFP 241

Query: 251 FLFETS-EQAWASLILSMEV--EHTKTFLKKWKTSTHFQKDVEHIVDVYRIREQMGLTKE 307
           F   TS E+ WA+L++++ +  +     FLK WK       K+  HI D +       L
Sbjct: 242 FFSLTSREELWAALLINLGIVLKDAPLFLKAWKLPGKVIKEAIHIADTF----GQSLDAM 297

Query: 308 HLYRYGKTIIKQAEGIRKAR-GLMVDFEKIEQLD---SELAIHDRHEIVVNGGTLIKKLG 363
            +YR GK  +  A  I + R    +D +K++ +       L I    ++ + G  L+
Sbjct: 298 TMYRAGKKALLSAAKISQLRQNEKLDEKKLKDIQYAYQNLPIKSLKDLDITGKDLLALRN 357

Query: 364 IKPGPQMGDIISQIELAIVLGQLINEEEAILHFVK 398
              G  + +  IE A+V G+L N+++ I  ++K
Sbjct: 358 RPAGKWVSEELQWIEQAVVTGKLSNQKKHIEEWLK 392
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6037> which encodes the amino acid sequence <SEQ ID 6038>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2023(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 = 256/400 (64%), Positives = 312/400 (78%)

Query: 2   MRLNYLPSEFQKALPILKKIKKAGYEAYFVGGSVRDVLLDRPIHDVDIATSSYPEETKQI 61
           M+L  +PSEFQKALPIL KIK+AGYEAYFVGGSVRDVLL+RPIHDVDIATSSYPEETK I
Sbjct: 1   MKLMTMPSEFQKALPILTKIKEAGYEAYFVGGSVRDVLLERPIHDVDIATSSYPEETKAI 60

Query: 62  FKRTVDVGIEHGTVLVLEKGGEYEITTFRTEEVYVDYRRPSQVNFVRSLEEDLKRRDFTV 121
           F RTVDVGIEHGTVLVLE GGEYEITTFRTE++YVDYRRPSQV+FVRSLEEDLKRRDFTV
Sbjct: 61  FNRTVDVGIEHGTVLVLENGGEYEITTFRTEDIYVDYRRPSQVSFVRSLEEDLKRRDFTV 120

Query: 122 NAFALNEDGEVIDLFHGLDDLDNHLLRAVGLASERFNEDALRIMRGLRFSASLNFDIETT 181
           NA AL+E+G+VID F GL DL     LRAVG A ERF EDALRIMRG RF+ASL+FDIE
Sbjct: 121 NALALDENGQVIDKFRGLIDLKQKRLRAVGKAEERFEEDALRIMRGFRFAASLDFDIEAI 180

Query: 182 TFEAMKKHASLLEKISVERSFIEFDKLLLAPYWRKGMLALIDSHAFNYLPCLKNRELQLS 241
           TFEAM+ H+ LLEKISVERSF EFDKLL+AP+WRKG+ A+I   A++YLP LK +E  L+
Sbjct: 181 TFEAMRSHSPLLEKISVERSFTEFDKLLMAPHWRKGISAMIACQAYDYLPGLKQQEAGLN 240

Query: 242 AFLSQLDKDFLFETSEQAWASLILSMEVEHTKTFLKKWKTSTHFQKDVEHIVDVYRIREQ 301
             +  L  +F F     QAWA +++S+ +E  K+FLK WKTS  FQ+ V  ++ +YRIR++
Sbjct: 241 HLIVSLKDNFTFSDYHQAWAYVMISLAIEDPKSFLKAWKTSNDFQRYVTKLIALYRIRQE 300

Query: 302 MGLTKEHLYRYGKTIIKQAEGIRKARGLMVDFEKIEQLDSELAIHDRHEIVVNGGTLIKK 361
               K  +Y+YGK +    E +RKA+ L VD ++I  LD  L IHD+H+IV+NG  LIK
Sbjct: 301 RSFEKLDIYQYGKKMASLVEDLRKAQSLSVDMDRINTLDQALVIHDKHDIVLNGSHLIKD 360

Query: 362 LGIKPGPQMGDIISQIELAIVLGQLINEEEAILHFVKQYL 401
            G+K GPQ+G ++ ++ELAIV G+L N+   I  FV++ L
Sbjct: 361 FGMKSGPQLGLMLEKVELAIVEGRLDNDFTTIEAFVREEL 400
```

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

Example 1949

A DNA sequence (GBSx2058) was identified in *S.agalactiae* <SEQ ID 6039> which encodes the amino acid sequence <SEQ ID 6040>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2939(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:BAB07346 GB:AP001519 unknown conserved protein [Bacillus halodurans]
 Identities = 94/274 (34%), Positives = 153/274 (55%), Gaps = 2/274 (0%)

Query: 2   KLALITDTSAYLPEAIENHEDVYVLDIPIIIDGKTYIEGQNLTLDQYYDKLAASKELPKT 61
```

```
            K+A++TD++AYL          V V+ + ++    + Y E    L+    +Y+KL    ++LP T
Sbjct: 3    KIAIVTDSTAYLGPKRAKELGVIVVPLSVVFGEEAYQEEVELSSADFYEKLKHEEKLPTT 62

Query: 62   SQPSLAELDDLLCQLEKEGYTHVLGLFIAAGISGFWQNIQFLIEEHPNLTIAFPDTKITS 121
            SQP++     +    +L KEG+  V+ + +++ ISG +Q+            + +    D+ I+
Sbjct: 63   SQPAVGLFVETFERLAKEGFEVVISIHLSSKISGTYQSALTAGSMVEGIEVIGYDSGISC 122

Query: 122  APQGNLVRNALMCSREGMDFDVIVNKIQSQIEKIEGFIVVNDLNHLVKGGRLSNGSAIIG 181
             PQ N V  A     +EG D    I++ +     ++      VV+DL+HL +GGRL+     ++G
Sbjct: 123  EPQANFVAEAAKLVKEGADPQTIIDHLDEVKKRTNALFVVHDLSHLHRGGRLNAAQLVVG 182

Query: 182  NLLSIKPVLHFNEEGKIVVYEKVRTEKKALKRLAEI-VKEMTADGEYDIAIIHSRAQDKA 240
            +LL IKP+LHF E+G IV  EKVRTEKKA  R+ E+  +E ++         +IH+    D A
Sbjct: 183  SLLKIKPILHF-EDGSIVPLEKVRTEKKAWARVKELFAEEASSASSVKATVIHANRLDGA 241

Query: 241  EQLYNLLAKAGLKDDLEIVSFGGVIATHLGEGAV 274
            E+L + +         D+ I   FG VI THLGEG++
Sbjct: 242  EKLADEIRSQFSHVDVSISHFGPVIGTHLGEGSI 275
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6041> which encodes the amino acid sequence <SEQ ID 6042>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3379(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 181/281 (64%), Positives = 233/281 (82%)

Query: 1    MKLALITDTSAYLPEAIENHEDVYVLDIPIIIDGKTYIEGQNLTLDQYYDKLAASKELPK 60
            MKLA+ITD++A LP   ++  + ++ LDIP+IID +TY EG+NL++D +Y K+A S+ LPK
Sbjct: 1    MKLAVITDSTATLPTDLKQDKAIFSLDIPVIIDDETYFEGRNLSIDDFYQKMADSQNLPK 60

Query: 61   TSQPSLAELDDLLCQLEKEGYTHVLGLFIAAGISGFWQNIQFLIEEHPNLTIAFPDTKIT 120
            TSQPSL+ELD+LL  L  +GYTHV+GLF+A GISGFWQNIQFL EEHP + +AFPD+KIT
Sbjct: 61   TSQPSLSELDNLLGLLSSKGYTHVIGLFLAGGISGFWQNIQFLAEEHPEIEMAFPDSKIT 120

Query: 121  SAPQGNLVRNALMCSREGMDFDVIVNKIQSQIEKIEGFIVVNDLNHLVKGGRLSNGSAII 180
            SAP G++V+N L   SR+GM F   I+NK+Q QI+     FI+V+DLNHLVKGGRLSNGSA++
Sbjct: 121  SAPLGSMVKNVLDWSRQGMTFQAILNKLQEQIDGTTAFIMVDDLNHLVKGGRLSNGSALL 180

Query: 181  GNLLSIKPVLHFNEEGKIVVYEKVRTEKKALKRLAEIVKEMTADGEYDIAIIHSRAQDKA 240
            GNLLSIKP+L F+EEGKIVVYEKVRTEKKA+KRL EI+ ++ ADG+Y++  IIHS+AQDKA
Sbjct: 181  GNLLSIKPILRFDEEGKIVVYEKVRTEKKAMKRLVEILNDLIADGQYNVFIIHSKAQDKA 240

Query: 241  EQLYNLLAKAGLKDDLEIVSFGGVIATHLGEGAVAFGITPK 281
            + L  LL  +G + D+E V FG VIATHLGEGA+AFG+TP+
Sbjct: 241  DYLKRLLQDSGYQYDIEEVHFGAVIATHLGEGAIAFGVTPR 281
```

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

Example 1950

A DNA sequence (GBSx2059) was identified in *S.agalactiae* <SEQ ID 6043> which encodes the amino acid sequence <SEQ ID 6044>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.59   Transmembrane   51 -  67 (  50 -  67)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.1638(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6045> which encodes the amino acid sequence <SEQ ID 6046>. Analysis of this protein sequence reveals the following:

```
      Possible site: 61

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.19   Transmembrane   50 -  66 (  49 -  67)

----- Final Results -----
              bacterial membrane --- Certainty=0.2275(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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 = 94/126 (74%), Positives = 115/126 (90%)

Query: 1   MEVIREQEFVNQYHYDARNLEWEEENGTPKTNFEVTFQLANRDEAAKVTSIVAVLQFVIV 60
           M+++RE+EFVNQYHYDARNLEWE+ENGTP+TNFEVTFQL ++DE  K T IV+VLQFVIV
Sbjct: 1   MQLVREKEFVNQYHYDARNLEWEKENGTPETNFEVTFQLIDKDEQQKETVIVSVLQFVIV 60

Query: 61  RDEFVISGVISQMAHIQGRLINEPSEFSQDEVENLAAPLLEIVKRLTYEVTEIALDRPGV 120
           ++EFVISGVISQM  I  RL+++PSEF+Q+EVE+LAAPLL++VKRLTYEVTEIALDRPG+
Sbjct: 61  KEEFVISGVISQMVRILDRLVDKPSEFTQEEVESLAAPLLDMVKRLTYEVTEIALDRPGI 120

Query: 121 TLEFNS 126
            LEF +
Sbjct: 121 HLEFKN 126
```

SEQ ID 6044 (GBS416) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 4; MW 17.5kDa).

GBS416-His was purified as shown in Figure 214, lane 7.

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

Example 1951

A DNA sequence (GBSx2060) was identified in *S.agalactiae* <SEQ ID 6047> which encodes the amino acid sequence <SEQ ID 6048>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3875(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 1952

A DNA sequence (GBSx2061) was identified in *S.agalactiae* <SEQ ID 6049> which encodes the amino acid sequence <SEQ ID 6050>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> 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 1953

A DNA sequence (GBSx2062) was identified in *S.agalactiae* <SEQ ID 6051> which encodes the amino acid sequence <SEQ ID 6052>. This protein is predicted to be PTS system, fructose-specific enzyme II, BC component (fruA-1). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.56   Transmembrane  630 - 646 ( 618 - 653)
   INTEGRAL    Likelihood = -7.43   Transmembrane  307 - 323 ( 303 - 331)
   INTEGRAL    Likelihood = -7.01   Transmembrane  415 - 431 ( 412 - 435)
   INTEGRAL    Likelihood = -7.01   Transmembrane  448 - 464 ( 444 - 474)
   INTEGRAL    Likelihood = -3.72   Transmembrane  595 - 611 ( 591 - 612)
   INTEGRAL    Likelihood = -3.61   Transmembrane  530 - 546 ( 529 - 553)
   INTEGRAL    Likelihood = -2.39   Transmembrane  350 - 366 ( 350 - 371)
   INTEGRAL    Likelihood = -1.70   Transmembrane  486 - 502 ( 486 - 506)
   INTEGRAL    Likelihood = -1.49   Transmembrane  376 - 392 ( 376 - 392)

----- Final Results -----
               bacterial membrane --- Certainty=0.5225(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 9959> which encodes amino acid sequence <SEQ ID 9960> was also identified.

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

```
>GP:BAB04547 GB:AP001510 PTS system, fructose-specific enzyme II, BC
          component [Bacillus halodurans]
 Identities = 320/659 (48%), Positives = 438/659 (65%), Gaps = 46/659 (6%)

Query: 1   MKIQDLLKKEVMIMDLKATSKEAAIDEMITKLVDTGVVTNFAIFKDGIMKREAQTSTGLG 60
           +KI +LLKK+ M+++L+A SKEA IDE++  L   G + +   FK  I++RE+Q++TG+G
Sbjct: 2   LKISELLKKDTMVLNLRAASKEAVIDELVRTLDKAGRLNDAQAFKRAILERESQSTTGVG 61

Query: 61  DGIAMPHSKNAAVKEATVLFAKSASGVDYEALDGQPTDLFFMIAAPDGANDTHLAALAEL 120
           +GIA+PH+K AAVK+  + F +S +G+DYE+LDGQP+ LFFMIAA +GAN+ HL  L+ L
Sbjct: 62  EGIAIPHAKTAAVKQPAIAFGRSDAGIDYESLDGQPSHLFFMIAASEGANNEHLETLSRL 121
```

```
Query: 121 SKYLLKEGFADQLRQAKTPDDIIATFDSNSISQETVAPQTVQSTSKGSDYIVAVTACTTG 180
           S +L+ E F   L +A++ D+I+A  D     +E       +   +G + ++AVT C TG
Sbjct: 122 STFLMDETFRSTLMKAQSEDEILAAID----KKEAETAGEAEEKQEGYE-LLAVTGCPTG 176

Query: 181 IAHTYMAEEALKKKAAEMGVGIKVETNGASGVGNKLTSSDIARAKGVIIAADKAVEMDRF 240
           IAHTYMA + LK KA E+GV IKVETNG+ GV N+LT  +I+ AK +I+AAD  VEMDRF
Sbjct: 177 IAHTYMAADNLKSKAQELGVSIKVETNGSGGVKNRLTDEEISAAKAIIVAADTKVEMDRF 236

Query: 241 DGKPLVSRPVADGIKKSEDLINIILDNKAQTYHAKNQNDKQSGESDGKSGLGS---AFYK 297
            GKP++  PV DGI++ ++LI+  L  KA  Y    +  Q+   DG +G G      FYK
Sbjct: 237 HGKPVIQVPVTDGIRRPKELIDQALAGKAPVY----EGGAQASGEDGSAGGGRPKLGFYK 292

Query: 298 HLMGGVSQMLPFVIGGGIMIAIAFLFDNILGVPKDQLSNLGSYHEIAALFKNIGGA-AFA 356
           HLM GVS MLPFV+GGGI+IAI+F+F      P D      SYH  A +   IGG  AF
Sbjct: 293 HLMNGVSNMLPFVVGGGILIAISFMFGIKAFDPSDP-----SYHPFAEMLMTIGGGNAFG 347

Query: 357 FMLPVLAGYIAYSIAEKPGLVAGFVAGSIASSGLAFGKVPFAEGGKATLALAGVPSGFLG 416
            M+PVLA +IA SIA++PG  AG + G IAS+G A                    GFLG
Sbjct: 348 LMIPVLAAFIAMSIADRPGFAAGMIGGLIASTGEA--------------------GFLG 386

Query: 417 ALVGGFLAGGVILLLRKLLSGLPKSLEGIKSILLYPLLGVLITGFLMLLVNIPMAAINTA 476
           L+ GFLAG V L ++K+L+ LP++L+GIK+IL YP+  + ITG +ML++  P+AA NT
Sbjct: 387 GLIAGFLAGYVALGVKKVLANLPQTLDGIKTILFYPVFNIFITGMIMLVIVGPLAAFNTG 446

Query: 477 LNTFLQGLSGSSAVLMGLLVGGMMAVDMGGPVNKAAYVFGTGTLAATVANGGSVVMAAVM 536
           L  +L  +  ++ V++G+++GGMMAVDMGGP+NKAA+ FG   + A    G     AAVM
Sbjct: 447 LQDWLGSMGTANMVILGVILGGMMAVDMGGPINKAAFTFGIAMIDA----GNFGPHAAVM 502

Query: 537 AGGMVPPLAVFVATLLFKDKFNNEERQSGLTNIVMGLSFITEGAIPFGAADPARAIPSFI 596
           AGGMVPPL + +AT LFK KF  +ER++G TN ++G SFITEGAIPF AADP R IPS I
Sbjct: 503 AGGMVPPLGIALATTLFKKKFTKQEREAGKTNYILGASFITEGAIPFAAADPGRVIPSII 562

Query: 597 VGSALTGALVGLAGIKLMAPHGGIFVI---ALTSNPLLYILFILIGAVVSGVLFGLFRK 652
           VGSA  G L  L  + L APHGG FVI     + +NPLLY++ I+ G++V+ +L G ++K
Sbjct: 563 VGSAFAGGLTALFNVTLSAPHGGAFVIFIGNIVNNPLLYLVAIIAGSIVTALLLGFWKK 621
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6053> which encodes the amino acid sequence <SEQ ID 6054>. Analysis of this protein sequence reveals the following:

```
      Possible site: 18
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.77   Transmembrane  624 - 640 ( 612 - 646)
   INTEGRAL    Likelihood = -7.59   Transmembrane  301 - 317 ( 297 - 321)
   INTEGRAL    Likelihood = -6.85   Transmembrane  442 - 458 ( 439 - 468)
   INTEGRAL    Likelihood = -5.95   Transmembrane  409 - 425 ( 406 - 426)
   INTEGRAL    Likelihood = -3.61   Transmembrane  524 - 540 ( 523 - 547)
   INTEGRAL    Likelihood = -2.50   Transmembrane  337 - 353 ( 337 - 353)
   INTEGRAL    Likelihood = -2.44   Transmembrane  589 - 605 ( 589 - 605)
   INTEGRAL    Likelihood = -1.70   Transmembrane  480 - 496 ( 480 - 500)
   INTEGRAL    Likelihood = -1.44   Transmembrane  370 - 386 ( 370 - 386)

----- Final Results -----
             bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:BAB04547 GB:AP001510 PTS system, fructose-specific enzyme II, BC
          component [Bacillus halodurans]
 Identities = 322/659 (48%), Positives = 431/659 (64%), Gaps = 48/659 (7%)

Query: 1   MKIQDLLRKDIMILDLQAISKEVAIDEMITKLVEKDIVHDFDVFKKSIMTREEQTSTGLG 60
           +KI +LL+KD M+L+L+A SKE  IDE++  L +   ++D   FK++I+ RE Q++TG+G
Sbjct: 2   LKISELLKKDTMVLNLRAASKEAVIDELVRTLDKAGRLNDAQAFKRAILERESQSTTGVG 61
```

```
Query: 61  DGIAMPHSKNIVVDKPAVLFAKSNKGVDYKALDGQPTDLFFMIAAPQGANDTHLAALAEL 120
           +GIA+PH+K   V +PA+ F +S+ G+DY++LDGQP+ LFFMIAA +GAN+ HL  L+ L
Sbjct: 62  EGIAIPHAKTAAVKQPAIAFGRSDAGIDYESLDGQPSHLFFMIAASEGANNEHLETLSRL 121

Query: 121 SQYLLKDGFADKLRAAATPEAVIAVFD--EASTAKEEVVAPTSGQDFIVAVTACPTGIAH 178
           S +L+ + F   L  A + + ++A  D  EA TA E        +  ++AVT CPTGIAH
Sbjct: 122 STFLMDETFRSTLMKAQSEDEILAAIDKKEAETAGEAEEKQEGYE--LLAVTGCPTGIAH 179

Query: 179 TYMAEEALKKQAAEMGVAIKVETNGASGVANRLTAEDIQRAKGVIVAADKAVEMDRFDGK 238
           TYMA + LK +A E+GV+IKVETNG+ GV NRLT E+I  AK +IVAAD  VEMDRF GK
Sbjct: 180 TYMAADNLKSKAQELGVSIKVETNGSGGVKNRLTDEEISAAKAIIVAADTKVEMDRFHGK 239

Query: 239 QFIARPVADGIKKSQELISLILNNEGNTYHAKNGKSETAVSTEKTSLGG-----AFYKHL 293
             I  PV DGI++ +ELI   L  +   Y     +     S E  S GG      FYKHL
Sbjct: 240 PVIQVPVTDGIRRPKELIDQALAGKAPVY-----EGGAQASGEDGSAGGGRPKLGFYKHL 294

Query: 294 MGGVSQMLPFVIGGGIMIALAFLLDNMLGVPNDQLGSLGSYHEIAAIFMNIGGA-AFSFM 352
           M GVS MLPFV+GGGI+IA++F+       P+D      SYH  A + M IGG  AF  M
Sbjct: 295 MNGVSNMLPFVVGGGILIAISFMFGIKAFDPSDP-----SYHPFAEMLMTIGGGNAFGLM 349

Query: 353 LPVLAGYIAYSIAEKPGLVAGFVAGAIASNGLAFGKVPFAAGGEVSLGLTGVPSGFLGAL 412
           +PVLA +IA SIA++PG  AG + G IAS G A                    GFLG L
Sbjct: 350 IPVLAAFIAMSIADRPGFAAGMIGGLIASTGEA--------------------GFLGGL 388

Query: 413 VGGFLAGGVILALRKLLAGLPRSLEGVKSILLYPLLGVLVTGFLMLFVNIPMAAINTALN 472
           + GFLAG V L ++K+LA LP++L+G+K+IL YP+  + +TG +ML +  P+AA NT L
Sbjct: 389 IAGFLAGYVALGVKKVLANLPQTLDGIKTILFYPVFNIFITGMIMLVIVGPLAAFNTGLQ 448

Query: 473 DFLQGLSGSSAVLMGLLVGGMMAVDMGGPVNKAAYVFGTGTLAATVANGGSVVMAAVMAG 532
           D+L  +  ++ V++G+++GGMMAVDMGGP+NKAA+ FG   + A    G     AAVMAG
Sbjct: 449 DWLGSMGTANMVILGVILGGMMAVDMGGPINKAAFTFGIAMIDA----GNFGPHAAVMAG 504

Query: 533 GMVPPLAVFVATLLFKDKFTKEERESGLTNIVMGLSFITEGAIPFGAADPARAIPSFIAG 592
           GMVPPL + +AT LFK KFTK+ERE+G TN ++G SFITEGAIPF AADP R IPS I G
Sbjct: 505 GMVPPLGIALATTLFKKKFTKQEREAGKTNYILGASFITEGAIPFAAADPGRVIPSIIVG 564

Query: 593 SALTGALVGLAGIKLMAPHGGIFVI---ALTSNPILYLVFVVIGALVSGILFGALRKKA 648
           SA  G L  L  + L APHGG FVI     + +NP+LYLV ++ G++V+ +L G  +K A
Sbjct: 565 SAFAGGLTALFNVTLSAPHGGAFVIFIGNIVNNPLLYLVAIIAGSIVTALLLGFWKKDA 623
```

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

```
 Identities = 526/652 (80%), Positives = 581/652 (88%), Gaps = 6/652 (0%)

Query: 1   MKIQDLLKKEVMIMDLKATSKEAAIDEMITKLVDTGVVTNFAIFKDGIMKREAQTSTGLG 60
           MKIQDLL+K++MI+DL+A SKE AIDEMITKLV+  +V +F +FK  IM RE QTSTGLG
Sbjct: 1   MKIQDLLRKDIMILDLQAISKEVAIDEMITKLVEKDIVHDFDVFKKSIMTREEQTSTGLG 60

Query: 61  DGIAMPHSKNAAVKEATVLFAKSASGVDYEALDGQPTDLFFMIAAPDGANDTHLAALAEL 120
           DGIAMPHSKN  V +  VLFAKS  GVDY+ALDGQPTDLFFMIAAP GANDTHLAALAEL
Sbjct: 61  DGIAMPHSKNIVVDKPAVLFAKSNKGVDYKALDGQPTDLFFMIAAPQGANDTHLAALAEL 120

Query: 121 SKYLLKEGFADQLRQAKTPDDIIATFDSNSISQETVAPQTVQSTSKGSDYIVAVTACTTG 180
           S+YLLK+GFAD+LR A TP+ +IA FD  S ++E V   T      G D+IVAVTAC TG
Sbjct: 121 SQYLLKDGFADKLRAAATPEAVIAVFDEASTAKEEVVAPT-----SGQDFIVAVTACPTG 175

Query: 181 IAHTYMAEEALKKKAAEMGVGIKVETNGASGVGNKLTSSDIARAKGVIIAADKAVEMDRF 240
           IAHTYMAEEALKK+AAEMGV IKVETNGASGV N+LT+ DI RAKGVI+AADKAVEMDRF
Sbjct: 176 IAHTYMAEEALKKQAAEMGVAIKVETNGASGVANRLTAEDIQRAKGVIVAADKAVEMDRF 235

Query: 241 DGKPLVSRPVADGIKKSEDLINIILDNKAQTYHAKNQNDKQSGESDGKSGLGSAFYKHLM 300
           DGK  ++RPVADGIKKS++LI++IL+N+  TYHAKN    ++  S  K+ LG AFYKHLM
Sbjct: 236 DGKQFIARPVADGIKKSQELISLILNNEGNTYHAKN-GKSETAVSTEKTSLGGAFYKHLM 294

Query: 301 GGVSQMLPFVIGGGIMIAIAFLFDNILGVPKDQLSNLGSYHEIAALFKNIGGAAFAFMLP 360
           GGVSQMLPFVIGGGIMIA+AFL DN+LGVP DQL +LGSYHEIAA+F NIGGAAF+FMLP
Sbjct: 295 GGVSQMLPFVIGGGIMIALAFLLDNMLGVPNDQLGSLGSYHEIAAIFMNIGGAAFSFMLP 354
```

```
Query: 361 VLAGYIAYSIAEKPGLVAGFVAGSIASSGLAFGKVPFAEGGKATLALAGVPSGFLGALVG 420
           VLAGYIAYSIAEKPGLVAGFVAG+IAS+GLAFGKVPFA GG+ +L L GVPSGFLGALVG
Sbjct: 355 VLAGYIAYSIAEKPGLVAGFVAGAIASNGLAFGKVPFAAGGEVSLGLTGVPSGFLGALVG 414

Query: 421 GFLAGGVILLLRKLLSGLPKSLEGIKSILLYPLLGVLITGFLMLLVNIPMAAINTALNTF 480
           GFLAGGVIL LRKLL+GLP+SLEG+KSILLYPLLGVL+TGFLML VNIPMAAINTALN F
Sbjct: 415 GFLAGGVILALRKLLAGLPRSLEGVKSILLYPLLGVLVTGFLMLFVNIPMAAINTALNDF 474

Query: 481 LQGLSGSSAVLMGLLVGGMMAVDMGGPVNKAAYVFGTGTLAATVANGGSVVMAAVMAGGM 540
           LQGLSGSSAVLMGLLVGGMMAVDMGGPVNKAAYVFGTGTLAATVANGGSVVMAAVMAGGM
Sbjct: 475 LQGLSGSSAVLMGLLVGGMMAVDMGGPVNKAAYVFGTGTLAATVANGGSVVMAAVMAGGM 534

Query: 541 VPPLAVFVATLLFKDKFNNEERQSGLTNIVMGLSFITEGAIPFGAADPARAIPSFIVGSA 600
           VPPLAVFVATLLFKDKF  EER+SGLTNIVMGLSFITEGAIPFGAADPARAIPSFI GSA
Sbjct: 535 VPPLAVFVATLLFKDKFTKEERESGLTNIVMGLSFITEGAIPFGAADPARAIPSFIAGSA 594

Query: 601 LTGALVGLAGIKLMAPHGGIFVIALTSNPLLYILFILIGAVVSGVLFGLFRK 652
           LTGALVGLAGIKLMAPHGGIFVIALTSNP+LY++F++IGA+VSG+LFG  RK
Sbjct: 595 LTGALVGLAGIKLMAPHGGIFVIALTSNPILYLVFVVIGALVSGILFGALRK 646
```

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

Example 1954

A DNA sequence (GBSx2063) was identified in *S.agalactiae* <SEQ ID 6055> which encodes the amino acid sequence <SEQ ID 6056>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1532(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:AAC24914 GB:AF012285 fructose-1-phosphate kinase [Bacillus subtilis]
 Identities = 146/303 (48%), Positives = 197/303 (64%)

Query: 1   MIYTVTLNPSIDFIVRLDTLLLGSVNRMTSDDKYVGGKGINVSRILKRLKIDNTATGFIG 60
           MIYTVTLNPS+D+IV ++    +G +NR + D KY GGKGINVSR+LKR  + + A GF+G
Sbjct: 1   MIYTVTLNPSVDYIVHVEDFTVGGLNRSSYDTKYPGGKGINVSRLLKRHHVASKALGFVG 60

Query: 61  GFTGHFVEDGLVLEGIKTDFVSVNEDTRINVKVKAKIETEINGGGPRITNEQLHRLEKLL 120
           GFTG +++  L  E ++T F  V  DTRINVK+K   ETEING GP I++E      +
Sbjct: 61  GFTGEYIKTFLREENLETAFSEVKGDTRINVKLKTGDETEINGQGPTISDEDFKAFLEQF 120

Query: 121 SRLTPEDTVVFAGSAPASLGNKVYNTLIPIAKKTGAEVVCDFEGQTLLDALAYQPLLVKP 180
             L   D VV AGS P+SL +  Y  +    K+  A VV D  G+ LL A   +P L+KP
Sbjct: 121 QSLQEGDIVVLAGSIPSSLPHDTYEKIAEACKQQNARVVLDISGEALLKATEMKPFLMKP 180

Query: 181 NNHELADIFGVELEGLPDIEKYAHKILDKGAKNVIVSMAGDGALLVTPEASYFAKPIKGE 240
           N+HEL ++FG  +  + +   Y  K++++GA++VIVSMAGDGALL T EA YFA   KG+
Sbjct: 181 NHHELGEMFGTAITSVEEAVPYGKKLVEQGAEHVIVSMAGDGALLFTNEAVYFANVPKGK 240

Query: 241 VKNSVGAGDSMVAGFTGEFVKSKNPVEALKWGVACGTATTFSDDLATAEFIQDIYNKVEV 300
           + NSVGAGDS+VAGF     K     EA + GV  G+AT FS++L T EF+Q +  +V+V
Sbjct: 241 LVNSVGAGDSVVAGFLAGISKQLPLEEAFRLGVTSGSATAFSEELGTEEFVQQLLPEVKV 300

Query: 301 EKL 303
            +L
Sbjct: 301 TRL 303
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6057> which encodes the amino acid sequence <SEQ ID 6058>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1738(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 = 222/302 (73%), Positives = 261/302 (85%)

Query: 1   MIYTVTLNPSIDFIVRLDTLLLGSVNRMTSDDKYVGGKGINVSRILKRLKIDNTATGFIG 60
           MIYTVTLNPSIDFIVR+D + LGSVNRM SDDK+ GGKGINVSRIL+RL I +TATGF+G
Sbjct: 1   MIYTVTLNPSIDFIVRIDQINLGSVNRMASDDKFAGGKGINVSRILQRLDIASTATGFLG 60

Query: 61  GFTGHFVEDGLVLEGIKTDFVSVNEDTRINVKVKAKIETEINGGGPRITNEQLHRLEKLL 120
           GFTG F+E+ L  EG+KTDFV  ++DTRINVK+K++ ETE+NG GP I+ EQL  L+  L
Sbjct: 61  GFTGRFIEESLSAEGVKTDFVKGDQDTRINVKIKSQEETELNGQGPIISQEQLEDLKTKL 120

Query: 121 SRLTPEDTVVFAGSAPASLGNKVYNTLIPIAKKTGAEVVCDFEGQTLLDALAYQPLLVKP 180
           S+LT EDTVVFAGSAPA+LGN VY  L+P+ +++GA+VVCDFEGQTL+DALAY PLLVKP
Sbjct: 121 SQLTAEDTVVFAGSAPANLGNAVYKELLPLVRQSGAQVVCDFEGQTLIDALAYNPLLVKP 180

Query: 181 NNHELADIFGVELEGLPDIEKYAHKILDKGAKNVIVSMAGDGALLVTPEASYFAKPIKGE 240
           NNHEL  IFG  L  L D+E YA ++L+ GA+NVI+SMAGDGALLVT EA+YFAKPIKGE
Sbjct: 181 NNHELEAIFGTILTSLDDVETYARRLLEMGAQNVIISMAGDGALLVTKEATYFAKPIKGE 240

Query: 241 VKNSVGAGDSMVAGFTGEFVKSKNPVEALKWGVACGTATTFSDDLATAEFIQDIYNKVEV 300
           VKNSVGAGDSMVAGFTGEF+KS+NP+EALKWGVACGTAT FSDDLAT  FI++ Y+KVEV
Sbjct: 241 VKNSVGAGDSMVAGFTGEFMKSQNPIEALKWGVACGTATAFSDDLATIAFIKETYHKVEV 300

Query: 301 EK 302
           EK
Sbjct: 301 EK 302
```

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

Example 1955

A DNA sequence (GBSx2064) was identified in *S.agalactiae* <SEQ ID 6059> which encodes the amino acid sequence <SEQ ID 6060>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2769(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 9961> which encodes amino acid sequence <SEQ ID 9962> was also identified.

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

```
>GP:AAC24913 GB:AF012285 FruR [Bacillus subtilis]
 Identities = 97/247 (39%), Positives = 148/247 (59%), Gaps = 4/247 (1%)

Query: 23  MLKSKRKEIILSRLEQNKSVTLDELTSILETSESTVRRDLDELESAGFLKRVHGGAELPY 82
```

```
            ML  +R ++I+ ++E++  V + EL ++   SEST+RRDL  LE  GFLKRVHGGA
Sbjct: 1    MLTPERHQLIIDQIEKHDVVKIQELINLTNASESTIRRDLSTLEERGFLKRVHGGAAKLS 60

Query: 83   SLGQELSNQEKAIKNVQKKLDIARQTAKLIAKQDVIFIDAGTTTELLIDFLPH-EQLTVV 141
             +  E    EK+ KN+  KL IA + A L+ + D I++DAGTTT  +IDF+   + + VV
Sbjct: 61   DIRLEPDMLEKSSKNLHDKLKIAEKAASLLEEGDCIYLDAGTTTLHMIDFMDKTKDIVVV 120

Query: 142  TNSIHHAAKLVDRGIKTIIIGGAVKHSTDASIGQVAINQIRQITVDKAFLGMNGID-EVY 200
            TN + H   L+ + I   ++GG VKH T A IG  ++  + Q   DK+FLG NG+  E
Sbjct: 121  TNGVMHIDALIRKEISFYLLGGYVKHRTGAIIGGASLVAMDQYRFDKSFLGTNGVHTEAG 180

Query: 201  LTTPDLEEAAIKEAIINNSQQTFILMDSSKIGQVTFAKVKEINDINLVTNKTDSELMTII 260
             TTPD +EA +K+  I  ++  ++L D SK G+++F+    I D  ++T  TD+E +T
Sbjct: 181  FTTPDPDEALLKQKAIKQAKHAYVLADPSKFGEISFSAFAGIGDATIIT--TDAEELTFD 238

Query: 261  KEKMKVI 267
              + K +
Sbjct: 239  NYQEKTV 245
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6061> which encodes the amino acid sequence <SEQ ID 6062>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2604(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 = 135/237 (56%), Positives = 184/237 (76%)

Query: 33   LSRLEQNKSVTLDELTSILETSESTVRRDLDELESAGFLKRVHGGAELPYSLGQELSNQE 92
            ++++ +   V+L++L  +L +SEST+RRDL ELE  G L RVHGGAEL +SL +ELSNQE
Sbjct: 1    MAKITEENYVSLEDLMQLLNSSESTIRRDLGELEQEGRLHRVHGGAELFHSLQEELSNQE 60

Query: 93   KAIKNVQKKLDIARQTAKLIAKQDVIFIDAGTTTELLIDFLPHEQLTVVTNSIHHAAKLV 152
            K++KN   K  IA++ ++LI   DVIFIDAGTTTE L+ FL  + LTVVTNSIHHAA+LV
Sbjct: 61   KSVKNSHIKKAIAQRASQLIYDNDVIFIDAGTTTEFLLPFLQAKNLTVVTNSIHHAARLV 120

Query: 153  DRGIKTIIIGGAVKHSTDASIGQVAINQIRQITVDKAFLGMNGIDEVYLTTPDLEEAAIK 212
            +  I+TII+GG VK +TDASIG VA+ QIRQ+  DKAFLGMNG+D+ YLTTPD+EEA IK
Sbjct: 121  ELSIETIIVGGYVKQTTDASIGNVALEQIRQMNFDKAFLGMNGVDDSYLTTPDMEEAVIK 180

Query: 213  EAIINNSQQTFILMDSSKIGQVTFAKVKEINDINLVTNKTDSELMTIIKEKMKVIQV 269
            +A+++N++  +IL+D +KIGQV+F KV  IND+ ++T    + ++  IKEK KVI++
Sbjct: 181  KAVLSNAKLAYILVDGTKIGQVSFVKVAPINDVTIITLGGSASILKQIKEKAKVIEL 237
```

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

Example 1956

A DNA sequence (GBSx2065) was identified in *S.agalactiae* <SEQ ID 6063> which encodes the amino acid sequence <SEQ ID 6064>. This protein is predicted to be beta-lactam resistance factor. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5777(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:CAB89121 GB:AJ277485 beta-lactam resistance factor
          [Streptococcus pneumoniae]
 Identities = 215/410 (52%), Positives = 283/410 (68%)

Query: 1   MTLRELTIEEFKEHSGNYDSQSFLQTPEMAKLLEKRGYDVRYLGYQVENKLEIISLSYIM 60
           M L  LT EEF+ +S    S+SF+Q+ +M  LLEKRG  + YL  + E ++++ +L Y +
Sbjct: 1   MALTTLTKEEFQTYSDQVSSRSFMQSVQMGDLLEKRGARIVYLALKQEGEIQVAALVYSL 60

Query: 61  PVTGGFQMKIDSGPVHSNSKYLKQFYKALQGYAKSNGVLELIVEPYDDYQLFTSSGVPSN 120
           P+ GG  M+++SGP+++    L  FY  L+ YAK NGVLEL+V+PY+ YQ F S G P +
Sbjct: 61  PMLGGLHMELNSGPIYTQQDALPVFYAELKEYAKQNGVLELLVKPYETYQTFDSQGNPID 120

Query: 121 QGNDNLIEDFTSSGYHHDGLTTGFTGKYLSWHYVKNLEGVTSETLLSSFSKTGRALVKKA 180
               ++I+D T  GY  DGLTTG+ G    W Y K+L  +T ++LL SFSK G+ LVKKA
Sbjct: 121 AEKKSIIQDLTDLGYQFDGLTTGYPGGEPDWLYYKDLTELTEKSLLKSFSKKGKPLVKKA 180

Query: 181 MSFGIKVRVLKRDELHLFKEITTSTSNRRDYMDKSLDYYQDFYDSFEGKAEFVIATLNFR 240
            +FGI+++ LKR+EL +FK IT  TS RR+Y DKSL+YY+ FYD+F  +AEF+IA+LNF
Sbjct: 181 ETFGIRLKKLKREELSIFKNITKETSERREYSDKSLEYYEHFYDTFGEQAEFLIASLNFS 240

Query: 241 EYDHNLQIKAEALENKLKLLDERFRENADSPKYHRQRSEIINQLASFETRRQEVQSFIQK 300
           +Y   LQ +   LE  L  L     +N  S K   Q  E  +Q  +FE R+ E +  I+K
Sbjct: 241 DYMSKLQGEQSKLEENLDKLRLDLSKNPHSEKKQNQLREYSSQFETFEVRKAEARDLIEK 300

Query: 301 YDNQDVVLAGSLFVYSLKETVYFFSGSYTEFNKFYAPAVLQEYVMQEALKRGSTFYNLLG 360
           Y  +D+VLAGSLFVY  +ET Y FSGSYTEFNKFYAPA+LQ+YVM E++KRG   YN LG
Sbjct: 301 YGEEDIVLAGSLFVYMPQETTYLFSGSYTEFNKFYAPALLQKYVMLESIKRGIPKYNFLG 360

Query: 361 IQGTFDGSDSILRFKQNFNGCIIRKMGTFNYYPSPFKYKGIQLLKKVLKR 410
           IQG FDGSD +LRFKQNFNG I+RK GTF Y+PSP KYK IQLLKK++ R
Sbjct: 361 IQGIFDGSDGVLRFKQNFNGYIVRKAGTFRYHPSPLKYKAIQLLKKIVGR 410
```

There is also homology to SEQ ID 5460.

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

Example 1957

A DNA sequence (GBSx2066) was identified in *S.agalactiae* <SEQ ID 6065> which encodes the amino acid sequence <SEQ ID 6066>. This protein is predicted to be cell wall protein, 40 kDa (sr 5' region). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.45   Transmembrane   25 -  41 (  23 -  42)

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

A related GBS nucleic acid sequence <SEQ ID 9963> which encodes amino acid sequence <SEQ ID 9964> was also identified.

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

```
!GB:AF278686 choline binding protein D; CbpD [Strept...
!GB:AF278686 choline binding protein D; CbpD [Strept...
```

```
>GP:AAF87768 GB:AF278686 choline binding protein D; CbpD
          [Streptococcus pneumoniae]
 Identities = 63/230 (27%), Positives = 108/230 (46%), Gaps = 34/230 (14%)

Query: 324 WTEQGGQDDIKWYTAVTTGDG------NYKVAVSFADHKNEKGLYNIHLYYQEASGTLVG 377
           W+  G    + W + V  GD       NY    S+     +       +++++  G  VG
Sbjct: 123 WSTAGTYGHVAWVSNVM-GDQIEIEEYNYGYTESYNKRVIKANTMTGFIHFKDLDGGSVG 181

Query: 378 VTGTKVTVAGTNSSQEPIENGLAKTGVYNIIGSTEVKNEAKISSQTQFTLEKGDKINYDQ 437
            + +  + GT+  +                    + +K E   S          G+K++YDQ
Sbjct: 182 NSQSSTSTGGTHYFKT----------------KSAIKTEPLASGTVIDYYYPGEKVHYDQ 225

Query: 438 VLTADGYQWISYKSYSGVRRYIPVKKLTTSSEKAKDEATKPTSYPNLPKTG-TYTFTKTV 496
           +L  DGY+W+SY +Y+G  RY+ ++ +  +           P     L  TG T+ F
Sbjct: 226 ILEKDGYKWLSYTAYNGSYRYVQLEAVNKN----------PLGNSVLSSTGGTHYFKTKS 275

Query: 497 DVKSQPKVSSPVEFNFQKGEKIHYDQVLVVDGHQWISYKSYSGIRRYIEI 546
            +K++P VS+ V   +  GEK+HYDQ+L  DG++W+SY +Y+G RRYI++
Sbjct: 276 AIKTEPLVSATVIDYYYPGEKVHYDQILEKDGYKWLSYTAYNGSRRYIQL 325
 Identities = 49/161 (30%), Positives = 85/161 (52%), Gaps = 14/161 (8%)

Query: 116 GNYVYSKETEVKNTPSKSAPVAFYAKKGDKVFYDQVFNKDNVKWISYKSFCGVRRYAAIE 175
           G + +  ++ +K  P  S  V  Y   G+KV YDQ+  KD   KW+SY ++ G  RY  +E
Sbjct: 191 GTHYFKTKSAIKTEPLASGTVIDYYYPGEKVHYDQILEKDGYKWLSYTAYNGSYRYVQLE 250

Query: 176 SLDPSGGSETKAPTPVTNSGSNNQEKIATQGNYTFSHKVEVKNEAKVASPTQFTLDKGDR 235
           +++ +         P+ NS  +     +T G + F   K  +K E   V++         G++
Sbjct: 251 AVNKN---------PLGNSVLS-----STGGTHYFKTKSAIKTEPLVSATVIDYYYPGEK 296

Query: 236 IFYDQILTIEGNQWLSYKSFNGVRRFVLLGKASSVEKTEDK 276
           + YDQIL  +G +WLSY ++NG RR++ L   +S +  +++
Sbjct: 297 VHYDQILEKDGYKWLSYTAYNGSRRYIQLEGVTSSQNYQNQ 337
 Identities = 52/192 (27%), Positives = 90/192 (46%), Gaps = 13/192 (6%)

Query: 295 ISNETTTGFDILITNIKDDNGIAAVKVPVWTEQGGQDDIKWYTAVTTGDGNYKVAVSFAD 354
           I   T TGF     + KD +G +        T  GG    K  +A+ T         + +
Sbjct: 161 IKANTMTGF----IHFKDLDGGSVGNSQSSTSTGGTHYFKTKSAIKTEPLASGTVIDYY- 215

Query: 355 HKNEKGLYNIHLY---YQEASGTLVGVTGTKVTVAGTNSSQEPIENGLAKT--GVYNIIG 409
           + EK  Y+  L     Y+  S T     +   V +   N +  P+ N +  +  G +
Sbjct: 216 YPGEKVHYDQILEKDGYKWLSYTAYNGSYRYVQLEAVNKN--PLGNSVLSSTGGTHYFKT 273

Query: 410 STEVKNEAKISSQTQFTLEKGDKINYDQVLTADGYQWISYKSYSGVRRYIPVKKLTTSSE 469
            + +K E  +S+         G+K++YDQ+L  DGY+W+SY +Y+G RRYI ++ +  TSS+
Sbjct: 274 KSAIKTEPLVSATVIDYYYPGEKVHYDQILEKDGYKWLSYTAYNGSRRYIQLEGV-TSSQ 332

Query: 470 KAKDEATKPTSY 481
            +++++  +SY
Sbjct: 333 NYQNQSGNISSY 344
 Identities = 33/113 (29%), Positives = 56/113 (49%), Gaps = 2/113 (1%)

Query: 91  NTATKDITTPLVETKPMVEKTLPEQGNYVYSK-ETEVKNTPSKSAPVAFYAKKGDKVFYD 149
           N + + +    V   P+    L   G   Y K ++ +K  P  SA V  Y   G+KV YD
Sbjct: 241 NGSYRYVQLEAVNKNPLGNSVLSSTGGTHYFKTKSAIKTEPLVSATVIDYYYPGEKVHYD 300

Query: 150 QVFNKDNVKWISYKSFCGVRRYAAIESLDPSGGSETKAPTPVTNSGSNNQEKI 202
           Q+  KD  KW+SY ++ G RRY  +E + S   + ++   +++ GS++   +
Sbjct: 301 QILEKDGYKWLSYTAYNGSRRYIQLEGVTSSQNYQNQSGN-ISSYGSHSSSTV 352
```

A related GBS gene <SEQ ID 8937> and protein <SEQ ID 8938> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      -6.74
GvH: Signal Score (-7.5): 1.26
     Possible site: 42
```

```
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -3.45 threshold:  0.0
  INTEGRAL    Likelihood = -3.45   Transmembrane   22 -  39 (  23 -  42)
  PERIPHERAL  Likelihood =  6.26     371
 modified ALOM score:   1.19

*** Reasoning Step: 3

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

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

```
41.2/57.9% over 283aa
                                                                    Streptococ
cus mutans
  EGAD|33594| cell wall protein, 40 kDa (sr 5' region) Insert characterized
  PIR|A60328|A60328 40K cell wall protein precursor (sr 5' region) - (strain OMZ
175, serotype f) Insert characterized

ORF02145(301 - 1803 of 2238)
EGAD|33594|34911(30 - 313 of 335) cell wall protein, 40 kDa (sr 5' region) (Stre
ptococcus mutans)PIR|A60328|A60328 40K cell wall protein precursor (sr 5' region
) - Streptococcus mutans (strain OMZ175, serotype f)
%Match = 8.0
%Identity = 41.1  %Similarity = 57.9
Matches = 81  Mismatches = 79  Conservative Sub.s = 33

156       186       216       246       276       306       336       366
*YA****FCYTKNNKSWVFFSRSIYSIKYYICITNISKIC*HVTKRIL***CK*IRK*VFMMKKGQVNDTKQSYSLRKYK
                                  :      :    :         : | :||:  :|   | :|:|||
                               MNQKIVVISSFYMLGAHSFSKAVYHNDRSVKLMKRIDINHQAQRFSIRKYA
                                       10        20        30        40        50

396       426       456       486       516       546       576       606
FGLASVILGSFIMVTSPVFADQTTSVQVNNQTGTSVDANNSSNETSASSVITSNNDSVQASDKVVNSQNTATKDITTPLV
|| |||::|    : :   : |    |:                  | ::|:    :||  |  | :  | |    || |
FGAASVLIGCVFFLGTQNVSAQEQGTQL---------------PASENAVVNVAENSVAISQAVADKAATQTTLTETPQV
                   70                          80        90       100       110

654       684       714
ETKPMVEK-------------------------------------------TLPEQGNYVYSKETEVKNTPSKSAPVAF
| :    |                                             ::| |||||: : | |||  | |:|
EVEEKESKVNAPALNVDDKGAKSKEDVN~~~~AEQNEKAVRENLMCRQAKAVSIPSQGNYVFQETTPVKNAASMSSP---
           130       140           200       210       220       230       240

744               1533      1563      1593      1623      1653      1683
YAKKGDKVFYDQVFNKD~~~~GVYNIIGSTEVKNEAKISSQTQFTLEKGDKINYDQVLTADGYQWISYKSYSGVRRYIPV
                                          ||| ::|||| || || |||:||||| ||||:||| |:
-------------------------------------TQFNFDKGDKVFYDNVLEADGHQWISYVSYSGIRRYAPI
                                              250       260       270

1713      1743      1773      1803      1833      1863      1893      1923
KKLTTSSEKAKDEATKPTSYPNLPKTGTYTFTKTVDVKSQPKVSSPVEFNFQKGEKIHYDQVLVVDGHQWISYKSYSGIR
      :   ::   |     |||  ||| |||   :| :          |         :
------AVTIEELKQKEIVQQNLPAQGTYHFTKQQSLKMKLNCLVRPNSRFTTEITFFMIRF
               290       300       310       320       330
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6067> which encodes the amino acid sequence <SEQ ID 6068>. 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 databases:

```
>GP:AAF87768 GB:AF278686 choline binding protein D; CbpD
          [Streptococcus pneumoniae]
 Identities = 93/217 (42%), Positives = 136/217 (61%), Gaps = 18/217 (8%)

Query: 42  GDNYPSKWKKGNG-IDSWNMYIRQCTSFAAFRLSSANGFQLPKGYGNACTWGHIAKNQGY 100
           GD+YP+ +K G+  ID W MY RQCTSF AFRLS+ NGF++P  YGNA  WGH A+ +GY
Sbjct: 51  GDDYPAYYKNGSQEIDQWRMYSRQCTSFVAFRLSNVNGFEIPAAYGNANEWGHRARREGY 110

Query: 101 PVNKTPSIGAIAWFDKNAYQSNAAYGHVAWVADIRGDTVTIEEYNYNAGQGPERYHKRQI 160
            V+ TP+IG+I W       +   YGHVAWV+++ GD + IEEYNY      E Y+KR I
Sbjct: 111 RVDNTPTIGSITW------STAGTYGHVAWVSNVMGDQIEIEEYNYGY---TESYNKRVI 161

Query: 161 PKSQVSGYIHFKDLSSQTSHSYPRQLKHISQASFDPSGTYHFTTRLPVKGQTSIDSPDLA 220
             + ++G+IHFKDL   +  +        SQ+S    GT++F T+  +K +       +
Sbjct: 162 KANTMTGFIHFKDLDGGSVGN--------SQSSTSTGGTHYFKTKSAIKTEPLASGTVID 213

Query: 221 YYEAGQSVYYDKVVTAGGYTWLSYLSFSGNRRYIPIK 257
           YY  G+ V+YD+++   GY WLSY +++G+ RY+ ++
Sbjct: 214 YYYPGEKVHYDQILEKDGYKWLSYTAYNGSYRYVQLE 250
```

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

```
 Identities = 34/94 (36%), Positives = 52/94 (55%)

Query: 453 SGVRRYIPVKKLTTSSEKAKDEATKPTSYPNLPKTGTYTFTKTVDVKSQPKVSSPVEFNF 512
           S V  YI  K L++ +  +     K  S  +   +GTY FT  + VK Q  + SP    +
Sbjct: 163 SQVSGYIHFKDLSSQTSHSYPRQLKHISQASFDPSGTYHFTTRLPVKGQTSIDSPDLAYY 222

Query: 513 QKGEKIHYDQVLVVDGHQWISYKSYSGIRRYIEI 546
           + G+ ++YD+V+   G+ W+SY S+SG RRYI I
Sbjct: 223 EAGQSVYYDKVVTAGGYTWLSYLSFSGNRRYIPI 256
 Identities = 30/78 (38%), Positives = 45/78 (57%), Gaps = 2/78 (2%)

Query: 402 TGVYNIIGSTEVKNEAKISSQTQFTLEKGDKINYDQVLTADGYQWISYKSYSGVRRYIPV 461
           +G Y+      VK +  I S      E G  + YD+V+TA GY W+SY S+SG RRYIP+
Sbjct: 197 SGTYHFTTRLPVKGQTSIDSPDLAYYEAGQSVYYDKVVTAGGYTWLSYLSFSGNRRYIPI 256

Query: 462 KKLTTSSEKAKDEATKPT 479
           K+   +    +++ TKP+
Sbjct: 257 KE--PAQSVVQNDNTKPS 272
 Identities = 27/94 (28%), Positives = 47/94 (49%)

Query: 198 NQEKIATQGNYTFSHKVEVKNEAKVASPTQFTLDKGDRIFYDQILTIEGNQWLSYKSFNG 257
           +Q      G Y F+ ++ VK +  + SP     + G  ++YD+++T  G  WLSY SF+G
Sbjct: 190 SQASFDPSGTYHFTTRLPVKGQTSIDSPDLAYYEAGQSVYYDKVVTAGGYTWLSYLSFSG 249

Query: 258 VRRFVLLGKASSVEKTEDKEKVSPQPQARITKTG 291
           RR++ + + +      D  K S +    +T  G
Sbjct: 250 NRRYIPIKEPAQSVVQNDNTKPSIKVGDTVTFPG 283
 Identities = 23/73 (31%), Positives = 35/73 (47%)

Query: 103 ETKPMVEKTLPEQGNYVYSKETEVKNTPSKSAPVAFYAKKGDKVFYDQVFNKDNVKWISY 162
           + K + + +    G Y ++    VK   S  +P   Y + G  V+YD+V       W+SY
Sbjct: 185 QLKHISQASFDPSGTYHFTTRLPVKGQTSIDSPDLAYYEAGQSVYYDKVVTAGGYTWLSY 244

Query: 163 KSFCGVRRYAAIE 175
           SF G RRY  I+
Sbjct: 245 LSFSGNRRYIPIK 257
```

SEQ ID 8938 (GBS91) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 7; MW 63kDa).

The GBS91-His fusion product was purified (Figure 195, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 283), 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 1958

A DNA sequence (GBSx2067) was identified in *S.agalactiae* <SEQ ID 6069> which encodes the amino acid sequence <SEQ ID 6070>. This protein is predicted to be thiamine biosynthesis protein. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0984(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:CAB49673 GB:AJ248285 PROBABLE 2-DEHYDROPANTOATE 2-REDUCTASE (EC
          1.1.1.169) [Pyrococcus abyssi]
 Identities = 85/301 (28%), Positives = 150/301 (49%), Gaps = 7/301 (2%)

Query: 1   MLVYIAGSGAMGCRFGYQISKTNHDVILLDNWADHIMAIKENGLKVTGDTEDLVKLPIMK 60
           M +YI G+GA+G  FG  ++    DV+L+     H+ AI E GLK+ G  +  VK+
Sbjct: 1   MKIYILGAGAIGSLFGGLLANAGEDVLLIGR-DPHVSAINEKGLKIVGIKDLNVKVEATT 59

Query: 61  PTDATEEADLIILFTKAMQLPNMLQDIKKIIGKKTKVLCLLNGLGHEDVIRQYIPEHNIL 120
               E+ DLI+L TK+      L+  + I+ K + VL + NG+G+ED I ++       +
Sbjct: 60  RVPE-EKPDLIVLATKSYSTIEALKSARHIV-KGSWVLSIQNGIGNEDKIIEF--GGKAI 115

Query: 121 MGVTVWTAGLKGPGHAHLEGVGSVNLQSIDPNNQEAGHRVTELLNEAKLQATYDENVLPN 180
            G+T   A ++ PG     G G   +       ++   +V ++ N A ++    EN++
Sbjct: 116 GGITTNGAMVEAPGVIKWTGKGVTIIGLYPQGKEKFIEKVADVFNSADIETHVSENIISW 175

Query: 181 IWRKACVNGTMNSTCALLDCTIGQLFASEDGVNMVHEIIHEFVTVGKAEGVELDEEEITK 240
           IW KA VN  +N    LL+     +  ++  ++M  E++ E   V    G+E D   +
Sbjct: 176 IWAKAIVNSAINPIGTLLEVKNKVIRENDFLLSMAMEVVKEGCRVALQNGIEFDVPPMDL 235

Query: 241 YVMDTSVKAAHHYPSMHQDLVQNQRLTEIDFLNGAVNKKGENLGIDTPYCRLITQLIHTKE 301
           +   T  +   +Y SM QD+ + ++ TE+D++NG + +  + + ++ P   L+ LI  KE
Sbjct: 236 F-FQTLEQTRENYNSMLQDIWRGKK-TEVDYINGKIVEYAKAVNLEAPMNLLLWGLIKGKE 294
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6071> which encodes the amino acid sequence <SEQ ID 6072>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1392(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 = 262/307 (85%), Positives = 288/307 (93%)

Query: 1   MLVYIAGSGAMGCRFGYQISKTNHDVILLDNWADHIMAIKENGLKVTGDTEDLVKLPIMK 60
           MLVYIAGSGAMGCRFGYQISKTN+DVILLDNW DHI AIKENGL VTGD E+ VKLPIMK
Sbjct: 1   MLVYIAGSGAMGCRFGYQISKTNNDVILLDNWEDHINAIKENGLVVTGDVEETVKLPIMK 60

Query: 61  PTDATEEADLIILFTKAMQLPNMLQDIKKIIGKKTKVLCLLNGLGHEDVIRQYIPEHNIL 120
           PT+AT+EADLIILFTKAMQLP MLQDIK IIGK+TKVLCLLNGLGHEDVIRQYIPEHNIL
Sbjct: 61  PTEATQEADLIILFTKAMQLPQMLQDIKGIIGKETKVLCLLNGLGHEDVIRQYIPEHNIL 120

Query: 121 MGVTVWTAGLKGPGHAHLEGVGSVNLQSIDPNNQEAGHRVTELLNEAKLQATYDENVLPN 180
           MGVTVWTAGL+GPG AHL+GVG++NLQS+DP+NQEAGH+V +LLNEA L ATYDENV+PN
Sbjct: 121 MGVTVWTAGLEGPGRAHLQGVGALNLQSMDPSNQEAGHQVADLLNEANLNATYDENVVPN 180

Query: 181 IWRKACVNGTMNSTCALLDCTIGQLFASEDGVNMVHEIIHEFVTVGKAEGVELDEEEITK 240
           IWRKACVNGTMNSTCALLDCTIG+LFASEDG+ MV EIIHEFV VG+AEGVEL+EEEIT+
Sbjct: 181 IWRKACVNGTMNSTCALLDCTIGELFASEDGLKMVKEIIHEFVIVGQAEGVELNEEEITQ 240

Query: 241 YVMDTSVKAAHHYPSMHQDLVQNQRLTEIDFLNGAVNKKGENLGIDTPYCRLITQLIHTK 300
           YVMDTSVKAAHHYPSMHQDLVQN RLTEIDF+NGAVN KGE LGI+TPYCR+IT+L+H K
Sbjct: 241 YVMDTSVKAAHHYPSMHQDLVQNHRLTEIDFINGAVNTKGEKLGINTPYCRMITELVHAK 300

Query: 301 ENVLSIK 307
           E VL+I+
Sbjct: 301 EAVLNIQ 307
```

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

Example 1959

A DNA sequence (GBSx2068) was identified in *S.agalactiae* <SEQ ID 6073> which encodes the amino acid sequence <SEQ ID 6074>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.03   Transmembrane   61 -  77 (  61 -  78)
   INTEGRAL    Likelihood = -1.33   Transmembrane   80 -  96 (  79 -  96)

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

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 1960

A DNA sequence (GBSx2069) was identified in *S.agalactiae* <SEQ ID 6075> which encodes the amino acid sequence <SEQ ID 6076>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.82   Transmembrane  317 - 333 ( 304 - 335)
   INTEGRAL    Likelihood = -7.64   Transmembrane  187 - 203 ( 183 - 217)
   INTEGRAL    Likelihood = -5.26   Transmembrane   24 -  40 (  18 -  44)
   INTEGRAL    Likelihood = -5.04   Transmembrane  143 - 159 ( 139 - 161)
```

```
   INTEGRAL    Likelihood = -2.34   Transmembrane  116 - 132 ( 115 - 136)
   INTEGRAL    Likelihood = -2.13   Transmembrane   55 -  71 (  55 -  71)
   INTEGRAL    Likelihood = -0.96   Transmembrane  268 - 284 ( 268 - 284)

----- Final Results -----
              bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema
          pallidum]
 Identities = 138/358 (38%), Positives = 220/358 (60%), Gaps = 18/358 (5%)

Query: 2   TNTVTPKETAGSFINKVLGGTATAIVVALIPNAILATFLKPFLSYG-LAAEFLHIVQVFQ 60
           T +++P++    F+ K+L G++  IV+ L+P AI     +       L A   H+V   Q
Sbjct: 3   TQSLSPRQ----FMMKILNGSSAGIVIGLVPPAIAGELFRALAPLSPLFAALYHVVLPIQ 58

Query: 61  FFTPIMAGFLIGQQFKFTPMQQLAVGGAAYIGSGAWAYTEVIQKGVATGSFQLRGIGDLI 120
           F  P + G L+G QF  +  +   +  + I SG          +  G++ + GIGD+I
Sbjct: 59  FSVPALIGTLVGLQFHCSAPEVATLAFVSVIASG--------NVTLQNGAWLITGIGDVI 110

Query: 121 NMMLTAALAVLAVKWFGNKFGSLTIILLPIIIGTGVGYLGWKLLPYVSYVTTLIGQGINS 180
           N+ML +ALA++ V+     K GSLTII LP+I+    G +G    LPYV  +T  +G+ I +
Sbjct: 111 NVMLISALAIILVRALRGKLGSLTIIALPVIVAVVAGGVGSFSLPYVKMITLFVGRVIAT 170

Query: 181 FTTLQPIAMSILIAMAFSMLIVSPISTVAIGLAIGLNGMSASAASMGVASTTAVLVWATM 240
           F  LQP+ MSIL++M+FS++I+SP+S+VA+G+A+GL G+++ AA++GV+S    L+  TM
Sbjct: 171 FIALQPLLMSILLSMSFSLIIISPVSSVAVGIAVGLTGLASGAANIGVSSCAMTLIVGTM 230

Query: 241 KANKSGVPIAIALGAMKMMMPNFLKHPVMAIPMLMTATVSSLTVPLFKLVGTPASSGFGL 300
           + NK GVP+A+  GAMKM+MPN++++P++ IP+L+   V  +   LF L GTPAS+GFG
Sbjct: 231 RVNKIGVPLAMFAGAMKMLMPNWIRYPILNIPLLLNGLVCGVLAWLFNLQGTPASAGFGF 290

Query: 301 VGAVGPIASFE--AGASML---IVILSWLVIPFAVGFVSHKICKDILKLYKDDIFVFE 353
           +G VGPI ++    A   M+   I+ L + V+ F    ++   I  D LKLY+ ++F+ E
Sbjct: 291 IGLVGPINAYRLMAYTPMVRAGILFLVYFVLSFLAAYLIDFILVDRLKLYRRELFIPE 348
```

There is also homology to SEQ ID 1280.

A related GBS gene <SEQ ID 8939> and protein <SEQ ID 8940> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     -7.24
GvH: Signal Score (-7.5): -2.94
     Possible site: 49
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 7 value:  -9.82 threshold:  0.0
   INTEGRAL    Likelihood = -9.82   Transmembrane  317 - 333 ( 304 - 335)
   INTEGRAL    Likelihood = -7.64   Transmembrane  187 - 203 ( 183 - 217)
   INTEGRAL    Likelihood = -6.37   Transmembrane  143 - 159 ( 136 - 161)
   INTEGRAL    Likelihood = -5.26   Transmembrane   24 -  40 (  18 -  44)
   INTEGRAL    Likelihood = -2.34   Transmembrane  116 - 132 ( 115 - 136)
   INTEGRAL    Likelihood = -2.13   Transmembrane   55 -  71 (  55 -  71)
   INTEGRAL    Likelihood = -0.96   Transmembrane  268 - 284 ( 268 - 284)
   PERIPHERAL  Likelihood =  0.69     205
 modified ALOM score:   2.46

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
ORF02147(337 - 1359 of 1668)
EGAD|138195|TP0038(10 - 348 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 = 21.6
%Identity = 40.1  %Similarity = 65.6
Matches = 135  Mismatches = 112  Conservative Sub.s = 86

87       117       147       177       207       237       267       297
LQQDMGKHQSL*TKLSIIFILIEITV*SIQHH**NNYK*N**VYKKGLYILLKK*QSFLFIL*YN*LCRYE*Y*INEARY

327       357       387       417       444       474       504       534
FMTNTVTPKETAGSFINKVLGGTATAIVVALIPNAILATFLKPFLSYG-LAAEFLHIVQVFQFFTPIMAGFLIGQQFKFT
             |: |:| |::  ||: |:| ||   ::: :     | | : |:|   ||  | : | |:| ||  :
    MHTQSLSPRQFMMKILNGSSAGIVIGLVPPAIAGELFRALAPLSPLFAALYHVVLPIQFSVPALIGTLVGLQFHCS
             10        20        30        40        50        60        70

564       594       624       654       684       714       744       774
PMQQLAVGGAAYIGSGAWAYTEVIQKGVATGSFQLRGIGDLINMMLTAALAVLAVKWFGNKFGSLXIILLPIIIGTGVGY
  :   :   : | ||          :  |:: : ||||:||:|| :|||:: |: :  |:||| || ||:|:    |
APEVATLAFVSVIASG--------NVTLQNGAWLITGIGDVINVMLISALAIILVRALRGKLGSLTIIALPVIVAVVAGG
             90               100       110       120       130       140

804       834       864       894       924       954       984       1014
LGWKLLPYVSYVTTLIGQGINSFTTLQPIAMSILIAMAFSMLIVSPISTVAIGLAIGLNGMSASAASMGVASTTAVLVWA
:|   ||||  :| ::|: | :|  |||: ||||::|:|::|:||:|:||:|:|:|| |::: ||::||:|    |:
VGSFSLPYVKMITLFVGRVIATFIALQPLLMSILLSMSFSLIIISPVSSVAVGIAVGLTGLASGAANIGVSSCAMTLIVG
         160       170       180       190       200       210       220

1044      1074      1104      1134      1164      1194      1224      1248
TMKANKSGVPIAIALGAMKMMMPNFLKHPVMAIPMLMTATVSSLTVPLFKLVGTPASSGFGLVGAVGPIASFE--AGASM
||: || |||:|:  |||||:|||::::|:: ||:|:   |  :   || | |||||:|||::| |||| ::   |  |
TMRVNKIGVPLAMFAGAMKMLMPNWIRYPILNIPLLLNGLVCGVLAWLFNLQGTPASAGFGFIGLVGPINAYRLMAYTPM
         240       250       260       270       280       290       300

1269      1299      1329      1359      1389      1419      1449      1479
L---IVILSWLVIPFAVGFVSHKICKDILKLYKDDIFVFEGQN*FGGCMLVYIAGSGAMGCRFGYQISKTNHDVILLDNW
:   |: | ::|: |   ::   |  | ||||: ::|: |
VRAGILFLVYFVLSFLAAYLIDFILVDRLKLYRRELFIPEQG
         320       330       340       350
```

There is also homology to SEQ ID 1276

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

Example 1961

A DNA sequence (GBSx2070) was identified in *S.agalactiae* <SEQ ID 6077> which encodes the amino acid sequence <SEQ ID 6078>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have a cleavable N-term signal seq.

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

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

```
>GP:BAB07127 GB:AP001518 thioredoxin reductase [Bacillus halodurans]
 Identities = 163/325 (50%), Positives = 222/325 (68%), Gaps = 3/325 (0%)
```

```
Query: 5   IYDITIVGGGPVGLFAAFYAGLRGVSVKIIESLSELGGQPAILYPEKKIYDIPGYPVITG 64
           +YDITI+GGGP GLFAAFY G+R   VKIIES+ +LGGQ A LYPEK IYD+ G+P +
Sbjct: 7   LYDITIIGGGPTGLFAAFYGGMRQAKVKIIESMPQLGGQLAALYPEKYIYDVAGFPKVKA 66

Query: 65  RELIDKHIEQLERFKDSIEICLKEEVLSFEK-VDDVFTIQTDKDQHLSRAIVFACGNGAF 123
           ++L++    Q E+F  +I   L++ V +  K  DD FTI+TDK+ H S+AI+   G GAF
Sbjct: 67  QDLVNDLKRQAEQFNPTI--ALEQSVQNVTKETDDTFTIKTDKETHYSKAIIITAGAGAF 124

Query: 124 APRLLGLENEENYADNNLFYNVTKLEQFAGKHVVICGGGDSAVDWANELDKIAASVAIVH 183
            PR L +E  + Y   NL Y V  L  +AGK+V+I GGGDSAVDWA  L+ +A +V ++H
Sbjct: 125 QPRRLEVEGAKQYEGKNLQYFVNDLNAYAGKNVLISGGGDSAVDWALMLEPVAKNVTLIH 184

Query: 184 RRDAFRAHEHSVDILKASGVRILTPYVPIGLNGDSQRVSSLVVQKVKGDEVIELPLDNLI 243
           RRD FRAHEHSV++L+ S V ILTP+    L+GD +++  + +Q+VKGD V  L +D +I
Sbjct: 185 RRDKFRAHEHSVELLQKSSVNILTPFAISELSGDGEKIHHVTIQEVKGDAVETLDVDEVI 244

Query: 244 VSFGFSTSNKNLRYWNLDYKRSSINVSSLFETTQEGVYAIGDAANYPGKVELIATGYGEA 303
           V+FGF +S   ++ W L+ +++SI V++  ET   G+YA GD   YPGKV+LIATG+GEA
Sbjct: 245 VNFGFVSSLGPIKGWGLEIEKNSIVVNTKMETNIPGIYAAGDICTYPGKVKLIATGFGEA 304

Query: 304 PVAINQAINYIYPDRDNRVVHSTSL 328
           P A+N A  +I P       HSTSL
Sbjct: 305 PTAVNNAKAFIDPTARVFPGHSTSL 329
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6079> which encodes the amino acid sequence <SEQ ID 6080>. Analysis of this protein sequence reveals the following:

```
     Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.37   Transmembrane    8 -  24 (   8 -  24)

----- Final Results -----
               bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB15201 GB:Z99120 similar to thioredoxin reductase [Bacillus subtilis]
 Identities = 173/328 (52%), Positives = 223/328 (67%), Gaps = 4/328 (1%)

Query: 4   KAYDITIIGGGPIGLFAAFYAGLRGVTVKIIESLSELGGQPAILYPEKMIYDIPAYPSLT 63
           K YDITIIGGGP+GLF AFY G+R  +VKIIESL +LGGQ + LYPEK IYD+  +P +
Sbjct: 6   KVYDITIIGGGPVGLFTAFYGGMRQASVKIIESLPQLGGQLSALYPEKYIYDVAGFPKIR 65

Query: 64  GVELTENLIKQLSRFEDRTTICLKEEVLTFDKVKGG-FSIRTNKAEHFSKAIIIACGNGA 122
             EL  NL +Q+++F+   TICL++ V + +K   G F +   K    K  I  GNGA
Sbjct: 66  AQELINNLKEQMAKFDQ--TICLEQAVESVEKQADGVFKLVQMKKPTTLKRSCITAGNGA 123

Query: 123 FAPRTLGLESEENFADHNLFYNVHQLDQFAGQKVVICGGGDSAVDWALALEDIAESVTVV 182
           F PR L LE+ E +   NL Y V  L +FAG++V I GGGDSAVDWAL LE IA+ V+++
Sbjct: 124 FKPRKLELENAEQYEGKNLHYFVDDLQKFAGRRVAILGGGDSAVDWALMLEPIAKEVSII 183

Query: 183 HRRDAFRAHEHSVELLKASTVNLLTPYVPKALKGIGNLAEKLVIQKVKEDEVLELELDSL 242
           HRRD FRAHEHSVE L AS VN+LTP+VP  L G   + E+LV+++VK D    LE+D L
Sbjct: 184 HRRDKFRAHEHSVENLHASKVNVLTPFVPAELIGEDKI-EQLVLEEVKGDRKEILEIDDL 242

Query: 243 IVSFGFSTSNKNLKNWNLDYKRSSITVSPLFQTSQEGIFAIGDAAAYNGKVDLIATGFGE 302
           IV++GF +S   +KNW LD +++SI V    +T+ EG FA GD   Y GKV+LIA+GFGE
Sbjct: 243 IVNYGFVSSLGPIKNWGLDIEKNSIVVKSTMETNIEGFFAAGDICTYEGKVNLIASGFGE 302

Query: 303 APTAVNQAINYIYPDRDNRVVHSTSLID 330
           APTAVN A  Y+ P     + +HSTSL +
Sbjct: 303 APTAVNNAKAYMDPKARVQPLHSTSLFE 330
```

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

```
 Identities = 242/324 (74%), Positives = 279/324 (85%)

Query: 6   YDITIVGGGPVGLFAAFYAGLRGVSVKIIESLSELGGQPAILYPEKKIYDIPGYPVITGR 65
           YDITI+GGGP+GLFAAFYAGLRGV+VKIIESLSELGGQPAILYPEK IYDIP YP +TG
Sbjct: 6   YDITIIGGGPIGLFAAFYAGLRGVTVKIIESLSELGGQPAILYPEKMIYDIPAYPSLTGV 65

Query: 66  ELIDKHIEQLERFKDSIEICLKEEVLSFEKVDDVFTIQTDKDQHLSRAIVFACGNGAFAP 125
           EL +  I+QL RF+D   ICLKEEVL+F+KV   F+I+T+K +H S+AI+ ACGNGAFAP
Sbjct: 66  ELTENLIKQLSRFEDRTTICLKEEVLTFDKVKGGFSIRTNKAEHFSKAIIIACGNGAFAP 125

Query: 126 RLLGLENEENYADNNLFYNVTKLEQFAGKHVVICGGGDSAVDWANELDKIAASVAIVHRR 185
           R LGLE+EEN+AD+NLFYNV +L+QFAG+ VVICGGGDSAVDWA  L+ IA SV +VHRR
Sbjct: 126 RTLGLESEENFADHNLFYNVHQLDQFAGQKVVICGGGDSAVDWALALEDIAESVTVVHRR 185

Query: 186 DAFRAHEHSVDILKASGVRILTPYVPIGLNGDSQRVSSLVVQKVKGDEVIELPLDNLIVS 245
           DAFRAHEHSV++LKAS V +LTPYVP  L G        LV+QKVK DEV+EL LD+LIVS
Sbjct: 186 DAFRAHEHSVELLKASTVNLLTPYVPKALKGIGNLAEKLVIQKVKEDEVLELELDSLIVS 245

Query: 246 FGFSTSNKNLRYWNLDYKRSSINVSSLFETTQEGVYAIGDAANYPGKVELIATGYGEAPV 305
           FGFSTSNKNL+ WNLDYKRSSI VS LF+T+QEG++AIGDAA Y GKV+LIATG+GEAP
Sbjct: 246 FGFSTSNKNLKNWNLDYKRSSITVSPLFQTSQEGIFAIGDAAAYNGKVDLIATGFGEAPT 305

Query: 306 AINQAINYIYPDRDNRVVHSTSLI 329
           A+NQAINYIYPDRDNRVVHSTSLI
Sbjct: 306 AVNQAINYIYPDRDNRVVHSTSLI 329
```

SEQ ID 6078 (GBS178) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 5; MW 37.4kDa). 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 8; MW 62.4kDa).

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

Example 1962

A DNA sequence (GBSx2071) was identified in *S.agalactiae* <SEQ ID 6081> which encodes the amino acid sequence <SEQ ID 6082>. This protein is predicted to be tRNA methyltransferase (trmD). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1496(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:BAB06198 GB:AP001515 tRNA methyltransferase [Bacillus halodurans]
 Identities = 144/246 (58%), Positives = 186/246 (75%), Gaps = 6/246 (2%)

Query: 2   MKIDILTLFPEMFAPLEHS-IVGKAKERGLLEINYHNFRENAE-KSRHVDDEPYGGGQGM 59
           MKID LTLFPEMF  + HS I+ +A+ERG +     NFRE +E K + VDD PYGGG GM
Sbjct: 1   MKIDFLTLFPEMFQGVLHSSILKQAQERGAVSFRVVNFREYSENKHKKVDDYPYGGGAGM 60

Query: 60  LLRAQPIFDTIDKIDAQKA---RVILLDPAGRTFDQDFAEELSKEDELIFICGHYEGYDE 116
           +L  QP+FD ++ +  + +   RVIL+ P G TF Q  AEEL++ + LI +CGHYEGYDE
Sbjct: 61  VLSPQPLFDAVEDLTKKSSSTPRVILMCPQGETFTQRKAEELAQAEHLILLCGHYEGYDE 120

Query: 117 RIKS-LVTDEVSLGDFVLTGGELAAMTMVDATVRLIPEVIGKETSHQDDSFSSGLLEYPQ 175
           RI+S LVTDE+S+GD+VLTGGEL AM + D+  RL+P V+G ETS Q DSFS+GLLEYPQ
Sbjct: 121 RIRSYLVTDELSIGDYVLTGGELGAMVIADSVTRLLPAVLGNETSAQTDSFSTGLLEYPQ 180
```

```
Query: 176 YTRPYDYLGMTVPDVLMSGHHENIRKWRLEQSLRKTLERRPDLLENYAMTDEERLILEKI 235
           YTRP D+ G  VPDVL+SGHH+NI +WR EQSL++TLERRPDLLE   +T+EE+ +L+ I
Sbjct: 181 YTRPADFRGWKVPDVLLSGHHQNIERWRKEQSLKRTLERRPDLLEGRKLTEEEQELLDSI 240

Query: 236 KTEIER 241
           + + E+
Sbjct: 241 RKQQEK 246
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6083> which encodes the amino acid sequence <SEQ ID 6084>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2705(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 = 195/240 (81%), Positives = 224/240 (93%)

Query: 2   MKIDILTLFPEMFAPLEHSIVGKAKERGLLEINYHNFRENAEKSRHVDDEPYGGGQGMLL 61
           MKIDILTLFPEMFAPLEHSIVGKAKE+GLL+I+YHNFR+ AEK+RHVDDEPYGGGQGMLL
Sbjct: 1   MKIDILTLFPEMFAPLEHSIVGKAKEKGLLDIHYHNFRDYAEKARHVDDEPYGGGQGMLL 60

Query: 62  RAQPIFDTIDKIDAQKARVILLDPAGRTFDQDFAEELSKEDELIFICGHYEGYDERIKSL 121
           RAQPIFDTI++I+A+K R+ILLDPAG+ F Q +AEEL+ E+ELIFICGHYEGYDERIK+L
Sbjct: 61  RAQPIFDTIEQIEAKKPRIILLDPAGKPFTQAYAEELALEEELIFICGHYEGYDERIKTL 120

Query: 122 VTDEVSLGDFVLTGGELAAMTMVDATVRLIPEVIGKETSHQDDSFSSGLLEYPQYTRPYD 181
           VTDE+SLGDFVLTGGELAAMTMVDATVRLIP+V+GKE+SHQDDSFSSGLLEYPQYTRPYD
Sbjct: 121 VTDEISLGDFVLTGGELAAMTMVDATVRLIPQVLGKESSHQDDSFSSGLLEYPQYTRPYD 180

Query: 182 YLGMTVPDVLMSGHHENIRKWRLEQSLRKTLERRPDLLENYAMTDEERLILEKIKTEIER 241
           Y GMTVPDVLMSGHHE IR WRLE+SL+KT  RRPDLLE+Y  ++EER +L+KIK  +++
Sbjct: 181 YRGMTVPDVLMSGHHERIRLWRLEESLKKTYLRRPDLLEHYNFSEEERKLLDKIKEALDQ 240
```

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

Example 1963

A DNA sequence (GBSx2072) was identified in *S.agalactiae* <SEQ ID 6085> which encodes the amino acid sequence <SEQ ID 6086>. This protein is predicted to be 16S rRNA processing protein. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.71   Transmembrane   32 -  48 (  32 -  52)

----- Final Results -----
               bacterial membrane --- Certainty=0.2084(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 9381> which encodes amino acid sequence <SEQ ID 9382> was also identified.

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

```
>GP:CAB13475 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]
```

```
Identities = 88/174 (50%), Positives = 128/174 (72%), Gaps = 1/174 (0%)

Query: 54  VTMEYFNVGKIVNTQGLQGEMRVLSVTDFVEERFKKGQVLALFDEKNQFVMDIEIASHRK 113
           +T  +FNVGKIVNT G++GE+RV+S TDF EER+K G  L LF +     +++ + +HR
Sbjct: 1   MTKRWFNVGKIVNTHGIKGEVRVISKTDFAEERYKPGNTLYLFMDGRNEPVEVTVNTHRL 60

Query: 114 QKNFDIIKFKGMYHINDIEKYKGFTLKVAEDQLSDLKDGEFYYHEIIGLDVYEGE-ELIG 172
            K F +++FK   ++N++E+ K   +KV E++L +L +GEFY+HEIIG +V+  E ELIG
Sbjct: 61  HKQFHLLQFKERQNLNEVEELKNAIIKVPEEELGELNEGEFYFHEIIGCEVFTEEGELIG 120

Query: 173 KIKEILQPGANDVWVVERHGKRDLLLPYIPPVVLEVDLSNQRVQVELMEGLDDE 226
           K+KEIL PGANDVWV+ R GK+D L+PYI  VV  +D+  +++++ELMEGL DE
Sbjct: 121 KVKEILTPGANDVWVIGRKGKKDALIPYIESVVKHIDVREKKIEIELMEGLIDE 174
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6087> which encodes the amino acid sequence <SEQ ID 6088>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2787(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 = 133/172 (77%), Positives = 153/172 (88%)

Query: 56  MEYFNVGKIVNTQGLQGEMRVLSVTDFVEERFKKGQVLALFDEKNQFVMDIEIASHRKQK 115
           MEYFNVGKIVNTQGLQGEMRVLSV+DF EERFKKG  LALFD+K++FV ++ I SHRKQK
Sbjct: 1   MEYFNVGKIVNTQGLQGEMRVLSVSDFAEERFKKGSQLALFDDKDRFVQEVTIVSHRKQK 60

Query: 116 NFDIIKFKGMYHINDIEKYKGFTLKVAEDQLSDLKDGEFYYHEIIGLDVYEGEELIGKIK 175
           +FDIIKFK MYHIN IEKYKG+TLKV++D   DL++GEFYYH+IIG+ VYE + LIG +K
Sbjct: 61  HFDIIKFKDMYHINAIEKYKGYTLKVSKDNQGDLQEGEFYYHQIIGMAVYEKDVLIGHVK 120

Query: 176 EILQPGANDVWVVERHGKRDLLLPYIPPVVLEVDLSNQRVQVELMEGLDDED 227
           EILQPGANDVW+V+R GKRDLLLPYIPPVVL VD+ N+RV VELMEGLDDED
Sbjct: 121 EILQPGANDVWIVKRQGKRDLLLPYIPPVVLNVDVPNKRVDVELMEGLDDED 172
```

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

Example 1964

A DNA sequence (GBSx2073) was identified in *S.agalactiae* <SEQ ID 6089> which encodes the amino acid sequence <SEQ ID 6090>. This protein is predicted to be similar to E. coli ykfC (11). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3488(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 9909> which encodes amino acid sequence <SEQ ID 9910> was also identified.

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

```
>GP:AAC38715 GB:AF030367 maturase-related protein [Streptococcus pneumoniae]
 Identities = 366/425 (86%), Positives = 396/425 (93%)

Query: 12  MSELLDKILSRNNMLEAYKQVKSNKGSAGINGVTIEQMDDYLHQNWRETKQLIKERSYKP 71
           MS+LLDKILSR NMLEAY QVKSNKGSAGI+G+TIE+MD+YL QNWR TK+LIK+R YKP
Sbjct: 1   MSKLLDKILSRENMLEAYNQVKSNKGSAGIDGMTIEEMDNYLRQNWRLTKELIKQRKYKP 60

Query: 72  QPVLRVEIPKPNGGVRNLGIPTAMDRMIQQAIVQVLSPLCEKHFSEYSYGFRPNRSCETA 131
           QPVL+VEIPKP+GG+R LGIPT MDRMIQQAIVQV+SP+CE HFS+ SYGFRPNRSCE A
Sbjct: 61  QPVLKVEIPKPDGGIRQLGIPTVMDRMIQQAIVQVMSPICEPHFSDTSYGFRPNRSCEKA 120

Query: 132 IVQLLEYLNDGYEWIVDIDLEKFFDTVPQDRLMSLVHNIIQDGDTESLIRKYLHSGVVIN 191
           I++LLEYLNDGYEWIVDIDLEKFFDTVPQDRLMSLVHNII+DGDTESLIRKYLHSGV+IN
Sbjct: 121 IMKLLEYLNDGYEWIVDIDLEKFFDTVPQDRLMSLVHNIIEDGDTESLIRKYLHSGVIIN 180

Query: 192 GQRHKTLVGTPQGGNLSPLLSNIMLNELDKGLEKRGLRFVRYADDCVITVGSEAAAKRVM 251
           GQR+KTLVGTPQGGNLSPLLSNIMLNELDK LEKRGLRFVRYADDCVITVGSEAAAKRVM
Sbjct: 181 GQRYKTLVGTPQGGNLSPLLSNIMLNELDKELEKRGLRFVRYADDCVITVGSEAAAKRVM 240

Query: 252 HSVSSYIEKRLGLKVNMTKTKIVRPNKLKYLGFGFWKSPKGWKCRPHQDSVQSFKRKLKQ 311
           +SVS +IEKRLGLKVNMTKTKI RP +LKYLGFGFWKS  GWK RPHQDSV+ FK KLK+
Sbjct: 241 YSVSRFIEKRLGLKVNMTKTKITRPRELKYLGFGFWKSSDGWKSRPHQDSVRRFKLKLKK 300

Query: 312 LTMRKWSIDLITRIERLNWVIRGWINYFSLGNMKSIMTQIDERLRTRIRVIIWKQWKKKA 371
           LT RKWSIDL  RIE+LN  IRGWINYFSLGNMKSI+  IDERLRTR+R+IIWKQWKKK+
Sbjct: 301 LTQRKWSIDLTRRIEQLNLSIRGWINYFSLGNMKSIVASIDERLRTRLRMIIWKQWKKKS 360

Query: 372 KRLWGLLKLGVARWIADKVSGWGDHYQLVAQKSVLKRAISKPALAKRGLVSCLDYYLERH 431
           +RLWGLLKLGV +WIADKVSGWGDHYQLVAQKSVLKRAISKP L KRGLVSCLDYYLERH
Sbjct: 361 RRLWGLLKLGVPKWIADKVSGWGDHYQLVAQKSVLKRAISKPVLEKRGLVSCLDYYLERH 420

Query: 432 ALKVS 436
           ALKVS
Sbjct: 421 ALKVS 425
```

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 1965

A DNA sequence (GBSx2074) was identified in *S.agalactiae* <SEQ ID 6091> which encodes the amino acid sequence <SEQ ID 6092>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.37   Transmembrane    7 -  23 (   7 -  23)

----- 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 DNA sequence was identified in *S.pyogenes* <SEQ ID 821> which encodes the amino acid sequence <SEQ ID 822>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.87   Transmembrane 1157 -1173 (1157 -1174)

----- 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>
```

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

```
Identities = 1031/1064 (96%), Positives = 1042/1064 (97%)

Query: 1    MRKKQKLPFDKLAIALISTSILLNAQSDIKANTVTEDTPATEQAVEPPQPIAVSEESPSS 60
            +RKKQKLPFDKLAIAL+STSILLNAQSDIKANTVTEDTPATEQAVE PQP AVSEE+PSS
Sbjct: 1    LRKKQKLPFDKLAIALMSTSILLNAQSDIKANTVTEDTPATEQAVETPQPTAVSEEAPSS 60

Query: 61   KETKTSQTPSDVGETVADDANDLAPQAPAKTADTPATSKATIRDLNDPSHVKTLQEKAGK 120
            KETKT QTP D  ET+ADDANDLAPQAPAKTADTPATSKATIRDLNDPS VKTLQEKAGK
Sbjct: 61   KETKTPQTPDDAEETIADDANDLAPQAPAKTADTPATSKATIRDLNDPSQVKTLQEKAGK 120

Query: 121  GVGTVVAVIDAGFDKNHEAWRLTDKTKARYQSKENLEKAKKEHGITYGEWVNDKVAYYHD 180
            G GTVVAVIDAGFDKNHEAWRLTDKTKARYQSKE+LEKAKKEHGITYGEWVNDKVAYYHD
Sbjct: 121  GAGTVVAVIDAGFDKNHEAWRLTDKTKARYQSKEDLEKAKKEHGITYGEWVNDKVAYYHD 180

Query: 181  YSKDGKNAVDQEHGTHVSGILSGNAPSEMKEPYRLEGAMPEAQLLLMRVEIVNGLADYAR 240
            YSKDGK AVDQEHGTHVSGILSGNAPSE KEPYRLEGAMPEAQLLLMRVEIVNGLADYAR
Sbjct: 181  YSKDGKTAVDQEHGTHVSGILSGNAPSETKEPYRLEGAMPEAQLLLMRVEIVNGLADYAR 240

Query: 241  NYAQAIRDAVNLGAKVINMSFGNAALAYANLPDETKKAFDYAKSKGVSIVTSAGNDSSFG 300
            NYAQAI DAVNLGAKVINMSFGNAALAYANLPDETKKAFDYAKSKGVSIVTSAGNDSSFG
Sbjct: 241  NYAQAIIDAVNLGAKVINMSFGNAALAYANLPDETKKAFDYAKSKGVSIVTSAGNDSSFG 300

Query: 301  GKPRLPLADHPDYGVVGTPAAADSTLTVASYSPDKQLTETATVKTDDHQDKEMPVLSTNR 360
            GK RLPLADHPDYGVVGTPAAADSTLTVASYSPDKQLTETATVKT D QDKEMPVLSTNR
Sbjct: 301  GKTRLPLADHPDYGVVGTPAAADSTLTVASYSPDKQLTETATVKTADQQDKEMPVLSTNR 360

Query: 361  FEPNKAYDYAYANRGTKEDDFKDVEGKIALIERGDIDFKDKIANAKKAGAVGVLIYDNQD 420
            FEPNKAYDYAYANRG KEDDFKDV+GKIALIERGDIDFKDKIANAKKAGAVGVLIYDNQD
Sbjct: 361  FEPNKAYDYAYANRGMKEDDFKDVKGKIALIERGDIDFKDKIANAKKAGAVGVLIYDNQD 420

Query: 421  KGFPIELPNVDQMPAAFISRRDGLLLKDNPQKTITFNATPKVLPTASGTKLSRFSSWGLT 480
            KGFPIELPNVDQMPAAFISR+DGLLLK+NPQKTITFNATPKVLPTASGTKLSRFSSWGLT
Sbjct: 421  KGFPIELPNVDQMPAAFISRKDGLLLKENPQKTITFNATPKVLPTASGTKLSRFSSWGLT 480

Query: 481  ADGNIKPDIAAPGQDILSSVANNKYAKLSGTSMSAPLVAGIMGLLQKQYETQYPDMTPSE 540
            ADGNIKPDIAAPGQDILSSVANNKYAKLSGTSMSAPLVAGIMGLLQKQYETQYPDMTPSE
Sbjct: 481  ADGNIKPDIAAPGQDILSSVANNKYAKLSGTSMSAPLVAGIMGLLQKQYETQYPDMTPSE 540

Query: 541  RLDLAKKVLMSSATALYDEDEKAYFSPRQQGAGAVDAKKASAATMYVTDKDNTSSKVHLN 600
            RLDLAKKVLMSSATALYDEDEKAYFSPRQQGAGAVDAKKASAATMYVTDKDNTSSKVHLN
Sbjct: 541  RLDLAKKVLMSSATALYDEDEKAYFSPRQQGAGAVDAKKASAATMYVTDKDNTSSKVHLN 600

Query: 601  NVSDKFEVTVTVHNKSDKPQELYYQVTVQTDKVDGKHFALAPKALYETSWQKITIPANSS 660
            NVSDKFEVTVTVHNKSDKPQELYYQ TVQTDKVDGK FALAPKALYETSWQKITIPANSS
Sbjct: 601  NVSDKFEVTVTVHNKSDKPQELYYQATVQTDKVDGKLFALAPKALYETSWQKITIPANSS 660

Query: 661  KQVTVPIDASRFSKDLLAQMKNGYFLEGFVRFKQDPTKEELMSIPYIGFRGDFGNLSALE 720
            KQVT+PID S+FSKDLLA MKNGYFLEGFVRFKQDPTKEELMSIPYIGFRGDFGNLSALE
Sbjct: 661  KQVTIPIDVSQFSKDLLAPMKNGYFLEGFVRFKQDPTKEELMSIPYIGFRGDFGNLSALE 720

Query: 721  KPIYDSKDGSSYYHEANSDAKDQLDGDGLQFYALKNNFTALTTESNPWTIIKAVKEGVEN 780
            KPIYDSKDGSSYYHEANSDAKDQLDGDGLQFYALKNNFTALTTESNPWTIIKAVKEGVEN
Sbjct: 721  KPIYDSKDGSSYYHEANSDAKDQLDGDGLQFYALKNNFTALTTESNPWTIIKAVKEGVEN 780

Query: 781  IEDIESSEITETIFAGTFAKQDDDSHYYIHRHANGKPYAAISPNGDGNRDYVQFQGTFLR 840
            IEDIESSEITETIFAGTFAKQDDDSHYYIHRHANGKPYAAISPNGDGNRDYVQFQGTFLR
Sbjct: 781  IEDIESSEITETIFAGTFAKQDDDSHYYIHRHANGKPYAAISPNGDGNRDYVQFQGTFLR 840

Query: 841  NAKNLVAEVLDKEGNVVWTSEVTEQVVKNYNNDLASTLGSTRFEKTRWDGKNKDGKVVAN 900
            NAKNLVAEVLDKEGNVVWTSEVTEQVVKNYNNDLASTLGSTRFEKTRWDGK+KDGKVVAN
Sbjct: 841  NAKNLVAEVLDKEGNVVWTSEVTEQVVKNYNNDLASTLGSTRFEKTRWDGKDKDGKVVAN 900

Query: 901  GTYTYRVRYTPISSGAKEQHTDFDVIVDNTTPEVATSATFSTEDSRLTLASKPKTSQPVY 960
            GTYTYRVRYTPISSGAKEQHTDFDVIVDNTTPEVATSATFSTED RLTLASKPKTSQPVY
Sbjct: 901  GTYTYRVRYTPISSGAKEQHTDFDVIVDNTTPEVATSATFSTEDRRLTLASKPKTSQPVY 960
```

```
Query: 961  RERIAYTYMDEDLPTTEYISPNEDGTFTLPEEAETMEGATVPLKMSDFTYVVEDMAGNIT 1020
            RERIAYTYMDEDLPTTEYISPNEDGTFTLPEEAETMEGATVPLKMSDFTYVVEDMAGNIT
Sbjct: 961  RERIAYTYMDEDLPTTEYISPNEDGTFTLPEEAETMEGATVPLKMSDFTYVVEDMAGNIT 1020

Query: 1021 YTPVTKLLEGHSNKPEQDGSDQAPDKKPEAKPEQDGSGQTPDKK 1064
            YTPVTKLLEGHSNKPEQDGSDQAPDKKPE KPEQDGSGQ PDKK
Sbjct: 1021 YTPVTKLLEGHSNKPEQDGSDQAPDKKPETKPEQDGSGQAPDKK 1064
```

A related GBS gene <SEQ ID 8941> and protein <SEQ ID 8942> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 10
McG: Discrim Score:      5.69
GvH: Signal Score (-7.5): -3.33
     Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -0.37 threshold:  0.0
   INTEGRAL    Likelihood = -0.37   Transmembrane    7 -  23 (   7 -  23)
   PERIPHERAL  Likelihood =  2.81     508
 modified ALOM score:   0.57

*** Reasoning Step: 3

----- 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>
```

SEQ ID 8942 (GBS276) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 2; MW 123kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 5; MW 46.5kDa).

The GBS276-His fusion product was purified (Figure 206, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 296), 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 1966

A DNA sequence (GBSx2075) was identified in *S.agalactiae* <SEQ ID 6093> which encodes the amino acid sequence <SEQ ID 6094>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4286(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 1967

A DNA sequence (GBSx2076) was identified in *S.agalactiae* <SEQ ID 6095> which encodes the amino acid sequence <SEQ ID 6096>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.15   Transmembrane   19 -  35 (  11 -  39)

----- 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 9911> which encodes amino acid sequence <SEQ ID 9912> 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 6096 (GBS654) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 142 (lane 8 & 10; MW 51.2kDa + lane 9; MW 27kDa). Purified GBS654-GST is shown in Figure 245, lane 11.

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

Example 1968

A DNA sequence (GBSx2077) was identified in *S.agalactiae* <SEQ ID 6097> which encodes the amino acid sequence <SEQ ID 6098>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4174(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 9913> which encodes amino acid sequence <SEQ ID 9914> was also identified.

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

```
>GP:AAF27324 GB:AF178424 unknown [Lactococcus lactis]
 Identities = 26/75 (34%), Positives = 45/75 (59%), Gaps = 4/75 (5%)

Query: 11 MAFEPKNSELTKVLKES-LDEEKKEIFSSEMNIRDFERTKQYQFTLQPSVRKKIDRLSKE 69
          MAF+  + ++  VL  S L + K E+      I   E  K Y FTL+PSV++ +++L+++
Sbjct: 1  MAFDVDDKKVKTVLSNSSLAKSKVEL---PKKIESEENKKSYSFTLEPSVKEGLEKLAEK 57

Query: 70 KGYRSASSFINDFFK 84
          + Y++ S F+ND  K
Sbjct: 58 QNYKNTSQFLNDLIK 72
```

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 1969

A DNA sequence (GBSx2078) was identified in *S.agalactiae* <SEQ ID 6099> which encodes the amino acid sequence <SEQ ID 6100>. This protein is predicted to be ParA. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> 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:AAF27325 GB:AF178424 ParA [Lactococcus lactis]
 Identities = 49/104 (47%), Positives = 72/104 (69%)

Query: 22  LSERLEEFKTEAFDFKTRASYVTAKLFFLGNMIKHNTNSSKELIRSLKNDKSVLAMIPHK 81
           L ERL+ FK E  D +TR +Y+TA  +F+GN I+HNT SS+E    +  DK  +AMIP K
Sbjct: 157 LIERLQNFKDEVIDARTRETYITAIPYFVGNRIRHNTKSSREFSEKISQDKGTIAMIPEK 216

Query: 82  ELFNRSTLDKKSLSYMMSDKELYSRDSKFFKEIDFTFRKITDKL 125
           ELFNRSTLD   L  M  DK++++ +  F+++++F F +IT+K+
Sbjct: 217 ELFNRSTLDGVPLVEMEKDKDVFNSNKVFYEKLNFAFNEITNKI 260
```

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 1970

A DNA sequence (GBSx2079) was identified in *S.agalactiae* <SEQ ID 6101> which encodes the amino acid sequence <SEQ ID 6102>. This protein is predicted to be transposase (orfA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2830(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

Example 1971

A DNA sequence (GBSx2080) was identified in *S.agalactiae* <SEQ ID 6103> which encodes the amino acid sequence <SEQ ID 6104>. This protein is predicted to be transposase (orfB). Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2618(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB90834 GB:AJ250837 putative transposase [Streptococcus dysgalactiae]
 Identities = 242/259 (93%), Positives = 249/259 (95%)

Query: 1   MCRWLNMPHSSYYYQAVESVSETEFEETIKRIFLDSESRYGSRKIKICLNNEGITLSRRR 60
           MCRWLN+P SSYYY+AVE VSE E EE+IK IFL+S++RYGSRKIKICLNNEGITLSRRR
Sbjct: 1   MCRWLNIPRSSYYYKAVEPVSEAELEESIKAIFLESKARYGSRKIKICLNNEGITLSRRR 60

Query: 61  IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFKQERPLQALVTDLTYVRVGNR 120
           IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFK ERPLQALVTDLTYVRVGNR
Sbjct: 61  IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFKPERPLQALVTDLTYVRVGNR 120

Query: 121 WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPYALTKVKMFHSDRKKEFDNQLID 180
           WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPY LTKVKMFHSDR KEF+NQLID
Sbjct: 121 WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPYPLTKVKMFHSDRGKEFNNQLID 180

Query: 181 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFQLLEELALKTKDYVHWWNY 240
           EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFQ LEELALKTK YVHWWNY
Sbjct: 181 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFQSLEELALKTKAYVHWWNY 240

Query: 241 HRIHGSLNYQTPMTKRLIA 259
           HRIHGSLNYQTPMTKRLIA
Sbjct: 241 HRIHGSLNYQTPMTKRLIA 259
```

There is also homology to SEQ ID 32.

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

Example 1972

A DNA sequence (GBSx2081) was identified in *S.agalactiae* <SEQ ID 6105> which encodes the amino acid sequence <SEQ ID 6106>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3325(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 1973

A DNA sequence (GBSx2082) was identified in *S.agalactiae* <SEQ ID 6107> which encodes the amino acid sequence <SEQ ID 6108>. Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4442(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 9917> which encodes amino acid sequence <SEQ ID 9918> was also identified.

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

```
>GP:AAD44095 GB:AF115103 orf359 gp [Streptococcus thermophilus
          bacteriophage Sfi21]
 Identities = 92/357 (25%), Positives = 162/357 (44%), Gaps = 33/357 (9%)

Query: 45  RKNQYGKTFETMKEAYDELVRIKYEFANKVSLENYNMTFENYMNKIYLRAYKQK-VQSVT 103
           RK +    F T  EA      ++ +  + V+++  ++T  +Y  K +   YK+  V  +T
Sbjct: 24  RKPKTKGGFRTKSEAIKAAAEMELKLQDNVNVDE-DITLYDYF-KQWCEVYKKPTVSKIT 81

Query: 104 YKTALPHHKLFIQYFGLKPLKAITPRDCEAFRLHIIENYSENYAKNLWSRF----KACMG 159
           YK  +   +    +FG K LK+IT  + +     ++ +Y++ +A++   RF    KAC+
Sbjct: 82  YKAYINSQRKIELFFGDKKLKSITATEYQ----RVLNSYAKTHAQDTVERFNVHVKACIE 137

Query: 160 YAERLGYISNMPCKALD---NPRGKHPETPFWTYAEFQTFIKSFDLHDYEELQRFTAIWL 216
            A   GYI    CK        +G+  ET F    E++  I  ++  + E   + A+++
Sbjct: 138 MAVHEGYIKRNFCKFAKINAKNKGRDIETKFLEVEEYERLI--YETSKHPEYASYAALYI 195

Query: 217 YYMTGVRVSEGLSLCWEDIDFDKKFLKVHTTLEKDENGNWYRKDQTKTPAGERLIELDDI 276
              TG+R +E L L  +DI  D   L V+ T +   N  +     TKT +  R I LDD
Sbjct: 196 IAKTGIRFAECLGLTVDDIKRDTGMLSVNKTWDYKNNTGFM---PTKTKSSIREIPLDDE 252

Query: 277 TIEVLQVWRKNQFANQDTDFIISRFGDPFCKSTICRIIKRKAQQVGVPVITGKGLRHSHA 336
           I  +     +Q    D  I+    +  T+ +I+ R+        +    LRH++A
Sbjct: 253 FINFI-----DQLPPTDDGRILPSLSNNAVNKTLRKIVGRE--------VRVHSLRHTYA 299

Query: 337 SYLINVLKKDILYVARRMGHADKSTTLNTYSHWFNALDKTVSEEITQNIKSAGLDSI 393
           SYLI     D++ V++ +GH + + TL  Y+H          E+I Q     G +++
Sbjct: 300 SYLI-AHDIDLISVSQVLGHENLNITLEVYAHQLQEQKSRNDEKIKQMWTECGRNAL 355
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6109> which encodes the amino acid sequence <SEQ ID 6110>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5549(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 = 111/127 (87%), Positives = 119/127 (93%)

Query: 242 LKVHTTLEKDENGNWYRKDQTKTPAGERLIELDDITIEVLQVWRKNQFANQDTDFIISRF 301
           LKVHTTLEKDENGNWYRKDQTKTPAGERLIELDD+TI VL+ WR+NQ  N DTDFIISRF
Sbjct: 1   LKVHTTLEKDENGNWYRKDQTKTPAGERLIELDDVTIVVLENWRRNQVVNTDTDFIISRF 60

Query: 302 GDPFCKSTICRIIKRKAQQVGVPVITGKGLRHSHASYLINVLKKDILYVARRMGHADKST 361
           G+PFCKSTICR+IK KAQ +GVPVITGKGLRHS+ASYLINVLKKDILYVA+ MGHADKST
Sbjct: 61  GEPFCKSTICRVIKHKAQSIGVPVITGKGLRHSYASYLINVLKKDILYVAKCMGHADKST 120

Query: 362 TLNTYSH 368
           TLNTYSH
Sbjct: 121 TLNTYSH 127
```

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

Example 1974

A DNA sequence (GBSx2083) was identified in *S.agalactiae* <SEQ ID 6111> which encodes the amino acid sequence <SEQ ID 6112>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

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

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

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 1975

A DNA sequence (GBSx2084) was identified in *S.agalactiae* <SEQ ID 6113> which encodes the amino acid sequence <SEQ ID 6114>. This protein is predicted to be repressor protein-related 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=0.2721(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 9919> which encodes amino acid sequence <SEQ ID 9920> was also identified.

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

```
>GP:AAC98432 GB:L29324 repressor protein [Streptococcus pneumoniae]
 Identities = 38/65 (58%), Positives = 52/65 (79%), Gaps = 1/65 (1%)

Query: 2  MYRRLRDLREDNDFTQKYVAEK-LSFTHSAYSKIERGERILSADVIIKLSNLYNVSTDYL 60
          M +R+RDLRED+D TQ+YVA+  L+ T SAYSK+E G R++S D +IKL++ YNVS DYL
Sbjct: 1  MLKRIRDLREDDDLTQEYVAKTILNCTRSAYSKMESGTRLISIDDLIKLADFYNVSLDYL 60

Query: 61 LGQTD 65
          +G+ D
Sbjct: 61 VGRVD 65
```

There is also homology to SEQ ID 582.

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

Example 1976

A DNA sequence (GBSx2085) was identified in *S.agalactiae* <SEQ ID 6115> which encodes the amino acid sequence <SEQ ID 6116>. This protein is predicted to be relaxase. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3160(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:AAC98434 GB:L29324 relaxase [Streptococcus pneumoniae]
 Identities = 223/417 (53%), Positives = 310/417 (73%), Gaps = 5/417 (1%)

Query: 1   MVITKHYAVHGKKYRRQLIKYILDPKKTRNLSLISDFGMSNYLDFPDYVELVKMYQNNFL 60
           MVITKH+A+HGK YR +LIKYIL+P KT+NL+L+SDFGM NYLDFP Y ELVKMY +NFL
Sbjct: 1   MVITKHFAIHGKNYRSKLIKYILNPSKTKNLTLVSDFGMRNYLDFPSYKELVKMYNDNFL 60

Query: 61  SNDQLYDSRFDRQEKKQQKIHAHHIIQSFSPEDKLSPEEINRIGYETIKELIGGQYKFIV 120
           SND LY+ R DRQE  Q+KIH+HHIIQSFSP+D L+PE+INRIGYE  KEL GG+++FIV
Sbjct: 61  SNDTLYEFRHDRQEVNQRKIHSHHIIQSFSPDDHLTPEQINRIGYEAAKELTGGRFRFIV 120

Query: 121 ATHVDQDHCHNHIIINSINSQSQKKLKWDYALERNLQMISDRISKVAGAKIIPPKRYSHR 180
           ATHVD+ H HNHII+NSI+  S KK  WDY  E NL+M+SDR+SK+AGAKII   RYSHR
Sbjct: 121 ATHVDKGHIHNHIILNSIDQNSDKKFLWDYKAEHNLRMVSDRLSKIAGAKII-ENRYSHR 179

Query: 181 DYEVYRRSNHKYELKQRLFFLMEHSIDFNDFMQKAEQLNVKIDFSRKHSRFFMTDRNMKQ 240
            YEVYR++N+KYE+KQR++FL+E+S +F D  +KA+ L++KIDF  KH  +FMTD NMKQ
Sbjct: 180 QYEVYRKTNYKYEIKQRVYFLIENSKNFEDLKKKAKALHLKIDFRHKHVTYFMTDSNMKQ 239

Query: 241 VIQGDKLNKREPYSKEYFQRYFAKKKIELILEFLLLRSNSFDDLVEKARLLGLELKSKKK 300
           V++  KL++++PY++ YF++ F +++I  ILEFLL +  + ++L+++A + GL++  K+K
Sbjct: 240 VVRDSKLSRKQPYNETYFEKKFVQREIINILEFLLPKMKNMNELIQRAEVFGLKIIPKEK 299

Query: 301 TIDFVLSDGKSCISIPNKSLRKKNLYDTTYFDSYFKEHDVFEVLHNNEVKIEFEKFETQQ 360
            + F   DG   I +  + L K NLY  +YF  YF   +   VL N  +   + + +  +
Sbjct: 300 HVLFEF-DG---IKLAEQELVKSNLYSVSYFQDYFNNKNETFVLDNKNLVELYNEEKIIK 355

Query: 361 LSEILTVEEITEAYETYKTKRDAVHEFEVEITEEQIEKIVLDGLFVKVWMGIGQEGL 417
            E+ + E + ++Y+ +K  RDAVHEFEVE+   QIE++V  G+++KV  GI ++ L
Sbjct: 356 EKELPSEEMVWKSYQDFKRNRDAVHEFEVELNLNQIEEVVEHGIYIKVQFGIDKKDL 412
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6117> which encodes the amino acid sequence <SEQ ID 6118>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3114(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 = 103/218 (47%), Positives = 170/218 (77%)

Query: 393 EEQIEKIVLDGLFVKVWMGIGQEGLIFIPNHQLNILEQENKKQYQVFIRETSSYFIYHKE 452
           E QIE+++ + +++KV   + Q GLIFIPN+QL+I ++EN K+Y+V+IRET+ +FIY+KE
Sbjct: 2   EHQIERLIAEDIYIKVSFSVKQSGLIFIPNYQLDIRKEENHKKYKVYIRETAQFFIYNKE 61
```

```
Query: 453 DSEMNRFMKGRDLIRQLTFDNKSLPYKRRISLVSLQQKIEEINLLMTLNIQNKSFLELKD 512
           SE+NR+M+G +LI QLT D+KS+P +RR ++ +L++KIEEI+LL+ L+ +NK + ++KD
Sbjct: 62  ASELNRYMRGHELICQLTNDSKSIPKRRRQTIDTLKKKIEEISLLIELDTENKPYQDIKD 121

Query: 513 ELVGDIAQLDIELTNLQDKNTTLNKMAEVVVNLQSDNQDTKQLAKYECSKMNLSQNVTIG 572
           ++V D+AQLD+ +T LQD    LNK+AEV++NL +++ + ++LA+Y+ +KMNL+  + I
Sbjct: 122 DIVKDMAQLDLTITELQDHIAHLNKVAEVLLNLNNNDIENRRLARYDYAKMNLTAAIKIE 181

Query: 573 QIESEIEMIQNQLDNKIEEYENAVRKLDEYVRVLNMDK 610
           ++E EIE  QN+L+  I+EYE  VR+L+++  +L+  K
Sbjct: 182 EVEKEIETSQNELNISIDEYEYLVRRLEKFGEILSDSK 219
```

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

Example 1977

A DNA sequence (GBSx2086) was identified in *S.agalactiae* <SEQ ID 6119> which encodes the amino acid sequence <SEQ ID 6120>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4006(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:AAC98436 GB:L29324 unknown [Streptococcus pneumoniae]
 Identities = 53/115 (46%), Positives = 77/115 (66%), Gaps = 2/115 (1%)

Query: 5   VREIRKEVNFSIEEYQQIQNFMEQEGYEQFSPFARGKLLKIDHQPSQQLEEWIKYLQHQK 64
           +R IRK+   +  E +QI + M ++G + FS F R  LL  D Q  +Q+E+W    + QK
Sbjct: 5   IRSIRKQFRLTETEEKQILDLMREKGDDNFSDFLRKSLLLSDGQ--KQMEKWFNLWKKQK 62

Query: 65  VEQIYRDVHEILVLAKLSQSVTMEHLEIILTCIKDLMKEIEVTIPLSYSFKDKYM 119
           +EQI RDVHE+ ++AK +  VT EH+ I+LTCI++L+KE+E T PLS  F +KYM
Sbjct: 63  LEQISRDVHEVFIIAKTNHQVTHEHVSILLTCIQELIKEVEKTGPLSEDFCNKYM 117
```

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 1978

A DNA sequence (GBSx2087) was identified in *S.agalactiae* <SEQ ID 6121> which encodes the amino acid sequence <SEQ ID 6122>. This protein is predicted to be TnpA. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2935(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAC82523 GB:AF027768 TnpA [Serratia marcescens]
```

```
 Identities = 176/413 (42%), Positives = 243/413 (58%), Gaps = 18/413 (4%)

Query: 26  MMFKVEAVGPPERCPECGFD-KLYKHSSRNQLIMDLPIRLKRVGLHLNRRRYKCRECGST 84
           M F+V+ V   P  C ECG   +  +   R+    DLPI  KRV L + RRRY CR C +T
Sbjct: 1   MHFQVD-VPDPIACEECGVQGEFVRFGKRDVPYRDLPIHGKRVTLWVVRRRYTCRACKTT 59

Query: 85  IS------VDEKRSMTKRLLKSIQEQSMSKTFVEVAESVGVDEKTIRNVFKDYVALKERE 138
                 VD  R MT RL + ++++S +  +  VA   G+DEKT+R++F        R
Sbjct: 60  FRPQLPEMVDGFR-MTLRLHEYVEKESFNHPYTFVAAQTGLDEKTVRDIFNARAEFLGRW 118

Query: 139 YQFETPKWLGIDEIHIIRRPRLVLTNIERRTIYDIKPNRNKETVIQRLSEISDRTYIEYV 198
           ++FETP+ LGIDE+++ +R R +LTNIE RT+ D+   R ++ V   L ++ DR  +E V
Sbjct: 119 HRFETPRILGIDELYLNKRYRCILTNIEERTLLDLLATRRQDVVTNYLMKLKDRQKVEIV 178

Query: 199 TMDMWKPYKDAVNTILPQAKVVVDKFHVVRMANQALDNVRKSLKAHMSQKERRTLMRERF 258
           +MDMW PY+ AV  +LPQA++VVDKFHVVRMAN AL+ VRK L+  +   + RTL  +R
Sbjct: 179 SMDMWNPYRAAVKAVLPQARIVVDKFHVVRMANDALERVRKGLRKELKPSQSRTLKGDRK 238

Query: 259 ILLKRKHDLNERESFLLDTWLGNLPALKEAYELKEEFYWIWDTPDPDEGHLRYSQWRHRC 318
           ILLKR H++++RE  +++TW G  P L  AYE KE FY IWD     +      +W
Sbjct: 239 ILLKRAHEVSDRERLIMETWTGAFPQLLAAYEHKERFYGIWDATTRLQAEAALDEW-IAT 297

Query: 319 MSSNSKDAYKDLVRAVDNWHVEIFNYF--DKRLTNAYTESINSIIRQVERMGRGYSFDAL 376
           +    K+ + DLVRAV NW  E   YF  D  +TNAYTESIN + +   R GRGYSF+ +
Sbjct: 298 IPKGQKEVWSDLVRAVGNWREETMTYFETDMPVTNAYTESINRLAKDKNREGRGYSFEVM 357

Query: 377 RAKILFNEKLHKKRKPRFNSSAFNKAMLYDTFNWYEVNDHDITDNLGVDFSTL 429
           RA++L+  K HKK+ P    S F K  +      Y + D     N GVD ST+
Sbjct: 358 RARMLYTTK-HKKKAPTAKVSPFYKKTI-----GYGLPDFAEELNYGVDLSTI 404
```

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 1979

A DNA sequence (GBSx2088) was identified in *S.agalactiae* <SEQ ID 6123> which encodes the amino acid sequence <SEQ ID 6124>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2115(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]
 Identities = 412/546 (75%), Positives = 484/546 (88%)

Query: 1   MNKFKVNISGMTCTGCEKHVESALEKIGAKNIESSYRRGEAVFELPDDIEVESAIKAIDE 60
           M K++V++ GMTCTGCE+HV  ALE +GA  IE  +RRGEAVFELP+ + VE+A KAI +
Sbjct: 1   MKKYRVDVQGMTCTGCEEHVAVALENMGATGIEVDFRRGEAVFELPNALGVETAKKAISD 60

Query: 61  ANYQAGEIEEVSSLENVALINEDNYDLLIIGSGAAAFSSAIKAIEYGAKVGMIERGTVGG 120
           A YQ G+ EEV S E V L NE +YD +IIGSG AAFSSAI+A++YGAKV MIERGT+GG
Sbjct: 61  AKYQPGKAEEVQSQEMVQLGNEGDYDYIIIGSGGAAFSSAIEAVKYGAKVAMIERGTIGG 120

Query: 121 TCVNIGCVPSKTLLRAGEINHLSKDNPFIGLQTSAGEVDLASLITQKDKLVSELRNQKYM 180
           TCVNIGCVPSKTLLRAGEINHL+K+NPF+GL TSAGEVDLA LI QK++LV+ELRN KY+
Sbjct: 121 TCVNIGCVPSKTLLRAGEINHLAKNNPFVGLHTSAGEVDLAPLIKQKNELVTELRNSKYV 180
```

```
Query: 181 DLIDEYNFDLIKGEAKFVDASTVEVNGTKLSAKRFLIATGASPSLPQISGLEKMDYLTST 240
           DLID+Y F+LI+GEAKFVD  TVEVNG  +SAKRFLIATGASP+ P I GL ++DYLTST
Sbjct: 181 DLIDDYGFELIEGEAKFVDEKTVEVNGAPISAKRFLIATGASPAKPNIPGLNEVDYLTST 240

Query: 241 TLLELKKIPKRLTVIGSGYIGMELGQLFHHLGSEITLMQRSERLLKEYDPEISESVEKAL 300
           +LLELKK+PKRL VIGSGYIGMELGQLFH+LGSE+TL+QRSERLLKEYDPEISESVEK+L
Sbjct: 241 SLLELKKVPKRLVVIGSGYIGMELGQLFHNLGSEVTLIQRSERLLKEYDPEISESVEKSL 300

Query: 301 IEQGINLVKGATFERVEQSGEIKRVYVTVNGSREVIESDQLLVATGRKPNTDSLNLSAAG 360
           +EQGINLVKGAT+ER+EQ+G+IK+V+V VNG + +IE+DQLLVATGR PNT +LNL AAG
Sbjct: 301 VEQGINLVKGATYERIEQNGDIKKVHVEVNGKKRIIEADQLLVATGRTPNTATLNLRAAG 360

Query: 361 VETGKNNEILINDFGQTSNEKIYAAGDVTLGPQFVYVAAYEGGIITDNAIGGLNKKIDLS 420
           VE G   EI+I+D+ +T+N +IYAAGDVTLGPQFVYVAAY+GG+   NAIGGLNKK++L
Sbjct: 361 VEIGSRGEIIIDDYSRTTNTRIYAAGDVTLGPQFVYVAAYQGGVAAPNAIGGLNKKLNLE 420

Query: 421 VVPAVTFTNPTVATVGLTEEQAKEKGYDVKTSVLPLDAVPRAIVNRETTGVFKLVADAET 480
           VVP VTFT P +ATVGLTE+QAKE GY+VKTSVLPLDAVPRA+VNRETTGVFKLVAD++T
Sbjct: 421 VVPGVTFTAPAIATVGLTEQQAKENGYEVKTSVLPLDAVPRALVNRETTGVFKLVADSKT 480

Query: 481 LKVLGVHIVSENAGDVIYAASLAVKFGLTIEDLTETLAPYLTMAEGLKLVALTFDKDISK 540
           +KVLG H+V+ENAGDVIYAA+LAVKFGLT++D+ ETLAPYLTMAEGLKL ALTFDKDISK
Sbjct: 481 MKVLGAHVVAENAGDVIYAATLAVKFGLTVDDIRETLAPYLTMAEGLKLAALTFDKDISK 540

Query: 541 LSCCAG 546
           LSCCAG
Sbjct: 541 LSCCAG 546
```

There is also homology to SEQ ID 1820.

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

Example 1980

A DNA sequence (GBSx2089) was identified in *S.agalactiae* <SEQ ID 6125> which encodes the amino acid sequence <SEQ ID 6126>. This protein is predicted to be regulatory protein. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4529(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAA83973 GB:AF138877 mercury resistance operon negative
           regulator MerR1 [Bacillus sp. RC607]
 Identities = 83/129 (64%), Positives = 104/129 (80%)

Query: 1   MIYRISEFADKCGVNKETIRYYERKNLLQEPHRTEAGYRIYSYDDVKRVGFIKRIQEFGF 60
           M +RI E ADKCGVNKETIRYYER  L+ EP RTE GYR+YS   V R+ FIKR+QE GF
Sbjct: 1   MKFRIGELADKCGVNKETIRYYERLGLIPEPERTEKGYRMYSQQTVDRLHFIKRMQELGF 60

Query: 61  SLSEIYKLLGVVDKDEVRCQDMFEFVSKKQKEVQKQIEDLKRIETMLDDLKQRCPDEKKL 120
           +L+EI KLLGVVD+DE +C+DM++F   K +++Q++IEDLKRIE ML DLK+RCP+ K +
Sbjct: 61  TLNEIDKLLGVVDRDEAKCRDMYDFTILKIEDIQRKIEDLKRIERMLMDLKERCPENKDI 120

Query: 121 HSCPIIETL 129
           + CPIIETL
Sbjct: 121 YECPIIETL 129
```

There is also homology to SEQ ID 1712.

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

Example 1981

A DNA sequence (GBSx2090) was identified in *S.agalactiae* <SEQ ID 6127> which encodes the amino acid sequence <SEQ ID 6128>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.86   Transmembrane   80 -  96 (  78 - 100)

----- Final Results -----
               bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

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

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

A related GBS gene <SEQ ID 8943> and protein <SEQ ID 8944> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:    -13.52
GvH: Signal Score (-7.5): -6.14
     Possible site: 44
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -7.86 threshold:  0.0
   INTEGRAL    Likelihood = -7.86   Transmembrane   80 -  96 (  78 - 100)
   PERIPHERAL  Likelihood =  1.80     136
 modified ALOM score:   2.07

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
ORF02021(439 - 666 of 1080)
GP|451734|gb|AAA18975.1||U05143(9   -   46   of   46)   envelope   glycoprotein   (Simian
immunodeficiency virus)  GP|451744|gb|AAA18980.1||U05148  envelope  glycoprotein  (Simian
immunodeficiency virus)
%Match = 3.2
%Identity = 38.5  %Similarity = 64.1
Matches = 15  Mismatches = 13  Conservative Sub.s = 10

336       366       396       426       456       486       516       546
RIPVQFKGCDDYYNENVGYPLSRINLEHYLTEGGVLYFVVYSKDVSPTVTYASLTPKVIKNVLPASDKKKRIKKKEDIFL
                                          :|| |  : ||:|::||:  |:
                                    WGLTGNAGTTPTATTTTTTPRVVENVINESN----------
                                           10        20        30
```

```
576       606       636       666       696       726       756       786
LFWMAIIAKLLILPYPALQTSYKSRPCLRRSSLRKLTQIPFSIVTKVGNTNMKSITAFLQVKAYILPCLAKGPARIMV*W
                            ||:: :|   | | |
------------------------PCIKDNSCAGLEQEP
                                  40
```

SEQ ID 8944 (GBS415) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 3; MW 21.2kDa).

Example 1982

A DNA sequence (GBSx2092) was identified in *S.agalactiae* <SEQ ID 6129> which encodes the amino acid sequence <SEQ ID 6130>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3402(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 1983

A DNA sequence (GBSx2093) was identified in *S.agalactiae* <SEQ ID 6131> which encodes the amino acid sequence <SEQ ID 6132>. This protein is predicted to be ATPase. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.08   Transmembrane  324 - 340 ( 317 - 343)
   INTEGRAL    Likelihood = -5.73   Transmembrane  662 - 678 ( 660 - 690)
   INTEGRAL    Likelihood = -5.41   Transmembrane  350 - 366 ( 346 - 378)
   INTEGRAL    Likelihood = -3.40   Transmembrane   94 - 110 (  93 - 110)
   INTEGRAL    Likelihood = -2.87   Transmembrane  681 - 697 ( 680 - 699)
   INTEGRAL    Likelihood = -1.38   Transmembrane  148 - 164 ( 148 - 164)

----- Final Results -----
               bacterial membrane --- Certainty=0.5034(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAA22858 GB:M90750 cadmium-efflux ATPase [Bacillus firmus]
 Identities = 486/725 (67%), Positives = 584/725 (80%), Gaps = 18/725 (2%)

Query: 1   MSRGKAKQSEKEMKAYRVQGFTCTNCAAIFENNVKELPGVQDAKVNFGASKVYVKGTTTI 60
            MS  KA  SE+EMKAYRVQGFTC NCA  FE NVK+L GV+DAKVNFGASK+ V G  TI
Sbjct: 1   MSDQKAITSEQEMKAYRVQGFTCANCAGKFEKNVKQLSGVEDAKVNFGASKIAVYGNATI 60

Query: 61  EELEKAGAFENLKIRDEKEQRVGGE-----------PFWKQKENIKVYISALLLVVSWFL 109
            EELEKAGAFENLK+  EK  R   +           PF+K K +  +Y S LL+   +
Sbjct: 61  EELEKAGAFENLKVTPEKSARQASQEVKEDTKEDKVPFYK-KHSTLLYAS-LLITFGYLS 118

Query: 110 GEQYGEEHVLPTIGYAASILIGGYSLFIKGLKNLRRLNFDMNTLMTIAIIGAAIIGEWGE 169
               GEE+++ T+ + AS+ IGG SLF  GL+NL R  FDM TLMT+A+IG AIIGEW E
```

```
Sbjct: 119 SYVNGEENIVTTLLFLASMFIGGLSLFKVGLQNLLRFEFDMKTLMTVAVIGGAIIGEWAE 178

Query: 170 GATVVILFAISEALERYSMDKARQSIESLMDIAPKEALIRRGNEEMMIHVDEIQVGDIMI 229
            A VVILFAISEALER+SMD+ARQSI SLMDIAPKEAL++R  +E+MIHVD+I VGDIMI
Sbjct: 179 VAIVVILFAISEALERFSMDRARQSIRSLMDIAPKEALVKRNGQEIMIHVDDIAVGDIMI 238

Query: 230 VKPGQKLAMDGIVVKGTSTLNQAAITGESVPVTKITNDEVFAGTLNEEGLLEVKVTKRVE 289
            VKPGQK+AMDG+VV G S +NQ AITGESVPV K  ++EVFAGTLNEEGLLEV++TK VE
Sbjct: 239 VKPGQKIAMDGVVVSGYSAVNQTAITGESVPVEKTVDNEVFAGTLNEEGLLEVEITKLVE 298

Query: 290 DTTLSKIIHLVEEAQAERAPSQAFVDKFAKYYTPAIVILALLIAVVPPL-FGGDWSQWIY 348
            DTT+SKIIHLVEEAQ ERAPSQAFVDKFAKYYTP I+I+A L+A+VPPL F G W  WIY
Sbjct: 299 DTTISKIIHLVEEAQGERAPSQAFVDKFAKYYTPIIMIIATLVAIVPPLFFDGSWETWIY 358

Query: 349 QGLAVLVVGCPCALVVSTPVAVVTAIGNAAKNGVLIKGGIHLEAAGHLKAIAFDKTGTLT 408
            QGLAVLVVGCPCALV+STP+++V+AIGNAAK GVL+KGG++LE  G LKAIAFDKTGTLT
Sbjct: 359 QGLAVLVVGCPCALVISTPISIVSAIGNAAKKGVLVKGGVYLEEMGALKAIAFDKTGTLT 418

Query: 409 KGIPAVTD--IVTYGRNENELITITSAIEKGSQHPLASAIMRKAEENGLKFNEVTVEDFQ 466
            KG+PAVTD  ++    NE EL++I +A+E  SQHPLASAIM+KAEE  + +++V VEDF
Sbjct: 419 KGVPAVTDYNVLNKQINEKELLSIITALEYRSQHPLASAIMKKAEEENITYSDVQVEDFS 478

Query: 467 SITGKGVKAKINNEMYYVGSQNLFEE-LHGSISSDKKEKIADMQTQGKTVMVLGTEKEIL 525
            SITGKG+K  +N   YY+GS  LF+E L      D ++ +  +Q QGKT M++GTEKEIL
Sbjct: 479 SITGKGIKGIVNGTTYYIGSPKLFKELLTNDFDKDLEQNVTTLQNQGKTAMIIGTEKEIL 538

Query: 526 SFIAVADEMRESSKEVIGKLNNMGI-ETVMLTGDNQRTATAIGKQVGVSDIKADLLPEDK 584
            + IAVADE+RESSKE++ KL+ +GI +T+MLTGDN+ TA AIG QVGVSDI+A+L+P+DK
Sbjct: 539 AVIAVADEVRESSKEILQKLHQLGIKKTIMLTGDNKGTANAIGGQVGVSDIEAELMPQDK 598

Query: 585 LNFIKELREKHQSVGMVGDGVNDAPALAASTVGVAMGGAGTDTALETADIALMSDDLSKL 644
            L+FIK+LR ++ +V MVGDGVNDAPALAASTVG+AMGGAGTDTALETAD+ALM DDL KL
Sbjct: 599 LDFIKQLRSEYGNVAMVGDGVNDAPALAASTVGIAMGGAGTDTALETADVALMGDDLRKL 658

Query: 645 PYTIKLSRKALAIIKQNITFSLAIKLVALLLVMPGWLTLWIAIFADMGATLLVTLNSLRL 704
            P T+KLSRK L IIK NITF++AIK +A LLV+PGWLTLWIAI +DMGATLLV LN LRL
Sbjct: 659 PSTVKLSRKTLNIIKANITFAIAIKFIASLLVIPGWLTLWIAILSDMGATLLVALNGLRL 718

Query: 705 LKIKE 709
            +++KE
Sbjct: 719 MRVKE 723
```

There is also homology to SEQ ID 3506.

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

Example 1984

A DNA sequence (GBSx2094) was identified in *S.agalactiae* <SEQ ID 6133> which encodes the amino acid sequence <SEQ ID 6134>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0779(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 1985

A DNA sequence (GBSx2095) was identified in *S.agalactiae* <SEQ ID 6135> which encodes the amino acid sequence <SEQ ID 6136>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.92   Transmembrane  123 - 139 ( 115 - 145)
   INTEGRAL    Likelihood = -6.74   Transmembrane  172 - 188 ( 167 - 190)
   INTEGRAL    Likelihood = -1.81   Transmembrane   80 -  96 (  80 -  96)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 9923> which encodes amino acid sequence <SEQ ID 9924> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 4216.

A related GBS gene <SEQ ID 8945> and protein <SEQ ID 8946> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:     -6.41
GvH: Signal Score (-7.5): -2.23
     Possible site: 58
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 3 value:  -8.92 threshold:  0.0
   INTEGRAL    Likelihood = -8.92   Transmembrane  123 - 139 ( 115 - 145)
   INTEGRAL    Likelihood = -6.74   Transmembrane  172 - 188 ( 167 - 190)
   INTEGRAL    Likelihood = -1.81   Transmembrane   80 -  96 (  80 -  96)
   PERIPHERAL  Likelihood =  2.92       46
 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>
```

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

Example 1986

A DNA sequence (GBSx2096) was identified in *S.agalactiae* <SEQ ID 6137> which encodes the amino acid sequence <SEQ ID 6138>. This protein is predicted to be histidine rich P type ATPase (HRA-1) (copB). Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.37   Transmembrane  318 - 334 ( 307 - 345)
   INTEGRAL    Likelihood = -5.84   Transmembrane  347 - 363 ( 335 - 364)
   INTEGRAL    Likelihood = -5.15   Transmembrane   88 - 104 (  86 - 112)
```

```
   INTEGRAL    Likelihood = -5.04   Transmembrane  651 - 667 ( 649 - 669)
   INTEGRAL    Likelihood = -4.30   Transmembrane  156 - 172 ( 155 - 173)
   INTEGRAL    Likelihood = -4.30   Transmembrane  669 - 685 ( 668 - 690)
   INTEGRAL    Likelihood = -3.03   Transmembrane   62 -  78 (  60 -  80)

----- Final Results -----
              bacterial membrane --- Certainty=0.6349(Affirmative) < succ>
               bacterial outside --- 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:AAA62113 GB:U16658 histidine rich P type ATPase [Escherichia
          coli]
 Identities = 598/731 (81%), Positives = 651/731 (88%), Gaps = 36/731 (4%)

Query: 1   MRNNKKHSSHSHHNHGDIDHSKHDHNEMEHSQMDHS---------------------- 36
           MRNNK+HSSHSHHNHGD++HSKHDHNEMEHSQMDHS
Sbjct: 1   MRNNKQHSSHSHHNHGDMEHSKHDHNEMEHSQMDHSAMGHCAMGGHAHHHHGDMDHSKHD 60

Query: 37  ------------NMDHSEMDHGAMGGHAHHHHGSFKEIFLKSLPLGIAILLITPMMDIQL 84
                       MD+SEMDHGAMGGHAHHHHGSFK+IFLKSLPLGIAILLITP+M IQL
Sbjct: 61  HNEMKHSQMDHSKMDYSEMDHGAMGGHAHHHHGSFKDIFLKSLPLGIAILLITPLMGIQL 120

Query: 85  PFQIIFPYADVVAAVLATILYIFGGKPFYMGAKDEFNSKAPGMMSLITLGITVSYAYSVY 144
           PFQIIFPYADVVAAVLATILYIFGGKPF MGAKDEFNSK PGMMSLITLGITVSYAYSVY
Sbjct: 121 PFQIIFPYADVVAAVLATILYIFGGKPFLMGAKDEFNSKVPGMMSLITLGITVSYAYSVY 180

Query: 145 AVAARYVTGEHVMDFFFEFTTLILIMLLGHWIEMKALGEAGDAQKALAELVPKDAHVVLE 204
           AVAARYVTGE VMDFFFEFTTLILIMLLGHWIEMKALGEAG+AQKALAELVPKDAHVVLE
Sbjct: 181 AVAARYVTGEPVMDFFFEFTTLILIMLLGHWIEMKALGEAGNAQKALAELVPKDAHVVLE 240

Query: 205 DDSIETRPVSELQIGDVIRVQAGENVPADGIIIRGESRVNEALVTGESKPIEKKTGDEVI 264
           DDSIETRPV++LQ+GD+IRVQAGENVPADG I RGESRVNEALVTGESKPIEK  GDEVI
Sbjct: 241 DDSIETRPVADLQVGDLIRVQAGENVPADGTIQRGESRVNEALVTGESKPIEKNPGDEVI 300

Query: 265 GGSTNGGGVLYVEIKQTGDQSFISQVQTLISQAQSQPSRAENVAQKVASWLFYIAVVVAL 324
           GGSTNG GVLYVEIKQTGD+SFISQVQTLISQAQSQPSRAEN+AQKVA WLFYIAV+ AL
Sbjct: 301 GGSTNGDGVLYVEIKQTGDKSFISQVQTLISQAQSQPSRAENLAQKVAGWLFYIAVIAAL 360

Query: 325 IALLIWTIIADLPTAVIFTVTALVIACPHALGLAIPLVVSRSTSLGASRGLLVKNREALE 384
           IAL+IW +IAD+PTAVIFTVT LVIACPHALGLAIPLV +RSTSLGASRGLLVK+R+ALE
Sbjct: 361 IALVIWMVIADVPTAVIFTVTTLVIACPHALGLAIPLVTARSTSLGASRGLLVKDRDALE 420

Query: 385 LTTKADVMVLDKTGTLTTGEFKVLDVTVLSDKYSEEEITGLLAGIEAGSSHPIAQSIVNH 444
           LTT ADVMVLDKTGTLTTGEFKVLDV + +DKY+++EI  LL+GIE GSSHPIAQSI+++
Sbjct: 421 LTTNADVMVLDKTGTLTTGEFKVLDVELFNDKYTKDEIVALLSGIEGGSSHPIAQSIISY 480

Query: 445 AEAKGIKSVSFDSIEIVSGAGIEGEANGHHYQLISQKAYGKALRMDIPKGATLSILVENN 504
           AE +GI+ VSFDSI+++SGAG+EG+ANGH YQLISQKAYG+ L MDIPKGAT+S+LVEN+
Sbjct: 481 AEQQGIRPVSFDSIDVMSGAGVEGQANGHRYQLISQKAYGRNLDMDIPKGATISVLVEND 540

Query: 505 EAIGAVALGDELKETSRNLIEVLKKYGIEPLMATGDNEEAAQGVAEVLGIQYQANQSPED 564
           EAIGAVALGDELK TS++LI+ LKK  I+P+MATGDNE+AAQG AE+LGI Y ANQSP+D
Sbjct: 541 EAIGAVALGDELKPTSKDLIQALKKNKIQPIMATGDNEKAAQGAAEILGIDYLANQSPQD 600

Query: 565 KYKLVESMKNQNKTVIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQSDPGDI 624
           KY+LVE +K + K VIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQ  PGDI
Sbjct: 601 KYELVEKLKAEGKKVIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQYSPGDI 660

Query: 625 ESFIELANKTTRKMKQNLVWGAGYNFIAIPIAAGLLAPIGITLGPAFGAVLMSLSTVIVA 684
            SFIELA KTTRKMK+NLVWGAGYNFIAIPIAAG+LAPIGITL PA  AVLMSLSTVIVA
Sbjct: 661 ASFIELAQKTTRKMKENLVWGAGYNFIAIPIAAGILAPIGITLSPAVAAVLMSLSTVIVA 720

Query: 685 INAMTLKLEPK 695
           INAMTLKLEPK
Sbjct: 721 INAMTLKLEPK 731
```

There is also homology to SEQ ID 3506.

A related GBS gene <SEQ ID 8947> and protein <SEQ ID 8948> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:    -19.12
GvH: Signal Score (-7.5): -3.71
     Possible site: 27
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 7 value: -13.37 threshold:  0.0
   INTEGRAL    Likelihood =-13.37   Transmembrane  291 - 307 ( 280 - 318)
   INTEGRAL    Likelihood = -5.84   Transmembrane  320 - 336 ( 308 - 337)
   INTEGRAL    Likelihood = -5.15   Transmembrane   61 -  77 (  59 -  85)
   INTEGRAL    Likelihood = -5.04   Transmembrane  624 - 640 ( 622 - 642)
   INTEGRAL    Likelihood = -4.30   Transmembrane  129 - 145 ( 128 - 146)
   INTEGRAL    Likelihood = -4.30   Transmembrane  642 - 658 ( 641 - 663)
   INTEGRAL    Likelihood = -3.03   Transmembrane   35 -  51 (  33 -  53)
   PERIPHERAL  Likelihood =  0.74     103
 modified ALOM score:   3.17

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6349(Affirmative) < succ>
                bacterial outside --- 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:

```
ORF02015(220 - 2304 of 2604)
EGAD|37454|38974(1 - 731 of 731) histidine rich P type ATPase (HRA-1) {Escherichia coli}
GP|643613|gb|AAA62113.1||U16658  histidine rich P type ATPase {Escherichia coli}
PIR|JC2464|JC2464  probable copper-transporting ATPase (EC 3.6.1.-) HRA-1 -
Enterobacteriaceae spp.
%Match = 67.4
%Identity = 85.9  %Similarity = 93.7
Matches = 598  Mismatches = 43  Conservative Sub.s = 54

162       192       222       252
PFRENYM*C*MRKF*NFKISL*YNKEELKMRNNKKHSSHSHHNHGDI--------------------------------
                            |||||:||||||||||:
                           MRNNKQHSSHSHHNHGDMEHSKHDHNEMEHSQMDHSAMGHCAMGGHAHHHH
                                    10        20        30        40        50

294       324       354       384       414       444       474       504
---DHSKHDHNEMEHSQMDHSNMDHSEMDHGAMGGHAHHHHGSFKEIFLKSLPLGIAILLITPMMDIQLPFQIIFPYADV
   ||||||||||:|||||| ||:||||||||||||||||||:||||||||||||||:| ||||||||||||||
GDMDHSKHDHNEMKHSQMDHSKMDYSEMDHGAMGGHAHHHHGSFKDIFLKSLPLGIAILLITPLMGIQLPFQIIFPYADV
                  70        80        90        100       110       120       130

534       564       594       624       654       684       714       744
VAAVLATILYIFGGKPFYMGAKDEFNSKAPGMMSLITLGITVSYAYSVYAVAARYVTGEHVMDFFFEFTTLILIMLLGHW
|||||||||||||||| |||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||
VAAVLATILYIFGGKPFLMGAKDEFNSKVPGMMSLITLGITVSYAYSVYAVAARYVTGEPVMDFFFEFTTLILIMLLGHW
                  150       160       170       180       190       200       210

774       804       834       864       894       924       954       984
IEMKALGEAGDAQKALAELVPKDAHVVLEDDSIETRPVSELQIGDVIRVQAGENVPADGIIIRGESRVNEALVTGESKPI
||||||||||:|||||||||||||||||||||||||::||:||:||||||||||||| | ||||||||||||||||||
IEMKALGEAGNAQKALAELVPKDAHVVLEDDSIETRPVADLQVGDLIRVQAGENVPADGTIQRGESRVNEALVTGESKPI
                  230       240       250       260       270       280       290

1014      1044      1074      1104      1134      1164      1194      1224
EKKTGDEVIGGSTNGGGVLYVEIKQTGDQSFISQVQTLISQAQSQPSRAENVAQKVASWLFYIAVVVALIALLIWTIIAD
||  |||||||||| ||||||||||:|||||||||||||||||||:||||| |||||||: |||||:|| :|||
EKNPGDEVIGGSTNGDGVLYVEIKQTGDKSFISQVQTLISQAQSQPSRAENLAQKVAGWLFYIAVIAALIALVIWMVIAD
```

```
                310        320        330        340        350        360        370

1254       1284       1314       1344       1374       1404       1434       1464
LPTAVIFTVTALVIACPHALGLAIPLVVSRSTSLGASRGLLVKNREALELTTKADVMVLDKTGTLTTGEFKVLDVTVLSD
:||||||||| |||||||||||||||| :|||||||||||||:|:|||||| |||||||||||||||||||||| :::|
VPTAVIFTVTTLVIACPHALGLAIPLVTARSTSLGASRGLLVKDRDALELTTNADVMVLDKTGTLTTGEFKVLDVELFND
                390        400        410        420        430        440        450

1494       1524       1554       1584       1614       1644       1674       1704
KYSEEEITGLLAGIEAGSSHPIAQSIVNHAEAKGIKSVSFDSIEIVSGAGIEGEANGHHYQLISQKAYGKALRMDIPKGA
||:::||  ||:||| ||||||||||:::|| :||: ||||||:::||||:||:|||| ||||||||||: | |||||||
KYTKDEIVALLSGIEGGSSHPIAQSIISYAEQQGIRPVSFDSIDVMSGAGVEGQANGHRYQLISQKAYGRNLDMDIPKGA
                470        480        490        500        510        520        530

1734       1764       1794       1824       1854       1884       1914       1944
TLSILVENNEAIGAVALGDELKETSRNLIEVLKKYGIEPLMATGDNEEAAQGVAEVLGIQYQANQSPEDKYKLVESMKNQ
|:|:||||:|||||||||||||| ||::||: ||| |:|:|||||||:|||| ||:||| | |||||:|||:||| :| :
TISVLVENDEAIGAVALGDELKPTSKDLIQALKKNKIQPIMATGDNEKAAQGAAEILGIDYLANQSPQDKYELVEKLKAE
                550        560        570        580        590        600        610

1974       2004       2034       2064       2094       2124       2154       2184
NKTVIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQSDPGDIESFIELANKTTRKMKQNLVWGAGYNFIAIPI
 | ||||||||||||||||||||||||||||||||||||||||||||||| |||| |||||| |||||||:|||||||||||||||
GKKVIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQYSPGDIASFIELAQKTTRKMKENLVWGAGYNFIAIPI
                630        640        650        660        670        680        690

2214       2244       2274       2304       2334       2364       2394       2424
AAGLLAPIGITLGPAFGAVLMSLSTVIVAINAMTLKLEPK*NEAGTKKHWLV*PPSRIGSDQLVCCIRKIIDR*IFDKNR
|||:|||||||| ||  ||||||||||||||||||||||
AAGILAPIGITLSPAVAAVLMSLSTVIVAINAMTLKLEPK
                710        720        730
```

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

Example 1987

A DNA sequence (GBSx2097) was identified in *S.agalactiae* <SEQ ID 6139> which encodes the amino acid sequence <SEQ ID 6140>. This protein is predicted to be CopA. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2197(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:CAA40599 GB:X57326 ORF-1 [Thiobacillus ferrooxidans]
 Identities = 26/65 (40%), Positives = 40/65 (61%), Gaps = 2/65 (3%)

Query: 1  MKQEILL--DGVKCAGCANTVQERFSAIEGVESVEVDLATKKAVLESQTEIDTETLNAAL 58
          M Q+I L   G+ CA CA++V++    I G++S +V LAT +A +  Q+ I TE L AA+
Sbjct: 1  MSQKIFLRITGMTCAHCAHSVEKALLGIHGIDSAQVSLATNQAEVFLQSSIPTEALLAAV 60

Query: 59 AETNY 63
           +  Y
Sbjct: 61 TQAGY 65
```

There is also homology to SEQ ID 3510.

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

Example 1988

A DNA sequence (GBSx2098) was identified in *S.agalactiae* <SEQ ID 6141> which encodes the amino acid sequence <SEQ ID 6142>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3220(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 1989

A DNA sequence (GBSx2099) was identified in *S.agalactiae* <SEQ ID 6143> which encodes the amino acid sequence <SEQ ID 6144>. This protein is predicted to be heavy-metal transporting P-type ATPase (b0484). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.09   Transmembrane  131 - 147 ( 130 - 150)

----- Final Results -----
               bacterial membrane --- Certainty=0.2635(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAB01764 GB:U42410 heavy-metal transporting P-type ATPase
           [Proteus mirabilis]
 Identities = 98/153 (64%), Positives = 123/153 (80%)

Query: 2   KAVKALRRRGVEVIMITGDNKRTAKAIAKQVGIDSVLSEVLPEDKAEEVKKLQEAGKKVA 61
           +A+KAL   G++V MITGDNK TAKAIAKQ+GID +++EVLP+ K   +K+L + G KVA
Sbjct: 649 EAIKALHALGLKVAMITGDNKATAKAIAKQLGIDEIVAEVLPDGKVAALKQLSQKGDKVA 708

Query: 62  MVGDGINDAPALAQANVGIAVGSGTDVAIESADIVLMRNDLTAVLTTIDLSHATLRNIKQ 121
            VGDGINDAPALAQA+VG+A+G+GTDVAIE+AD+VLM  DL  V+  I LS AT+RNIKQ
Sbjct: 709 FVGDGINDAPALAQADVGLAIGTGTDVAIEAADVVLMSGDLRGVVDAIALSQATIRNIKQ 768

Query: 122 NLFWAFAYNLVGIPVAMGLLYIFGGLLMSPMLA 154
           NLFW FAYN + IPVA G+LY   G+L+SP+ A
Sbjct: 769 NLFWTFAYNALLIPVAAGMLYPINGMLLSPIFA 801
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3505> which encodes the amino acid sequence <SEQ ID 3506>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.83   Transmembrane  328 - 344 ( 314 - 348)
```

```
INTEGRAL    Likelihood = -7.01   Transmembrane  354 - 370 ( 347 - 377)
INTEGRAL    Likelihood = -3.24   Transmembrane  101 - 117 ( 100 - 117)
INTEGRAL    Likelihood = -2.97   Transmembrane  165 - 181 ( 165 - 185)
INTEGRAL    Likelihood = -2.34   Transmembrane  665 - 681 ( 662 - 684)
INTEGRAL    Likelihood = -2.18   Transmembrane   67 -  83 (  66 -  83)
INTEGRAL    Likelihood = -0.64   Transmembrane  491 - 507 ( 490 - 508)
INTEGRAL    Likelihood = -0.59   Transmembrane  691 - 707 ( 691 - 707)
INTEGRAL    Likelihood = -0.43   Transmembrane  140 - 156 ( 139 - 156)

----- 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 = 92/152 (60%), Positives = 123/152 (80%)

Query: 4   VKALRRRGVEVIMITGDNKRTAKAIAKQVGIDSVLSEVLPEDKAEEVKKLQEAGKKVAMV 63
           V+AL + G+  IM+TGD+  TAKAIA QVGI  V+S+VLP+ KA  +  L+  G+KVAMV
Sbjct: 544 VEALHQLGIHTIMLTGDHDATAKAIASQVGITDVISQVLPDQKAGVIADLRSQGRKVAMV 603

Query: 64  GDGINDAPALAQANVGIAVGSGTDVAIESADIVLMRNDLTAVLTTIDLSHATLRNIKQNL 123
           GDGINDAPALA A++GIA+GSGTD+AIESAD++LM+ D+  ++  + LS  T+R +K+NL
Sbjct: 604 GDGINDAPALAVADIGIAMGSGTDIAIESADVILMKPDMLDLVKAMSLSRVTMRIVKENL 663

Query: 124 FWAFAYNLVGIPVAMGLLYIFGGLLMSPMLAG 155
           FWAF YN++ IPVAMGLL++FGG L++PMLAG
Sbjct: 664 FWAFIYNVLMIPVAMGLLHLFGGPLLNPMLAG 695
```

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

Example 1990

A DNA sequence (GBSx2100) was identified in *S.agalactiae* <SEQ ID 6145> which encodes the amino acid sequence <SEQ ID 6146>. This protein is predicted to be CopY. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2067(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:AAG10085 GB:AF296446 CopY [Streptococcus mutans]
 Identities = 63/139 (45%), Positives = 96/139 (68%)

Query: 8   TSITDAEWEVMRVVWANDLVTSKTVISVLKEKMDWTESTIKTILGRLVEKGVLNTEQEGR 67
           TSI++AEWEVMRVVWA  + +S  +I++L     W+ STIKT++ RL EKG L ++++GR
Sbjct: 2   TSISNAEWEVMRVVWAKQMTSSSEIIAILSRTYCWSASTIKTLITRLSEKGYLTSQRQGR 61

Query: 68  KFIYTANIVEKEAVRDFAEDIFNRICKKKVGNVIGSIIEDHVLSFDDIDRLEKILEIKKS 127
           K+IY++ I E+EA+     ++F+RIC  K   +I  ++E+  ++  DI++LE +L  KK+
Sbjct: 62  KYIYSSLISEEEALEQQVSEVFSRICVTKHQALIRHLVEETPMTLSDIEKLEALLLSKKA 121

Query: 128 FAVEEVDCQCTEGQCDCHE 146
            AV EV C C  GQC C+E
Sbjct: 122 NAVPEVKCNCIVGQCSCYE 140
```

There is also homology to SEQ ID 3502.

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

Example 1991

A DNA sequence (GBSx2101) was identified in *S.agalactiae* <SEQ ID 6147> which encodes the amino acid sequence <SEQ ID 6148>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2829(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 1992

A DNA sequence (GBSx2102) was identified in *S.agalactiae* <SEQ ID 6149> which encodes the amino acid sequence <SEQ ID 6150>. This protein is predicted to be DS RF protein. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-13.21   Transmembrane  142 - 158 ( 136 - 169)
   INTEGRAL    Likelihood = -3.45   Transmembrane   70 -  86 (  66 -  88)
   INTEGRAL    Likelihood = -3.13   Transmembrane  178 - 194 ( 176 - 195)

----- Final Results -----
             bacterial membrane --- Certainty=0.6286(Affirmative) < succ>
              bacterial outside --- 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:AAA26611 GB:L10909 putative [Staphylococcus aureus]
 Identities = 98/204 (48%), Positives = 148/204 (72%), Gaps = 3/204 (1%)

Query: 4   TIISAIGVYISTSIDYLIVLIILFAQLSQNKQKWHIYAGQYLGTGLLVGASLVAAY-VVN 62
           TI++A  VY++T IDYL++LI+LF+Q+ + + K HI+ GQY+GT +++GASL+ A  VVN
Sbjct: 18  TILTATAVYVATGIDYLVILILLFSQVKKGQVK-HIWIGQYIGTAIVIGASLLVAQGVVN 76

Query: 63  FVPEAWMVGLLGLIPIYLGIRFAIVGEGEEEEEEEEIIERLEQSKANQLFWTVTLLTIASG 122
            +P+ W++GLLGL+P+YLG++  I GE E+E+E  I+       K NQLF T+  + +AS
Sbjct: 77  LIPQQWVIGLLGLLPLYLGVKIWIKGE-EDEDESSILSLFSSGKFNQLFLTMIFIVLASS 135

Query: 123 GDNLGIYIPYFASLDWSQTLVVLLVFAIGIIIFCELSWVLSSIPLISETIEKYQRIIVPL 182
            D+  IYIPYF +L  S+  +V +VF I + + C +S+ L+S   ISETIEKY+R IVP+
Sbjct: 136 ADDFSIYIPYFTTLSMSEIFIVTIVFLIMVGVLCYVSYRLASFDFISETIEKYERWIVPI 195

Query: 183 VFIPLGLYIMYESGTIETFLNFIL 206
           VFI LG+YI++E+GT    ++F+L
Sbjct: 196 VFIGLGIYILFENGTSNALISFLL 219
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6151> which encodes the amino acid sequence <SEQ ID 6152>. Analysis of this protein sequence reveals the following:

```
      Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.16   Transmembrane  143 - 159 ( 135 - 165)
   INTEGRAL    Likelihood = -9.13   Transmembrane   49 -  65 (  43 -  71)
   INTEGRAL    Likelihood = -7.17   Transmembrane   73 -  89 (  72 -  94)
   INTEGRAL    Likelihood = -6.00   Transmembrane   13 -  29 (   9 -  33)
   INTEGRAL    Likelihood = -2.71   Transmembrane  180 - 196 ( 179 - 197)
   INTEGRAL    Likelihood = -0.59   Transmembrane  112 - 128 ( 109 - 128)

----- 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:AAF42284 GB:AE002544 cadmium resistance protein [Neisseria
          meningitidis MC58]
 Identities = 201/208 (96%), Positives = 205/208 (97%)

Query: 1   MRCFMIQNVVTSIILYSGTAVDLLIILMLFFAKRKSRKDIINIYLGQFLGSVSLILLSLL 60
           MRCFMIQNVVTSIILYSGTAVDLLIILMLFFAKRKSRKDIINIYLGQFLGSVSLILLSLL
Sbjct: 1   MRCFMIQNVVTSIILYSGTAVDLLIILMLFFAKRKSRKDIINIYLGQFLGSVSLILLSLL 60

Query: 61  FAFVLDYIPSKEILGLLGLIPIFLGLKVLLLGDSDGEAIAKEGLSKDNKNLIFLVAMITF 120
           FAFVLDYIPSKEILGLLGLIPI LG+KVLLLGDSDGEAIAKEGL KDNKNLIFLVAMITF
Sbjct: 61  FAFVLDYIPSKEILGLLGLIPILLGIKVLLLGDSDGEAIAKEGLRKDNKNLIFLVAMITF 120

Query: 121 ASCGADNIGVFVPYFTTLNLANLIVALLTFLVMIYLLVFSAQKLAQVPSVGETLEKYSRW 180
           ASCGADNIGVFVPYFTTLNLANLIVALLTFLVMIYLLVFSAQKLAQVPSVGETLEKYSRW
Sbjct: 121 ASCGADNIGVFVPYFTTLNLANLIVALLTFLVMIYLLVFSAQKLAQVPSVGETLEKYSRW 180

Query: 181 FIAVVYLGLGMYILIENNSFDMLWAVLG 208
           F+AVVYLGLG+YIL+ENNSFDMLW VLG
Sbjct: 181 FVAVVYLGLGIYILVENNSFDMLWTVLG 208
```

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

```
 Identities = 71/200 (35%), Positives = 130/200 (64%), Gaps = 4/200 (2%)

Query: 1   MGQTIISAIGVYISTSIDYLIVLIILFAQLSQNKQKWHIYAGQYLGTGLLVGASLVAAYV 60
           M Q ++++I +Y  T++D LI+L++ FA+    K   +IY GQ+LG+  L+  SL+ A+V
Sbjct: 5   MIQNVVTSIILYSGTAVDLLIILMLFFAKRKSRKDIINIYLGQFLGSVSLILLSLLFAFV 64

Query: 61  VNFVPEAWMVGLLGLIPIYLGIRFAIVGEGEEEEEEEIIERLEQSKANQLFWTVTLLTIA 120
           ++++P   ++GLLGLIPI+LG++  ++G+ + E   +  E L +   N +F  V ++T A
Sbjct: 65  LDYIPSKEILGLLGLIPIFLGLKVLLLGDSDGEAIAK--EGLSKDNKNLIF-LVAMITFA 121

Query: 121 S-GGDNLGIYIPYFASLDWSQTLVVLLVFAIGIIIFCELSWVLSSIPLISETIEKYQRII 179
           S G DN+G+++PYF +L+ +  +V LL F + I +     +  L+ +P + ET+EKY R
Sbjct: 122 SCGADNIGVFVPYFTTLNLANLIVALLTFLVMIYLLVFSAQKLAQVPSVGETLEKYSRWF 181

Query: 180 VPLVFIPLGLYIMYESGTIE 199
           + +V++ LG+YI+ E+ + +
Sbjct: 182 IAVVYLGLGMYILIENNSFD 201
```

SEQ ID 6150 (GBS174) was expressed in and purified from *E.coli*. The purified protein is shown in lane 7 of Figure 223.

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

Example 1993

A DNA sequence (GBSx2103) was identified in *S.agalactiae* <SEQ ID 6153> which encodes the amino acid sequence <SEQ ID 6154>. This protein is predicted to be Pgm. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4324(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:CAB96418 GB:AJ243290 phosphoglucomutase [Streptococcus thermophilus]
 Identities = 65/76 (85%), Positives = 71/76 (92%)

Query: 1  MTYTENLQKWLDFEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI 60
          M+YTEN QKWLDF +LP YLR EL+SMDEKTKEDAFYTNLEFGTAGMRG IGAGTNRINI
Sbjct: 1  MSYTENYQKWLDFAELPAYLRDELVSMDEKTKEDAFYTNLEFGTAGMRGLIGAGTNRINI 60

Query: 61 YVVRQATEGLAKLIET 76
          YVVRQATEGLA+LI++
Sbjct: 61 YVVRQATEGLAQLIDS 76
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6155> which encodes the amino acid sequence <SEQ ID 6156>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4324(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 = 75/76 (98%), Positives = 75/76 (98%)

Query: 1  MTYTENLQKWLDFEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI 60
          MTYTEN QKWLDFEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI
Sbjct: 1  MTYTENFQKWLDFEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI 60

Query: 61 YVVRQATEGLAKLIET 76
          YVVRQATEGLAKLIET
Sbjct: 61 YVVRQATEGLAKLIET 76
```

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

Example 1994

A DNA sequence (GBSx2104) was identified in *S.agalactiae* <SEQ ID 6157> which encodes the amino acid sequence <SEQ ID 6158>. This protein is predicted to be a membrane protein. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.21   Transmembrane   94 - 110 (  93 - 115)
   INTEGRAL    Likelihood = -4.14   Transmembrane  172 - 188 ( 166 - 188)
```

```
   INTEGRAL    Likelihood = -1.97   Transmembrane  130 - 146 ( 129 - 149)
   INTEGRAL    Likelihood = -0.16   Transmembrane   62 -  78 (  62 -  79)

----- Final Results -----
                bacterial membrane --- Certainty=0.3484(Affirmative) < succ>
                 bacterial outside --- 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:CAA80247 GB:Z22520 membrane protein [Bacillus acidopullulyticus]
 Identities = 47/185 (25%), Positives = 80/185 (42%), Gaps = 23/185 (12%)

Query: 1   MKKKNKSSNIAIIAIFFAIMLVIHFLSSFIFSFWLVPIKPTLMHIPVIIASIAYGPRIGA 60
           MKK     +I I  +  A+ +++                    T+MHIP II  I  GP +G
Sbjct: 1   MKKSLTVRDIVIAGVLGAVAILLGVTRLGYIPVPTAAGNATIMHIPAIIGGIMQGPVVGL 60

Query: 61  TLGALMGGISVANSSIVLLPTSYLFSPFVENGNFYSLIIALVPRILIGIIPYFVYKLLHN 120
            +GA+ G  S  N+++ L                F   +++++PR+ IG++ + VY  +
Sbjct: 61  IVGAIFGISSFLNATVPL---------------FKDPLVSILPRLFIGVVAWLVYIGIRR 105

Query: 121 R---FGLAISGAIGSLTNTVFVLSGIFIFFSSTYNGNIKLMLAGIISSNSLAEMVIAAII 177
           +     + +S  IG+LTNT  VL+     F       +   +A    +N L E V+  I+
Sbjct: 106 KSEYVAVGLSAFIGTLTNTALVLA--MAVFRHYLTAGVAWTVA---ITNGLPEAVVGTIV 160

Query: 178 VYLTV 182
               V
Sbjct: 161 TLAVV 165
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6159> which encodes the amino acid sequence <SEQ ID 6160>. Analysis of this protein sequence reveals the following:

```
     Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.97   Transmembrane   18 -  34 (  10 -  41)
   INTEGRAL    Likelihood = -7.43   Transmembrane  170 - 186 ( 160 - 191)
   INTEGRAL    Likelihood = -5.63   Transmembrane   96 - 112 (  94 - 117)
   INTEGRAL    Likelihood = -4.67   Transmembrane  140 - 156 ( 131 - 158)
   INTEGRAL    Likelihood = -3.66   Transmembrane   64 -  80 (  63 -  84)
   INTEGRAL    Likelihood = -0.22   Transmembrane   39 -  55 (  39 -  55)

----- Final Results -----
               bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAA80247 GB:Z22520 membrane protein [Bacillus acidopullulyticus]
 Identities = 47/193 (24%), Positives = 86/193 (44%), Gaps = 28/193 (14%)

Query: 8   RKSADISRIAIFFAIMLVIHFVSSLVFNIWPIPI---KPTLVHIPVIIASVLYGPRIGAI 64
           +KS  +  I I   +  V   +        P+P      T++HIP II  ++ GP +G I
Sbjct: 2   KKSLTVRDIVIAGVLGAVAILLGVTRLGYIPVPTAAGNATIMHIPAIIGGIMQGPVVGLI 61

Query: 65  LGGLMGIISVITNTIILLPTNYLFSPFVDHGTFASLIIAIIPRILIGITPYYCYKLIPNQ 124
           +G + GI S +  T+ L               F   +++I+PR+ IG+  + CY  +  +
Sbjct: 62  VGAIFGISSFLNATVPL---------------FKDPLVSILPRLFIGVVAWLVYIGIRRK 106

Query: 125 FGLIVSGI---IGSLTNTIFVLS-GIFIFFATVFDGNIKALLTAIISSNAIVEMIISAII 180
              + G+   IG+LTNT  VL+  +F  + T        +   +  +N + E ++  I+
Sbjct: 107 SEYVAVGLSAFIGTLTNTALVLAMAVFRHYLTA------GVAWTVAITNGLPEAVVGTIV 160

Query: 181 TFVLIPTLSRLKR 193
           T  ++    ++ R
Sbjct: 161 TLAVVLAWKQIGR 173
```

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

```
 Identities = 121/184 (65%), Positives = 157/184 (84%)

Query: 6   KSSNIAIIAIFFAIMLVIHFLSSFIFSFWLVPIKPTLMHIPVIIASIAYGPRIGATLGAL 65
           KS++I+ IAIFFAIMLVIHF+SS +F+ W +PIKPTL+HIPVIIAS+ YGPRIGA LG L
Sbjct: 9   KSADISRIAIFFAIMLVIHFVSSLVFNIWPIPIKPTLVHIPVIIASVLYGPRIGAILGGL 68

Query: 66  MGGISVANSSIVLLPTSYLFSPFVENGNFYSLIIALVPRILIGIIPYFVYKLLHNRFGLA 125
           MG ISV  ++I+LLPT+YLFSPFV++G F SLIIA++PRILIGI PY+ YKL+ N+FGL
Sbjct: 69  MGIISVITNTIILLPTNYLFSPFVDHGTFASLIIAIIPRILIGITPYYCYKLIPNQFGLI 128

Query: 126 ISGAIGSLTNTVFVLSGIFIFFSSTYNGNIKLMLAGIISSNSLAEMVIAAIIVYLTVPRI 185
           +SG IGSLTNT+FVLSGIFIFF++ ++GNIK +L  IISSN++ EM+I+AII ++ +P +
Sbjct: 129 VSGIIGSLTNTIFVLSGIFIFFATVFDGNIKALLTAIISSNAIVEMIISAIITFVLIPTL 188

Query: 186 LNIK 189
             +K
Sbjct: 189 SRLK 192
```

A related GBS gene <SEQ ID 8949> and protein <SEQ ID 8950> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:     13.42
GvH: Signal Score (-7.5): -1.93
     Possible site: 53
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 2 value:  -6.21 threshold:  0.0
   INTEGRAL    Likelihood = -6.21   Transmembrane   94 - 110 (  93 - 115)
   INTEGRAL    Likelihood = -0.16   Transmembrane   62 -  78 (  62 -  79)
   PERIPHERAL  Likelihood =  1.70      123
 modified ALOM score:   1.74

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3484(Affirmative) < succ>
                 bacterial outside --- 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:

```
ORF01561(301 - 723 of 1017)
EGAD|38021|39600(1 - 129 of 183) hypothetical membrane protein {Bacillus acidopullulyticus}
GP|806536|emb|CAA80247.1||Z22520 membrane protein {Bacillus acidopullulyticus}
%Match = 7.6
%Identity = 29.7  %Similarity = 53.9
Matches = 38  Mismatches = 57  Conservative Sub.s = 31

162       192       222       252       282       312       342       372
KKIGYQEIEPRISLLACGDTGQGALADISTILKCIQEVAN*AVNLYTISSLI*GVIMKKKNKSSNIAIIAIFFAIMLVIH
                                                    |||     :| |  :: |: :::
                                                   MKKSLTVRDIVIAGVLGAVAILLG
                                                           10        20

402       432       462       492       522       552       582       612
FLSSFIFSFWLVPIKPTLMHIPVIIASIAYGPRIGATLGALMGGISVANSSIVLLPTSYLFSPFVENGNFYSLIIALVPR
                  |:|||| ||  |  || :|  :||: |  |  |::: |               |   :::::||
VTRLGYIPVPTAAGNATIMHIPAIIGGIMQGPVVGLIVGAIFGISSFLNATVPL---------------FKDPLVSILPR
               40        50        60        70                        80

642       663       693       723       753       783       813       843
ILIGIIPYFVY---KLLHNRFGLAISGAIGSLTNTVFVXSGIFIFFSSTYNGNIKLMLAGIISXNSLAEMVIAAIIVYLT
::||:: ::||   :       : :|  ||:|||| :| :
LFIGVVAWLVYIGIRRKSEYVAVGLSAFIGTLTNTALVLAMAVFRHYLTAGVAWTVAITNGLPEAVVGTIVTLAVVLAWK
        100       110       120       130       140       150       160
```

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

Example 1995

A DNA sequence (GBSx2105) was identified in *S.agalactiae* <SEQ ID 6161> which encodes the amino acid sequence <SEQ ID 6162>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence (or aa 1-18)

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0165(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:AAC44502 GB:U48885 DNA/pantothenate metabolism flavoprotein
          [Streptococcus mutans]
 Identities = 101/145 (69%), Positives = 122/145 (83%)

Query: 1   MIKRITLAVTGSISAYKAADLTSQLTKIGYDVHIIMTQAATEFITPLTLQVLSKNPIHLD 60
           M K+I LAV+GSI+AYKAADL+ QLTK+GY V++ MT AA +FI PLTLQVLSKNP++ +
Sbjct: 1   MTKKILLAVSGSIAAYKAADLSHQLTKLGYHVNVFMTNAAKQFIPPLTLQVLSKNPVYSN 60

Query: 61  VMDEHNPKIINHIELAKRTDLFIVAPASANTIAHLAYGFADNIVTSVALAMPDETPKLIA 120
           VM E +P++INHI LAK+ DLF++ PASANT+AHLA+GFADNIVTSVALA+P E PK  A
Sbjct: 61  VMKEDDPQVINHIALAKQADLFLLPPASANTLAHLAHGFADNIVTSVALALPLEVPKFFA 120

Query: 121 PAMNTKMYHNTITQRNIDILKKIGY 145
           PAMNTKMY N ITQ NI +LKK GY
Sbjct: 121 PAMNTKMYENPITQSNITLLKKFGY 145
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6163> which encodes the amino acid sequence <SEQ ID 6164>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0076(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 = 146/178 (82%), Positives = 155/178 (87%)

Query: 1   MIKRITLAVTGSISAYKAADLTSQLTKIGYDVHIIMTQAATEFITPLTLQVLSKNPIHLD 60
           M K ITLAV+GSISAYKAADLTSQLTKIGYDVHIIMTQAAT+FITPLTLQVLSKN IHLD
Sbjct: 1   MTKHITLAVSGSISAYKAADLTSQLTKIGYDVHIIMTQAATQFITPLTLQVLSKNAIHLD 60

Query: 61  VMDEHNPKIINHIELAKRTDLFIVAPASANTIAHLAYGFADNIVTSVALAMPDETPKLIA 120
           VMDEH+PK+INHIELAKRTDLFIVAPASANTIAHLAYGFADN+VTSVALA+P  TPKLIA
Sbjct: 61  VMDEHDPKVINHIELAKRTDLFIVAPASANTIAHLAYGFADNLVTSVALALPATTPKLIA 120

Query: 121 PAMNTKMYHNTITQRNIDILKKIGYQEIEPRISLLACGDTGQGALADISTILKCIQEV 178
           PAMNTKMY N ITQ NI  L  IG+ EI P+ SLLACGD G GALADI  IL  I  +
Sbjct: 121 PAMNTKMYQNPITQENIKRLSTIGFTEIPPKSSLLACGDKGPGALADIDVILATIDTI 178
```

SEQ ID 6162 (GBS236) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 5; MW 21.6kDa).

Purified GBS236-GST is shown in Figure 208 (lane 6) and in Figure 225 (lanes 4-5).

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

Example 1996

A DNA sequence (GBSx2106) was identified in *S.agalactiae* <SEQ ID 6165> which encodes the amino acid sequence <SEQ ID 6166>. This protein is predicted to be pantothenate metabolism flavoprotein homolog (dfp). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2325(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 9835> which encodes amino acid sequence <SEQ ID 9836> was also identified.

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

```
>GP:AAG39941 GB:AF301375 MTW1216 [Methanothermobacter wolfeii
          prophage psiM100]
 Identities = 71/229 (31%), Positives = 117/229 (51%), Gaps = 27/229 (11%)

Query: 6   MKILITSGGTTEKIDTVRSITNHATGTLGKIIAEKYLREGHQVTLVTTKNAVKPESATNL 65
           +++L++ GGT E ID VR ITN ++G +G  +A +   +G  VTLV     V  +  + L
Sbjct: 172 LRVLVSLGGTLEPIDPVRVITNRSSGRMGLAVAREAYIQGADVTLVA--GTVSVDIPSQL 229

Query: 66  STFEIEDVDSLIKTLKPLVKEHDILIHSMAVSDYTPVYMADFEKVKSSDHLDTFLRKDNH 125
            T   E    + + +  L+ EHD+ + + AVSD+ PVY
Sbjct: 230 RTVRAETAHEMAEAVAELIGEHDVFVSAAAVSDFRPVYS---------------------- 268

Query: 126 EGKISSESEYQVLFLKKTPKVISLVKKWNPQITLVGFKLLVNVTKENLFKVARHSLIKNK 185
           E KISS+SE   L LK  PK+I + ++ NP+  +VGFK    V++E L   AR  +  +
Sbjct: 269 EEKISSDSEI-TLRLKPNPKIIRMARETNPEAFIVGFKAEHGVSEEELIAAARKQIEDSV 327

Query: 186 ATFILANDL-IDITSKHHIAYLLDHDNVYKATT--KEDIAQLIYEKVKK 231
           A  ++AND+ ++     +   ++  + V +  T  KE++A LI  ++ K
Sbjct: 328 ADMVVANDVSVEGFGSENNRAIIVSEGVTELPIMKKEELAGLIIGEIMK 376
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6167> which encodes the amino acid sequence <SEQ ID 6168>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1737(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 = 142/230 (61%), Positives = 170/230 (73%)

Query: 4   MAMKILITSGGTTEKIDTVRSITNHATGTLGKIIAEKYLREGHQVTLVTTKNAVKPESAT 63
           M MK++ITSGGTTE ID VR ITNH+TG LGK+I E++L+  H VTLVTTK A KP
Sbjct: 1   MTMKLIITSGGTTEPIDAVRGITNHSTGQLGKLITERFLQYHHDVTLVTTKTATKPLPNK 60
```

```
Query: 64  NLSTFEIEDVDSLIKTLKPLVKEHDILIHSMAVSDYTPVYMADFEKVKSSDHLDTFLRKD 123
           L   E+E V+ L+  LK  V  HDILIHSMAVSDYTPVYM D E+V  +D+L+ FL +
Sbjct: 61  RLRIIEVETVNDLMAALKDQVPHHDILIHSMAVSDYTPVYMTDLEQVSQADNLNCFLCEH 120

Query: 124 NHEGKISSESEYQVLFLKKTPKVISLVKKWNPQITLVGFKLLVNVTKENLFKVARHSLIK 183
           N E KISS S+YQVLFLKKTPKVIS VK+WNP I LVGFKLLVNV +E L KVAR SL K
Sbjct: 121 NSEPKISSASDYQVLFLKKTPKVISYVKQWNPNIKLVGFKLLVNVPQEELIKVARASLAK 180

Query: 184 NKATFILANDLIDITSKHHIAYLLDHDNVYKATTKEDIAQLIYEKVKKYD 233
           N A +ILANDL+DI +  H A L+ ++ V  A TKE IA L+YE++ K+D
Sbjct: 181 NHADYILANDLVDIQTGMHKALLISNNEVASADTKEAIADLLYERMTKHD 230
```

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

Example 1997

A DNA sequence (GBSx2107) was identified in *S.agalactiae* <SEQ ID 6169> which encodes the amino acid sequence <SEQ ID 6170>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22   Transmembrane  117 - 133 ( 117 - 133)

----- Final Results -----
               bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9833> which encodes amino acid sequence <SEQ ID 9834> was also identified.

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

```
>GP:BAB07541 GB:AP001520 unknown conserved protein in B. subtilis
          [Bacillus halodurans]
 Identities = 94/221 (42%), Positives = 133/221 (59%), Gaps = 2/221 (0%)

Query: 52  AEKPFIWTEVFLREINRSNQEIILHIWPMTKTVILGMLDRELPHLELAKKEIISRGYEPV 111
           A + F + +     I +S     L  W    TV+LG+ D  LP ++   + +    ++ +
Sbjct: 27  ALQSFAYDDTLCTSIGKSQSPPTLRAWVHHNTVVLGIQDSRLPQIKAGIEALKGFQHDVI 86

Query: 112 VRNFGGLAVVADEGILNFSLVIPDVFERKLSISDGYLIMVDFIRSIFSDFYQPIEHFEVE 171
           VRN GGLAVV D GILN SLV+ +  E+  SI DGY +M + I S+F D  + IE  E+
Sbjct: 87  VRNSGGLAVVLDSGILNLSLVLKE--EKGFSIDDGYELMYELICSMFQDHREQIEAREIV 144

Query: 172 TSYCPGKFDLSINGKKFAGLAQRRIKNGIAVSIYLSVCGDQKGRSQMISDFYKIGLGDTG 231
            SYCPG +DLSI+GKKFAG++QRRI+ G+AV IYL V G    R++MI  FY   +
Sbjct: 145 GSYCPGSYDLSIDGKKFAGISQRRIRGGVAVQIYLCVSGSGAERAKMIRTFYDKAVAGQP 204

Query: 232 SPIAYPNVDPEIMANLSDLLDCPMTVEDVIDRMLISLKQVG 272
           +  YP + PE MA+LS+LL  P  V DV+ + L++L+Q G
Sbjct: 205 TKFVYPRIKPETMASLSELLGQPHNVSDVLLKALMTLQQHG 245
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6171> which encodes the amino acid sequence <SEQ ID 6172>. Analysis of this protein sequence reveals the following:

```
      Possible site: 61

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.22   Transmembrane   95 - 111 (  95 - 111)

----- Final Results -----
               bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
```

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

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

```
>GP:BAB07541 GB:AP001520 unknown conserved protein in B. subtilis
          [Bacillus halodurans]
 Identities = 97/228 (42%), Positives = 138/228 (59%), Gaps = 2/228 (0%)

Query: 30  ALSPFVWTEVFLKTINQEPNQLILHIWPMTRTVILGMLDRQLPYFELAKTEIGNNGYVPV 89
           AL  F + +    +I +  +   L  W    TV+LG+ D +LP  +     +   + +
Sbjct: 27  ALQSFAYDDTLCTSIGKSQSPPTLRAWVHHNTVVLGIQDSRLPQIKAGIEALKGFQHDVI 86

Query: 90  TRNIGGLAVVADDGILNFSLVIPDHFSESISISNAYLIMVDVIRESFSDYYQRIEYHEIK 149
            RN GGLAVV D GILN SLV+ +   + SI + Y +M ++I   F D+ ++IE  EI
Sbjct: 87  VRNSGGLAVVLDSGILNLSLVLKEE--KGFSIDDGYELMYELICSMFQDHREQIEAREIV 144

Query: 150 NSYCPGNFDLSIAGRKFAGIAQRRIKKGIVVSIYLSVCGDQAARGQLIKDFYEAGTQGEV 209
            SYCPG++DLSI G+KFAGI+QRRI+ G+ V IYL V G  A R ++I+ FY+    G+
Sbjct: 145 GSYCPGSYDLSIDGKKFAGISQRRIRGGVAVQIYLCVSGSGAERAKMIRTFYDKAVAGQP 204

Query: 210 TKVNYPQIDPECMATLSELLETPFTVAEVLERLRLTLRQLGFSLTEKS 257
           TK  YP+I PE MA+LSELL  P  V++VL +  +TL+Q G SL  +S
Sbjct: 205 TKFVYPRIKPETMASLSELLGQPHNVSDVLLKALMTLQQHGASLLTES 252
```

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

```
 Identities = 155/275 (56%), Positives = 199/275 (72%), Gaps = 8/275 (2%)

Query: 32  QDLAQLPVSIFKDYVTDAQDAEKPFIWTEVFLREINRSNQEIILHIWPMTKTVILGMLDR 91
           +DLA LP+ ++ D       A  PF+WTEVFL+ IN+   ++ILHIWPMT+TVILGMLDR
Sbjct: 10  RDLASLPIFVYGDGNKKVPGALSPFVWTEVFLKTINQEPNQLILHIWPMTRTVILGMLDR 69

Query: 92  ELPHLELAKKEIISRGYEPVVRNFGGLAVVADEGILNFSLVIPDVFERKLSISDGYLIMV 151
           +LP+ ELAK EI + GY PV RN GGLAVVAD+GILNFSLVIPD F   +SIS+ YLIMV
Sbjct: 70  QLPYFELAKTEIGNNGYVPVTRNIGGLAVVADDGILNFSLVIPDHFSESISISNAYLIMV 129

Query: 152 DFIRSIFSDFYQPIEHFEVETSYCPGKFDLSINGKKFAGLAQRRIKNGIAVSIYLSVCGD 211
           D IR  FSD+YQ IE+ E++ SYCPG FDLSI G+KFAG+AQRRIK GI VSIYLSVCGD
Sbjct: 130 DVIRESFSDYYQRIEYHEIKNSYCPGNFDLSIAGRKFAGIAQRRIKKGIVVSIYLSVCGD 189

Query: 212 QKGRSQMISDFYKIGLGDTGSPIAYPNVDPEIMANLSDLLDCPMTVEDVIDRMLISLKQV 271
           Q  R Q+I DFY+ G     + + YP +DPE MA LS+LL+ P TV +V++R+ ++L+Q+
Sbjct: 190 QAARGQLIKDFYEAGTQGEVTKVNYPQIDPECMATLSELLETPFTVAEVLERLRLTLRQL 249

Query: 272 GFN------DRLLMIRPDLVAEFNRFQAKSMANKG 300
           GF+      D+ L+   D V  + R Q + +  +G
Sbjct: 250 GFSLTEKSPDQALLTNFDAV--YERMQLEVVRKEG 282
```

A related GBS gene <SEQ ID 8951> and protein <SEQ ID 8952> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:    -16.85
GvH: Signal Score (-7.5): -5.07
     Possible site: 49
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -0.22 threshold:  0.0
   INTEGRAL    Likelihood = -0.22   Transmembrane  117 - 133 ( 117 - 133)
   PERIPHERAL  Likelihood =  0.47      73
 modified ALOM score:   0.54

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

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

```
ORF01564(451 - 1116 of 1518)
EGAD|13388|BS3758(27 - 249 of 281) hypothetical 31.4 kd protein in pta 3'region (Bacillus
subtilis) OMNI|NT01BS4391 hypothetical protein SP|P39648|YWFL_BACSU HYPOTHETICAL 31.4 KDA
PROTEIN IN PTA 3'REGION. GP|414014|emb|CAA51646.1||X73124 ipa-90d (Bacillus subtilis)
GP|2636300|emb|CAB15791.1||Z99123 alternate gene name: ipa-90d (Bacillus subtilis)
PIR|S39745|S39745 ywfL protein - Bacillus subtilis
%Match = 15.8
%Identity = 40.8  %Similarity = 61.0
Matches = 91  Mismatches = 82  Conservative Sub.s = 45

321       351       381       411       441       471       501       531
*WNLRETYWKISSDCDKINLAEFSRERMSDLLEWQDLAQLPVSIFKDYVTDAQDAEKPFIWTEVFLREINRSNQEIILHI
                                                    ||:: |   : :   : :
                                MANQPIDLLMQPKWRVIDQSSLGPLFDAKQSFAMDDTLCMSVGKGVSPATARS
                                          10        20        30        40        50

561       591       621       651       681       711       738       768
WPMTKTVILGMLDRELPHLELAKKEIISRGYEPVVRNFGGLAVVADEGILNFSLVIPDVFERK-LSISDGYLIMVDFIRS
|    |::||: |  || |:     : | ||  :||| |||||| |:|:|| ||:  |  |:| : |  ||  ||:::|
WVHHDTIVLGIQDTRLPFLQDGISLLESEGYRVIVRNSGGLAVVLDDGVLNISLIFED--EKKGIDIDKGYEAMVELMRR
                  70        80        90        100       110       120       130

798       828       858       888       918                 972       996
IFSDFYQPIEHFEVETSYCPGKFDLSINGKKFAGLAQRRIKNGIAVSIYLSVCGDQKG--RSQMISDFYKIGLGD--TGS
::  :   || :|:| ||||| :||||||||||::|||:: |:|| |||  | |: |  |: :|  ||:  | |
MLRPYNAKIEAYEIEGSYCPGSYDLSINGKKFAGISQRRVRGGVAVQIYL--CADKSGSERADLIRRFYQAALKDKQNDK
                  150       160       170       180       190       200

1026      1056      1086      1116      1146      1176      1206      1236
PIAYPNVDPEIMANLSDLLDCPMTVEDVIDRMLISLKQVGFNDRLLMIRPDLVAEFNRFQAKSMANKGMVSRDE*CPR*F
   || : || ||:||:||   ::|:|::  :|  || :
KGVYPEIRPETMASLSELLQKDISVQDLMFALLTELKALSTHLYSAGLSIDEEMEFEKNLVRMAERNAKVFG
         220       230       240       250       260       270       280
```

SEQ ID 8952 (GBS390) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 7; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 82 (lane 3; MW 62kDa).

GBS390-GST was purified as shown in Figure 216, lane 12.

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

Example 1998

A DNA sequence (GBSx2108) was identified in *S.agalactiae* <SEQ ID 6173> which encodes the amino acid sequence <SEQ ID 6174>. This protein is predicted to be probable trimethylamine dehydrogenase (nemA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2218(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:CAA83700 GB:Z33015 similar to trimethylamine DH [Mycoplasma
```

```
            capricolum]
 Identities = 162/311 (52%), Positives = 219/311 (70%), Gaps = 1/311 (0%)

Query: 3   NVQGNLFRPLTLPNGLSLENRFVLSPMVTNSSTSEGFVTDDDIAYAVRRAKSAPLQITGA 62
            N    LF P  L NG  LENRFVLSPM  + +T +G +TD +  Y  RR+ SAPLQITG
Sbjct: 2   NKYEKLFEPFYL-NGFKLENRFVLSPMTLSLATLDGKITDKEADYVKRRSHSAPLQITGG 60

Query: 63  AYITEYGQLFEYGFSVSKDEDIPGLTKLAKAMKSKGAKAVLQLTHAGRFSSHTLARHGYV 122
             Y  E+GQLFEYG S   D+DIP LT+L + MK+      +LQL HAG+FS  +L ++GY+
Sbjct: 61  VYFDEFGQLFEYGISAKSDDDIPSLTRLYQEMKTDSNCVILQLAHAGKFSKTSLKKYGYL 120

Query: 123 YGPSPMQLQSPYPHQVKELTHKDILRIIDEYVQATRRAIQAGFDGVEISSAQRLLIQTFF 182
            YGPS  +  +P  H+V EL  + I +II +Y  AT R I+AGF+G+EIS AQRLLIQTFF
Sbjct: 121 YGPSYEKNHTPIEHEVLELPKEKIKQIIQDYKDATLRVIKAGFNGIEISMAQRLLIQTFF 180

Query: 183 STFSNQRKDEYGPQTLTNRCRLGLEVFKAVQKVIREEAESDFILGFRATPEETRGSQIGY 242
            S   N+R DEY      NR R  LEV KA+++VI + A  +FI GFRATPEET G  +GY
Sbjct: 181 SQIINKRTDEYSATNFENRSRFCLEVVKAIREVIDKYAPKNFIFGFRATPEETYGDILGY 240

Query: 243 SIEEFMEFLEKILAIAQVDYLAIASWGHDVFRNTIRSEGVYKGQLVNQVIFEHFGDRVPI 302
            +IE+F++  ++KI+ I ++ YLAIASWGHD++ N +RS   YKGQLVN+VI++ + +++PI
Sbjct: 241 TIEDFIQLVDKIIEIGKISYLAIASWGHDIYLNKVRSNTKYKGQLVNKVIYDIYKNKLPI 300

Query: 303 MATGGINSASK 313
            +++GGIN+ +K
Sbjct: 301 ISSGGINTPTK 311
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6175> which encodes the amino acid sequence <SEQ ID 6176>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3055(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 265/390 (67%), Positives = 321/390 (81%)

Query: 8   LFRPLTLPNGLSLENRFVLSPMVTNSSTSEGFVTDDDIAYAVRRAKSAPLQITGAAYITE 67
            LF PLTLPNG  L+NRFVLSPMVTNSST +G+VT DD++YA+RRA SAPLQITGAAY+
Sbjct: 8   LFEPLTLPNGSQLDNRFVLSPMVTNSSTKDGYVTQDDVSYALRRAASAPLQITGAAYVDP 67

Query: 68  YGQLFEYGFSVSKDEDIPGLTKLAKAMKSKGAKAVLQLTHAGRFSSHTLARHGYVYGPSP 127
            YGQLFEYGFSV+KD DI GL +LA+AMK+KGAKAVLQLTHAGRF+SH L ++G+VYGPS
Sbjct: 68  YGQLFEYGFSVTKDADISGLKELAQAMKAKGAKAVLQLTHAGRFASHALTKYGFVYGPSY 127

Query: 128 MQLQSPYPHQVKELTHKDILRIIDEYVQATRRAIQAGFDGVEISSAQRLLIQTFFSTFSN 187
            MQL+SP PH+VK LT + I  +I  Y QATRRAIQAGFDGVE+SSAQRLLIQTFFSTFSN
Sbjct: 128 MQLRSPQPHEVKPLTGQQIEELIAAYAQATRRAIQAGFDGVEVSSAQRLLIQTFFSTFSN 187

Query: 188 QRKDEYGPQTLTNRCRLGLEVFKAVQKVIREEAESDFILGFRATPEETRGSQIGYSIEEF 247
            +R D YG QTL NR +L L V +AVQ+VI++EA   FI GFRATPEETRG+ IGYSI+EF
Sbjct: 188 KRTDSYGCQTLFNRSKLTLAVLQAVQQVIKQEAPDGFIFGFRATPEETRGNDIGYSIDEF 247

Query: 248 MEFLEKILAIAQVDYLAIASWGHDVFRNTIRSEGVYKGQLVNQVIFEHFGDRVPIMATGG 307
            ++ ++ +L +A++DYLAIASWG  VFRNT+RS G Y G+ VNQV+ ++  +++P+MATGG
Sbjct: 248 LQLMDWVLNVAKLDYLAIASWGRHVFRNTVRSPGPYYGRRVNQVVRDYLRNKLPVMATGG 307

Query: 308 INSASKVFEALQHAHMIGASTPLVVDPEFLQKIKAKCSDQINLRIKVSDLEGLAIPKASF 367
            +N+  K  EAL HA  IG STP VVDPEF  KIK  C + I+LRI+ +DL+ LAIP+ASF
Sbjct: 308 MNTPDKAIEALAHADFIGVSTPFVVDPEFAHKIKEGCEESIHLRIRPADLKSLAIPQASF 367

Query: 368 KDIVPLMDYGESLPKEAREVFRELRSNYRE 397
```

```
          KDIVPLMDYGESLPKE+R +FR L  NY+E
Sbjct: 368 KDIVPLMDYGESLPKESRTLFRSLTHNYKE 397
```

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

Example 1999

A DNA sequence (GBSx2109) was identified in *S.agalactiae* <SEQ ID 6177> which encodes the amino acid sequence <SEQ ID 6178>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3748(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:BAB04594 GB:AP001510 unknown conserved protein [Bacillus halodurans]
 Identities = 121/333 (36%), Positives = 192/333 (57%), Gaps = 12/333 (3%)

Query: 1   MKLSVLDYGLIDYGKTASDAIQETILLSQEAERLGYHQFWVAEHHGVKAFSISNPELMIM 60
           MKLSVLD   I YG  A +A+++T  L++  E LGYH+FWV+EHH      + S+PE++I
Sbjct: 1   MKLSVLDQSPIAYGSNAKEALRQTTELAKVTEALGYHRFWVSEHHDASTLAGSSPEVLIA 60

Query: 61  HLANQTKSIKIGSGGIMPLHYSSFKLAETLKTLETCHPNRVSIGLGNSLGTVKVSNALRS 120
           HLA  TK I++GSGG+M  HYS++K+AE  K LE  HP R+ +GLG + G + ++
Sbjct: 61  HLAAHTKKIRLGSGGVMLPHYSAYKVAENFKLLEALHPGRIDVGLGRAPGGMPIAKMALQ 120

Query: 121 LHK---AHDYEEVLEELKSWLIDESSSKEPL----VQPTLSSFPDLYVLGSGQKSAYLAA 173
             K    H Y   ++++  +L D+  +           P + + PD++LGS   SA +AA
Sbjct: 121 EGKEQNIHKYPLQVKDVIGYLQDDLPTDHRFHGLKATPLIDTVPDVWLLGSSGGSANVAA 180

Query: 174 KLGLGFTFGVFPFMDKDPLTEAKKLSSLYYHQFEEYYPNKSPNLMVAAFVVIADTSEEAE 233
           + G GF F    F++ +   +A +    Y   F+        P   VA FV+ ADT E+A+
Sbjct: 181 ENGTGFAFA--HFINGEGGVQAVE---SYRETFQPSALFDRPQTSVAIFVICADTDEQAD 235

Query: 234 NIAKTLDIWMLGNKDFNEFATFPTIEEANHYQLTPEQKAKIKSNRDRMIVGDPKQVKESL 293
            IA +LD+ ++  ++        P+IE A  Y  +P ++A+I+ NR RMIVG PK V++ L
Sbjct: 236 QIASSLDLSLIMLENGQLSKGTPSIESALSYPYSPFERARIRENRKRMIVGSPKAVRQQL 295

Query: 294 DALVNASQAEELLLIPLVPGLDQRIKSLKLLSQ 326
             L  A + EE++++ +    + RI+S +LL +
Sbjct: 296 VELARAYETEEVIVVTITHRFEDRIRSYELLGE 328
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6179> which encodes the amino acid sequence <SEQ ID 6180>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60   Transmembrane  212 - 228 ( 210 - 229)

----- 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>
```

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

```
 Identities = 173/329 (52%), Positives = 241/329 (72%), Gaps = 1/329 (0%)

Query: 1   MKLSVLDYGLIDYGKTASDAIQETILLSQEAERLGYHQFWVAEHHGVKAFSISNPELMIM 60
```

```
            MK+S+LDYG+ID  KT  +A+ ET  L+Q A++LG+H+FWVAEHH + AF+IS+PEL++M
Sbjct: 1    MKVSILDYGVIDKEKTPQEALLETRCLAQVADKLGFHRFWVAEHHNIYAFAISSPELLMM 60

Query: 61   HLANQTKSIKIGSGGIMPLHYSSFKLAETLKTLETCHPNRVSIGLGNSLGTVKVSNALRS 120
            HLA+ TK I+IGSGGIMPLHYSSFK+AE + TLE  HPNR+ +G+GNSLGT  V  AL S
Sbjct: 61   HLADHTKQIRIGSGGIMPLHYSSFKIAEWIMTLEALHPNRIDLGIGNSLGTTLVQRALSS 120

Query: 121  LHKAHDYEEVLEELKSWLIDESSSKEPL-VQPTLSSFPDLYVLGSGQKSAYLAAKLGLGF 179
             +H    Y +V+ EL  +L  +  S  P+ V P  +++P ++ L +  ++A LA +LGLG+
Sbjct: 121  IHCKDSYSQVVTELYQYLNPDHLSPLPIFVNPRGNTYPQIWTLSNSLETAELAGQLGLGY 180

Query: 180  TFGVFPFMDKDPLTEAKKLSSLYYHQFEEYYPNKSPNLMVAAFVVIADTSEEAENIAKTL 239
            TFG+FP++ KDP+TEAK++S+ Y   F      K P L++A F+V++DT E+AE +AK L
Sbjct: 181  TFGIFPYIPKDPITEAKRVSAHYRKAFRPSKLLKIPKLILAVFIVLSDTDEKAEALAKPL 240

Query: 240  DIWMLGNKDFNEFATFPTIEEANHYQLTPEQKAKIKSNRDRMIVGDPKQVKESLDALVNA 299
            DIWMLG +DFNEF T+P +EEA +Y LT +Q+  I +NR RM++G P  VK+ LD L+ A
Sbjct: 241  DIWMLGQQDFNEFKTYPDVEEARNYHLTEKQREAIAANRSRMVIGSPHTVKKQLDRLIEA 300

Query: 300  SQAEELLLIPLVPGLDQRIKSLKLLSQLY 328
             QA+ELL IPLVP     R ++L+LL+ LY
Sbjct: 301  CQADELLAIPLVPEFANRQRTLELLADLY 329
```

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

Example 2000

A DNA sequence (GBSx2110) was identified in *S.agalactiae* <SEQ ID 6181> which encodes the amino acid sequence <SEQ ID 6182>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2384(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAF81345 GB:AC007767 Identical to a glycine cleavage system
          H-protein precursor from Arabidopsis thaliana gb|P25855.
          It contains a glycine cleavage H-protein domain
          PF|01597. ESTs gb|R90208, gb|AI
 Identities = 30/91 (32%), Positives = 53/91 (57%), Gaps = 1/91 (1%)

Query: 18   TISLTPELQDDLGTVGYVEFTD-DANLEVDDVILNIEASKTVMAILSPLTGKVVKVNTAA 76
            TI +T   QD LG V +VE  + ++++  +     +E+ K    ILSP++G+V++VNT
Sbjct: 59   TIGITDHAQDHLGEVVFVELPEANSSVSKEKSFGAVESVKATSEILSPISGEVIEVNTKL 118

Query: 77   SQEPTLLNSEKADENWLVVLTEVDYAAFEAL 107
            ++ P L+NS   ++ W++ +     A  EAL
Sbjct: 119  TESPGLINSSPYEDGWMIKVKPSSPAELEAL 149
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6183> which encodes the amino acid sequence <SEQ ID 6184>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3544(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 80/110 (72%), Positives = 98/110 (88%)

Query: 1   MKKIANYLLIEKNEELYTISLTPELQDDLGTVGYVEFTDDANLEVDDVILNIEASKTVMA 60
           MKKIANYLLIEK ++ YTIS+TPELQDD+GT+GY EFTD+ +L VDD+ILN+EASKTVM+
Sbjct: 1   MKKIANYLLIEKTDDRYTISMTPELQDDIGTIGYAEFTDNDHLAVDDIILNLEASKTVMS 60

Query: 61  ILSPLTGKVVKVNTAASQEPTLLNSEKADENWLVVLTEVDYAAFEALENA 110
           +LSPL G VV+ N AA+  PTLLNSEKA+ENW+VVLT+VD AAF+ALE+A
Sbjct: 61  VLSPLAGAVVERNEAATLTPTLLNSEKAEENWIVVLTDVDQAAFDALEDA 110
```

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

Example 2001

A DNA sequence (GBSx2111) was identified in *S.agalactiae* <SEQ ID 6185> which encodes the amino acid sequence <SEQ ID 6186>. This protein is predicted to be LRP16 (b1045). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0608(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:AAF15294 GB:AF202922 LRP16 [Homo sapiens]
 Identities = 73/171 (42%), Positives = 98/171 (56%), Gaps = 13/171 (7%)

Query: 88  DICLLQVDAIVNAANSKLLGCFIPNHHCIDNQIHTFAGSRLRLACHQLMTQQGRMEAVGQ 147
           DI  L+VDAIVNAANS LLG         +D  IH  AG  L   C  L + +      G+
Sbjct: 78  DITKLEVDAIVNAANSSLLG-----GGGVDGCIHRAAGPLLTDECRTLQSCK-----TGK 127

Query: 148 AKLTESYHLPCKYVIHTVGPYVKVDQKPSRIREDLLKSSYKSCLQLAVRANLKTIVFPCI 207
           AK+T  Y LP KYVIHTVGP    +   S+  E  L+S Y S L L +   L+++ FPCI
Sbjct: 128 AKITGGYRLPAKYVIHTVGPIAYGEPSASQAAE--LRSCYLSSLDLLLEHRLRSVAFPCI 185

Query: 208 STGEFGFPNQRAAELAVQAILEWQRENQHKL-YIIFNTFTPKDQDIYQKLL 257
           STG FG+P + AAE+ +  + EW  +++ K+  +I   F  KD+DIY+  L
Sbjct: 186 STGVFGYPCEAAAEIVLATLREWLEQHKDKVDRLIICVFLEKDEDIYRSRL 236
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6187> which encodes the amino acid sequence <SEQ ID 6188>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1992(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 = 139/266 (52%), Positives = 178/266 (66%), Gaps = 6/266 (2%)

Query: 1   MPNQKQLLLAMIEYLQSEKLTDVDDL----RTTDLQTVWRGLVNQQDPQNISQEYLSLED 56
           MP+   LL  MI  LQ+E+LT          T  Q +WR L+NQ+    +S++YL+LED
```

```
Sbjct: 1   MPSSFDLLGEMIGLLQTEQLTSSWACPLPNALTKRQDLWRALINQRPALPLSKDYLNLED 60

Query: 57  RYLSHWWNTQKVKTIDVCHQTVYSNVFTYHGDICLLQVDAIVNAANSKLLGCFIPNHHCI 116
           YL  W  +    ++  C +T Y+++F YHGDI  L VDAIVNAANS+LLGCF PNH CI
Sbjct: 61  AYLDDWRASFVPVSVKDCQKTNYTSLFLYHGDIRYLAVDAIVNAANSELLGCFSPNHGCI 120

Query: 117 DNQIHTFAGSRLRLACHQLMTQQGRMEAVGQAKLTESYHLPCKYVIHTVGPYVKVDQKPS 176
           DN IHTFAGSRLRLAC  +MT+QGR EA+GQAKLT +YHLP  Y+IHTVGP +      S
Sbjct: 121 DNAIHTFAGSRLRLACQAIMTEQGRKEAIGQAKLTSAYHLPASYIIHTVGPRITKGHHVS 180

Query: 177 RIREDLLKSSYKSCLQLAVRANLKTIVFPCISTGEFGFPNQRAAELAVQAILEWQRENQH 236
           IR DLL   Y+S L LAV+A L ++ F  ISTGEFGFP + AA++A++ +L+WQ E+
Sbjct: 181 PIRADLLARCYRSSLDLAVKAGLTSLAFCSISTGEFGFPKKEAAQIAIKTVLKWQAEHPE 240

Query: 237 K--LYIIFNTFTPKDQDIYQKLLLKE 260
             L  IFNTFT +D+ +Y   L KE
Sbjct: 241 SKTLTTIFNTFTSEDKALYDTYLQKE 266
```

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

Example 2002

A DNA sequence (GBSx2112) was identified in *S.agalactiae* <SEQ ID 6189> which encodes the amino acid sequence <SEQ ID 6190>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2171(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 6191> which encodes the amino acid sequence <SEQ ID 6192>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2477(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 = 218/284 (76%), Positives = 250/284 (87%)

Query: 4   WKTLEKTNHSQSEILSQLIEESDAIVVGIGAGMSAADGFTYIGPRFEEAFPDFIAKYQLL 63
           W T  + N +Q+E L+QLI+E+DA+VVGIGAGMSAADGFTYIG RFE AFPDFIAKYQ L
Sbjct: 4   WTTYPQKNLTQAEQLAQLIKEADALVVGIGAGMSAADGFTYIGSRFETAFPDFIAKYQFL 63

Query: 64  DMLQASLYDFEDWEEYWAFQSRFVALNYLDQPVGQAYLDLKDILAKKEYHIITTNADNAF 123
           DMLQASL+DFEDW+EYWAFQSRFVALNYLDQPVGQ+YLDLK+IL  K+YHIITTNADNAF
Sbjct: 64  DMLQASLFDFEDWQEYWAFQSRFVALNYLDQPVGQSYLDLKEILGTKDYHIITTNADNAF 123

Query: 124 AVADYNLEKVFHIQGEYGLWQCSQHCHQQTYRNDQAIRQMIAQQKDMKIPSNLIPKCPKC 183
            VA Y+   +FHIQGEYGLWQCSQHCHQQTY++D  IRQMIA+QK+MK+P  LIP CP+C
Sbjct: 124 WVAGYDPHNIFHIQGEYGLWQCSQHCHQQTYKDDTVIRQMIAEQKNMKVPGQLIPHCPEC 183

Query: 184 DQPFEINKRNEEKGMVEDADFHAQRQRYENFLSQHQNDKVLYLEIGVGHTTPQFIKHPFW 243
           + PFEINKRNEEKGMVEDADFHAQ+ RYE FLS+H+  KVLYLEIGVGHTTPQFIKHPFW
```

```
Sbjct: 184 EAPFEINKRNEEKGMVEDADFHAQKARYEAFLSEHKEGKVLYLEIGVGHTTPQFIKHPFW 243

Query: 244 RFVSLNENSLFVTLNHKHYRIPQKIRSRSVQLTQHIAELIAEAK 287
           + VS N N+LFVTLNHKHYRIP  IR +S++LT+HIA+LI+  K
Sbjct: 244 KRVSENPNALFVTLNHKHYRIPLSIRRQSLELTEHIAQLISATK 287
```

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

Example 2003

A DNA sequence (GBSx2113) was identified in *S.agalactiae* <SEQ ID 6193> which encodes the amino acid sequence <SEQ ID 6194>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1086(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:CAB12865 GB:Z99109 similar to lipoate-protein ligase [Bacillus subtilis]
 Identities = 130/331 (39%), Positives = 206/331 (61%), Gaps = 5/331 (1%)

Query: 9   NGKRITDGAIALAMQVYILQNVFLDDDILFPYYCDPKVEIGKFQNAVIETNQEYLKEHDI 68
           + + I D  I LA++ Y ++++  +   L  Y   P + IGK QN + E N +Y++E+ I
Sbjct: 5   DNQNINDPRINLAIEEYCVKHLDPEQQYLLFYVNQPSIIIGKNQNTIEEINTKYVEENGI 64

Query: 69  PVVRRDTGGGAVYVDSGAVNICYLMKDHGQ-FGDFKRAYEPAIKALKTLGASSVEMRERN 127
            VVRR +GGGAVY D G +N  ++ KD G  F +FK+  EP I+AL  LG  + E+  RN
Sbjct: 65  IVVRRLSGGGAVYHDLGNLNFSFITKDDGDSFHNFKKFTEPVIQALHQLGVEA-ELSGRN 123

Query: 128 DLVIDGKKVSGAAMTIVNGRIYGGYSLLLDVDFDAMEKVLNPNRKKIESKGIKSVRSRVG 187
           D+V+DG+K+SG A     GRI+   +L+ D   D +   L   + KIESKGIKS+RSRV
Sbjct: 124 DIVVDGRKISGNAQFATKGRIFSHGTLMFDSAIDHVVSALKVKKDKIESKGIKSIRSRVA 183

Query: 188 DIRSHLSEDYRHITTDQFKDLMVCQLLHIDHIDQAKRYHLTEKDWAAIDALADEKYKNWD 247
           +I   L +    +TT++F+  ++  + + + +     Y LTEKDW  I  ++ E+Y+NWD
Sbjct: 184 NISEFLDDK---MTTEEFRSHLLRHIFNTNDVGNVPEYKLTEKDWETIHQISKERYQNWD 240

Query: 248 WNYGNSPQYSYHRDARFPSGTYDFHLEIEKGIITNCRIYGDFFSSKDISDIENLLIGCPM 307
           WNYG SP+++ +   R+P G+ D HLE++KG I +C+I+GDFF   D+S+IENLL+G
Sbjct: 241 WNYGRSPKFNLNHSKRYPVGSIDLHLEVKKGKIEDCKIFGDFFGVGDVSEIENLLVGKQY 300

Query: 308 KEELVLEKLSTLSLEDYFGQTSPEEIKAVLF 338
           +  ++ + L  ++L+ YFG  + E+   +++
Sbjct: 301 ERSVIADVLEGVNLKHYFGNITKEDFLDLIY 331
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6195> which encodes the amino acid sequence <SEQ ID 6196>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0939(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 = 248/339 (73%), Positives = 283/339 (83%)
```

```
Query: 1   MYLIEPIRNGKRITDGAIALAMQVYILQNVFLDDDILFPYYCDPKVEIGKFQNAVIETNQ 60
           MYLIEPIRNGKRITDGA+ALAMQVY+ +N+FLDDDILFPYYCDPKVEIGKFQNAV+ETNQ
Sbjct: 1   MYLIEPIRNGKRITDGAVALAMQVYVQENLFLDDDILFPYYCDPKVEIGKFQNAVVETNQ 60

Query: 61  EYLKEHDIPVVRRDTGGGAVYVDSGAVNICYLMKDHGQFGDFKRAYEPAIKALKTLGASS 120
           EYLKEH IPVVRRDTGGGAVYVDSGAVNICYL+ D+G FGDFKR Y+PAI+AL  LGA+
Sbjct: 61  EYLKEHHIPVVRRDTGGGAVYVDSGAVNICYLINDNGIFGDFKRTYQPAIEALHHLGATE 120

Query: 121 VEMRERNDLVIDGKKVSGAAMTIVNGRIYGGYSLLLDVDFDAMEKVLNPNRKKIESKGIK 180
           VEM  RNDLVIDGKKVSGAAMTI NGR+YGGYSLLLDVDF+AMEK L PNRKKIESKGI+
Sbjct: 121 VEMSGRNDLVIDGKKVSGAAMTIANGRVYGGYSLLLDVDFEAMEKALKPNRKKIESKGIR 180

Query: 181 SVRSRVGDIRSHLSEDYRHITTDQFKDLMVCQLLHIDHIDQAKRYHLTEKDWAAIDALAD 240
           SVRSRVG+IR HL+  Y+ IT ++FKDLMVCQLL I+ I QAKRY LTEKDW  IDAL +
Sbjct: 181 SVRSRVGNIREHLAPQYQGITIEEFKDLMVCQLLQIETISQAKRYDLTEKDWQQIDALTE 240

Query: 241 EKYKNWDWNYGNSPQYSYHRDARFPSGTYDFHLEIEKGIITNCRIYGDFFSSKDISDIEN 300
            KY NW+WNYGN+PQY YHRD RF  GT D HL+I+KG I  CRIYGDFF   DI+++E
Sbjct: 241 RKYHNWEWNYGNAPQYRYHRDGRFTGGTVDIHLDIKKGYIAACRIYGDFFGKADIAELEG 300

Query: 301 LLIGCPMKEELVLEKLSTLSLEDYFGQTSPEEIKAVLFS 339
            LIG  M++E VL  L+ + L  Y G  + EE+  ++FS
Sbjct: 301 HLIGTRMEKEDVLATLNAIDLAPYLGAITAEELGDLIFS 339
```

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

Example 2004

A DNA sequence (GBSx2114) was identified in *S.agalactiae* <SEQ ID 6197> which encodes the amino acid sequence <SEQ ID 6198>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.49   Transmembrane  196 - 212 ( 196 - 212)

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

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

```
>GP:AAB49329 GB:U39612 formyl-tetrahydrofolate synthetase
          [Streptococcus mutans]
 Identities = 432/556 (77%), Positives = 493/556 (87%)

Query: 1   MKTDIEIAQSVALKPIAEIVEQVGIGFDDIELYGKYKAKLSFDKIEAVKSQKVGKLILVT 60
           MKTDIEIAQSV L+PI  +V+++GI FDD+ELYGKYKAKL+FDKI+AV+   GKL+LVT
Sbjct: 1   MKTDIEIAQSVDLRPITNVVKKLGIDFDDLELYGKYKAKLTFDKIKAVEENAPGKLVLVT 60

Query: 61  AINPTPAGEGKSTMSIGLADALNKIGKKTMIALREPSLGPVMGIKGGAAGGGYAQVLPME 120
           AINPTPAGEGKST++IGLADALNKIGKKTMIA+REPSLGPVMGIKGGAAGGGYAQVLPME
Sbjct: 61  AINPTPAGEGKSTITIGLADALNKIGKKTMIAIREPSLGPVMGIKGGAAGGGYAQVLPME 120

Query: 121 DINLHFTGDMHAITTANNALSALLDNHIHQGNELDIDQRRVIWKRVVDLNDRALRQVIVG 180
           DINLHFTGDMHAITTANNALSAL+DNH+HQGNEL IDQRR+IWKRVVDLNDRALR V VG
Sbjct: 121 DINLHFTGDMHAITTANNALSALIDNHLHQGNELGIDQRRIIWKRVVDLNDRALRHVTVG 180

Query: 181 LGSPVNGIPREDGFDITVASEIMAILCLATDLSDLKKRLSNIVVAYSRNRKPIYVKDLKI 240
           LGSP+NGIPREDGFDITVASEIMAILCLAT++ DLK+RL+NIV+ Y  +R P+YV+DL++
Sbjct: 181 LGSPINGIPREDGFDITVASEIMAILCLATNVEDLKERLANIVIGYRFDRSPVYVRDLEV 240

Query: 241 EGALTLILKDTIKPNLVQTIYGTPALVHGGPFANIAHGCNSVLATSTALRLADYVVTEAG 300
           +GAL LILK+ IKPNLVQTIYGTPA VHGGPFANIAHGCNSVLATSTALRLADY +TEAG
```

```
Sbjct: 241 QGALALILKEAIKPNLVQTIYGTPAFVHGGPFANIAHGCNSVLATSTALRLADYTITEAG 300

Query: 301 FGADLGAEKFLDIKTPNLPTSPDAIVIVATLRALKMHGGVSKEDLSQENVEAVKRGFTNL 360
           FGADLGAEKFLDIK PNLPTSPDA+VIVAT+RALKM+GGV+K+ L+QENVEAVK GF NL
Sbjct: 301 FGADLGAEKFLDIKAPNLPTSPDAVVIVATIRALKMNGGVAKDALNQENVEAVKAGFANL 360

Query: 361 ERHVNNMRQYGVPVVVAINQFTADTESEIATLKTLCSNIDVAVELASVWEDGADGGLELA 420
            RHV NMR+YGVPVVVAIN+F  DT  EIA L+ LC+ IDV VELASVW +GADGG++LA
Sbjct: 361 ARHVENMRKYGVPVVVAINEFITDTNDEIAVLRNLCAAIDVPVELASVWANGADGGVDLA 420

Query: 421 QTVANVIETQSSNYKRLYNDEDTIEEKIKKIVTKIYGGNKVHFGPKAQIQLKEFSDNGWD 480
            T+ N IE   S+YKRLY++  ++EEK+ +I  +IY  +KV F  KA+ Q+ +   NGWD
Sbjct: 421 NTLINTIENNPSHYKRLYDNNLSVEEKVTEIAKEIYRADKVIFEKKAKTQIAQIVKNGWD 480

Query: 481 KMPICMAKTQYSFSDNPNLLGAPTDFDITVREFVPKTGAGFIVALTGDVLTMPGLPKKPA 540
           +PICMAKTQYSFSD+P LLGAPT FDIT+RE VPK GAGFIVALTGDV+TMPGLPKKPA
Sbjct: 481 NLPICMAKTQYSFSDDPKLLGAPTGFDITIRELVPKLGAGFIVALTGDVMTMPGLPKKPA 540

Query: 541 ALNMDVLEDGTAIGLF 556
           ALNMDV  DGTA+GLF
Sbjct: 541 ALNMDVAADGTALGLF 556
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6199> which encodes the amino acid sequence <SEQ ID 6200>. Analysis of this protein sequence reveals the following:

```
     Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.49   Transmembrane  196 - 212 ( 196 - 212)

----- 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:AAB49329 GB:U39612 formyl-tetrahydrofolate synthetase
           [Streptococcus mutans]
 Identities = 432/556 (77%), Positives = 490/556 (87%)

Query: 1   MKSDIEIAQSVALQPITDIVKKVGIDGDDIELYGKYKAKLSFEKMKAVEANEPGKLILVT 60
           MK+DIEIAQSV L+PIT++VKK+GID DD+ELYGKYKAKL+F+K+KAVE N PGKL+LVT
Sbjct: 1   MKTDIEIAQSVDLRPITNVVKKLGIDFDDLELYGKYKAKLTFDKIKAVEENAPGKLVLVT 60

Query: 61  AINPTPAGEGKSTMSIGLADALNQMGKKTMLALREPSLGPVMGIKGGAAGGGYAQVLPME 120
           AINPTPAGEGKST++IGLADALN++GKKTM+A+REPSLGPVMGIKGGAAGGGYAQVLPME
Sbjct: 61  AINPTPAGEGKSTITIGLADALNKIGKKTMIAIREPSLGPVMGIKGGAAGGGYAQVLPME 120

Query: 121 DINLHFTGDMHAITTANNALSALIDNHLQQGNDLGIDPRRIIWKRVLDLNDRALRQVIVG 180
           DINLHFTGDMHAITTANNALSALIDNHL QGN+LGID RRIIWKRV+DLNDRALR V VG
Sbjct: 121 DINLHFTGDMHAITTANNALSALIDNHLHQGNELGIDQRRIIWKRVVDLNDRALRHVTVG 180

Query: 181 LGSPVNGVPREDGFDITVASEIMAILCLATDLKDLKKRLADIVVAYTYDRKPVYVRDLKV 240
           LGSP+NG+PREDGFDITVASEIMAILCLAT+++DLK+RLA+IV+ Y +DR PVYVRDL+V
Sbjct: 181 LGSPINGIPREDGFDITVASEIMAILCLATNVEDLKERLANIVIGYRFDRSPVYVRDLEV 240

Query: 241 EGALTLILKDAIKPNLVQTIYGTPALIHGGPFANIAHGCNSVLATSTALRLADYTVTEAG 300
           +GAL LILK+AIKPNLVQTIYGTPA +HGGPFANIAHGCNSVLATSTALRLADYT+TEAG
Sbjct: 241 QGALALILKEAIKPNLVQTIYGTPAFVHGGPFANIAHGCNSVLATSTALRLADYTITEAG 300

Query: 301 FGADLGAEKFLNIKVPNLPKAPDAIVIVATLRALKMHGGVAKSDLAAENCEAVRLGFANL 360
           FGADLGAEKFL+IK PNLP +PDA+VIVAT+RALKM+GGVAK  L  EN EAV+ GFANL
Sbjct: 301 FGADLGAEKFLDIKAPNLPTSPDAVVIVATIRALKMNGGVAKDALNQENVEAVKAGFANL 360

Query: 361 KRHVENMRQFKVPVVVAINEFVADTEAEIATLKALCEEIKVPVELASVWANGAEGGLALA 420
           KRHVENMR++ VPVVVAINEF+ DT  EIA L+ LC  I VPVELASVWANGA+GG+ LA
Sbjct: 361 ARHVENMRKYGVPVVVAINEFITDTNDEIAVLRNLCAAIDVPVELASVWANGADGGVDLA 420
```

```
Query: 421 KTVVRVIDQEAADYKRLYSDEDTLEEKVINIVTQIYGGKAVQFGPKAKTQLKQFAEFGWD 480
           T++  I+   + YKRLY +  ++EEKV  I  +IY    V F  KAKTQ+ Q  + GWD
Sbjct: 421 NTLINTIENNPSHYKRLYDNNLSVEEKVTEIAKEIYRADKVIFEKKAKTQIAQIVKNGWD 480

Query: 481 KLPVCMAKTQYSFSDNPSLLGAPTDFDITIREFVPKTGAGFIVGLTGDVMTMPGLPKVPA 540
            LP+CMAKTQYSFSD+P LLGAPT FDITIRE VPK GAGFIV LTGDVMTMPGLPK PA
Sbjct: 481 NLPICMAKTQYSFSDDPKLLGAPTGFDITIRELVPKLGAGFIVALTGDVMTMPGLPKKPA 540

Query: 541 AMAMDVAENGTALGLF 556
           A+ MDVA +GTALGLF
Sbjct: 541 ALNMDVAADGTALGLF 556
```

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

```
 Identities = 452/556 (81%), Positives = 513/556 (91%)

Query: 1   MKTDIEIAQSVALKPIAEIVEQVGIGFDDIELYGKYKAKLSFDKIEAVKSQKVGKLILVT 60
           MK+DIEIAQSVAL+PI +IV++VGI  DDIELYGKYKAKLSF+K++AV++ + GKLILVT
Sbjct: 1   MKSDIEIAQSVALQPITDIVKKVGIDGDDIELYGKYKAKLSFEKMKAVEANEPGKLILVT 60

Query: 61  AINPTPAGEGKSTMSIGLADALNKIGKKTMIALREPSLGPVMGIKGGAAGGGYAQVLPME 120
           AINPTPAGEGKSTMSIGLADALN++GKKTM+ALREPSLGPVMGIKGGAAGGGYAQVLPME
Sbjct: 61  AINPTPAGEGKSTMSIGLADALNQMGKKTMLALREPSLGPVMGIKGGAAGGGYAQVLPME 120

Query: 121 DINLHFTGDMHAITTANNALSALLDNHIHQGNELDIDQRRVIWKRVVDLNDRALRQVIVG 180
           DINLHFTGDMHAITTANNALSAL+DNH+ QGN+L ID RR+IWKRV+DLNDRALRQVIVG
Sbjct: 121 DINLHFTGDMHAITTANNALSALIDNHLQQGNDLGIDPRRIIWKRVLDLNDRALRQVIVG 180

Query: 181 LGSPVNGIPREDGFDITVASEIMAILCLATDLSDLKKRLSNIVVAYSRNRKPIYVKDLKI 240
           LGSPVNG+PREDGFDITVASEIMAILCLATDL DLKKRL++IVVAY+ +RKP+YV+DLK+
Sbjct: 181 LGSPVNGVPREDGFDITVASEIMAILCLATDLKDLKKRLADIVVAYTYDRKPVYVRDLKV 240

Query: 241 EGALTLILKDTIKPNLVQTIYGTPALVHGGPFANIAHGCNSVLATSTALRLADYVVTEAG 300
           EGALTLILKD IKPNLVQTIYGTPAL+HGGPFANIAHGCNSVLATSTALRLADY VTEAG
Sbjct: 241 EGALTLILKDAIKPNLVQTIYGTPALIHGGPFANIAHGCNSVLATSTALRLADYTVTEAG 300

Query: 301 FGADLGAEKFLDIKTPNLPTSPDAIVIVATLRALKMHGGVSKEDLSQENVEAVKRGFTNL 360
           FGADLGAEKFL+IK PNLP +PDAIVIVATLRALKMHGGV+K DL+ EN EAV+ GF NL
Sbjct: 301 FGADLGAEKFLNIKVPNLPKAPDAIVIVATLRALKMHGGVAKSDLAAENCEAVRLGFANL 360

Query: 361 ERHVNNMRQYGVPVVVAINQFTADTESEIATLKTLCSNIDVAVELASVWEDGADGGLELA 420
           +RHV NMRQ+ VPVVVAIN+F ADTE+EIATLK LC  I V VELASVW +GA+GGL LA
Sbjct: 361 KRHVENMRQFKVPVVVAINEFVADTEAEIATLKALCEEIKVPVELASVWANGAEGGLALA 420

Query: 421 QTVANVIETQSSNYKRLYNDEDTIEEKIKKIVTKIYGGNKVHFGPKAQIQLKEFSDNGWD 480
           +TV  VI+ ++++YKRLY+DEDT+EEK+  IVT+IYGG  V FGPKA+ QLK+F++ GWD
Sbjct: 421 KTVVRVIDQEAADYKRLYSDEDTLEEKVINIVTQIYGGKAVQFGPKAKTQLKQFAEFGWD 480

Query: 481 KMPICMAKTQYSFSDNPNLLGAPTDFDITVREFVPKTGAGFIVALTGDVLTMPGLPKKPA 540
           K+P+CMAKTQYSFSDNP+LLGAPTDFDIT+REFVPKTGAGFIV LTGDV+TMPGLPK PA
Sbjct: 481 KLPVCMAKTQYSFSDNPSLLGAPTDFDITIREFVPKTGAGFIVGLTGDVMTMPGLPKVPA 540

Query: 541 ALNMDVLEDGTAIGLF 556
           A+ MDV E+GTA+GLF
Sbjct: 541 AMAMDVAENGTALGLF 556
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9057> which encodes amino acid sequence <SEQ ID 9058>. Analysis of this protein sequence reveals the following:

```
     Possible site: 13
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.49   Transmembrane  516 - 532 ( 516 - 533)

----- 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>
```

An alignment of the GAS and GBS sequences follows:

```
 Score =  604 bits (1540), Expect = e-174
 Identities = 304/555 (54%), Positives = 389/555 (69%), Gaps = 2/555 (0%)

Query: 4   SDIEIANSVTMEPISKVADQLGIDKEALCLYGKYKAKIDARQLVALKNKPDGKLILVTAI 63
           +DIEIA SV ++PI+++ +Q+GI  + + LYGKYKAK+   ++ A+K++  GKLILVTAI
Sbjct: 3   TDIEIAQSVALKPIAEIVEQVGIGFDDIELYGKYKAKLSFDKIEAVKSQKVGKLILVTAI 62

Query: 64  SPTPAGEGKTTTSVGLVDALSAIGKKAVIALREPSLXXXXXXXXXXXXXXXXXXXPMEDI 123
           +PTPAGEGK+T S+GL DAL+ IGKK +IALREPSL                   PMEDI
Sbjct: 63  NPTPAGEGKSTMSIGLADALNKIGKKTMIALREPSLGPVMGIKGGAAGGGYAQVLPMEDI 122

Query: 124 NLHFTGDFHAIGVANNLLAALIDNHIHHGNSLGIDSRRITWKRVVDMNDRQLRHIVDGLQ 183
           NLHFTGD HAI  ANN L+AL+DNHIH GN L ID RR+ WKRVVD+NDR LR ++ GL
Sbjct: 123 NLHFTGDMHAITTANNALSALLDNHIHQGNELDIDQRRVIWKRVVDLNDRALRQVIVGLG 182

Query: 184 GKVNGIPREDGYDITVASEIMAILCLSENISDLKARLEKIIIGYNYQGEPVTXXXXXXXX 243
             VNGIPREDG+DITVASEIMAILCL+ ++SDLK RL  I++ Y+   +P+
Sbjct: 183 SPVNGIPREDGFDITVASEIMAILCLATDLSDLKKRLSNIVVAYSRNRKPIYVKDLKIEG 242

Query: 244 XXXXXXXXXIHPNLVQTLEHTPALIHGGPFANIAHGCNSVLATKLALKYGDYAVTEAGFG 303
                  I PNLVQT+  TPAL+HGGPFANIAHGCNSVLAT  AL+  DY VTEAGFG
Sbjct: 243 ALTLILKDTIKPNLVQTIYGTPALVHGGPFANIAHGCNSVLATSTALRLADYVVTEAGFG 302

Query: 304 ADLGAEKFIDIKCRMSGLRPAAVVLVATIRALKMHGGVPKADLATENVQAVVDGLPNLDK 363
           ADLGAEKF+DIK       P A+V+VAT+RALKMHGGV K DL+ ENV+AV  G  NL++
Sbjct: 303 ADLGAEKFLDIKTPNLPTSPDAIVIVATLRALKMHGGVSKEDLSQENVEAVKRGFTNLER 362

Query: 364 HLANIQDVYGLPVVVAINKFPLDTDAELQAVYDACDKRGVDVVISDVWANGGAGGRELAE 423
           H+ N++  YG+PVVVAIN+F  DT++E+  +   C    V V ++ VW +G  GG ELA+
Sbjct: 363 HVNNMRQ-YGVPVVVAINQFTADTESEIATLKTLCSNIDVAVELASVWEDGADGGLELAQ 421

Query: 424 KVVTLAE-QDNQFRFVYEEDDSIETKLTKIVTKVYGGKGINLSSAAKRELADLERLGFGN 482
            V  + E Q + ++ +Y ++D+IE K+ KIVTK+YGG  ++    A+ +L +   G+
Sbjct: 422 TVANVIETQSSNYKRLYNDEDTIEEKIKKIVTKIYGGNKVHFGPKAQIQLKEFSDNGWDK 481

Query: 483 YPICMAKTQYSFSDDAKKLGAPTDFTVTISNLKVSAGAGFIVALTGAIMTMPGLPKVPAS 542
            PICMAKTQYSFSD+   LGAPTDF +T+      GAGFIVALTG ++TMPGLPK PA+
Sbjct: 482 MPICMAKTQYSFSDNPNLLGAPTDFDITVREFVPKTGAGFIVALTGDVLTMPGLPKKPAA 541

Query: 543 ETIDIDEEGNITGLF 557
             +D+ E+G   GLF
Sbjct: 542 LNMDVLEDGTAIGLF 556
```

SEQ ID 6198 (GBS131) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 6; MW 64.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 4; MW 90kDa).

GBS131-GST was purified as shown in Figure 201, lane 5.

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

Example 2005

A DNA sequence (GBSx2115) was identified in *S.agalactiae* <SEQ ID 6201> which encodes the amino acid sequence <SEQ ID 6202>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.03   Transmembrane   34 -  50 (  29 -  56)
```

```
    INTEGRAL     Likelihood = -7.70   Transmembrane   90 - 106 (  84 - 110)
    INTEGRAL     Likelihood = -1.97   Transmembrane   62 -  78 (  62 -  78)
    INTEGRAL     Likelihood = -0.69   Transmembrane  275 - 291 ( 275 - 291)

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

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

```
>GP:AAA88609 GB:M37842 unknown protein [Streptococcus mutans]
 Identities = 243/373 (65%), Positives = 302/373 (80%), Gaps = 1/373 (0%)

Query: 71  IGAVLYLVNSEMDALSRVTWLILVMIAPLLGAMFLMYTKFDWGYRGLKQRLETLIDESQI 130
           IG+VLYLVNS+MD LS +TWL++++  P+LG +FL+YTK DWGYR LK  ++       +
Sbjct: 2   IGSVLYLVNSQMDTLSIITWLLVILPFPILGTLFLIYTKQDWGYRELKSLIKKSTQAIKP 61

Query: 131 YLEDDPETLNQLKSSTSTTYHLVQYFEKAHGNFPVYRNTDVTFLPTGEAFFEKMKEELLK 190
           Y + D    L +LK S + TY+L QY  ++ G FPVY+NT VT+ P G++ FE+MK++LLK
Sbjct: 62  YFQYDQRILYKLKESHARTYNLAQYLHRS-GGFPVYKNTKVTYFPNGQSKFEEMKKQLLK 120

Query: 191 AKKYIFLEFFIIDEGIMWGEILSILEQKVEEGVEVRILYDGMIEITKLSFDYTKRLEKIG 250
           A+K+IFLE+FII EG+MWGEILSILEQKV+EGVEVR++YDGM+E++ LSFDY KRLEKIG
Sbjct: 121 AEKFIFLEYFIIAEGLMWGEILSILEQKVQEGVEVRVMYDGMLELSTLSFDYAKRLEKIG 180

Query: 251 IKAKAFSPISPFISTYYNYRDHRKIVVIDGVVGMTGGVNLADEYINHIELFGHWKDSGIM 310
           IKAK FSPI+PF+STYYNYRDHRKI+VID V   GG+NLADEYIN IE FG+WKD+ +M
Sbjct: 181 IKAKVFSPITPFVSTYYNYRDHRKILVIDNKVAFNGGINLADEYINQIERFGYWKDTAVM 240

Query: 311 LKGKAVDSFLLLFLQMWSITEEKMLVAPYLGVHDDLVENEGYVIPYGDSPLDTDKVGENV 370
           L+G+ V SF L+FLQMWS T +    APYL  +   +   GYVIPY DSPLD +KVGENV
Sbjct: 241 LEGEGVASFTLMFLQMWSTTNKDYEFAPYLTQNFHEIVANGYVIPYSDSPLDHEKVGENV 300

Query: 371 YIDILNHAREYVYIMTPYLILDSELEHAIQFAAERGVDVRIIMPGIPDKPIPYALAKTYY 430
           YIDILN AR+YVYIMTPYLILDSE+EHA+QFAAERGVDV+IIMPGIPDK +P+ALAK Y+
Sbjct: 301 YIDILNQARDYVYIMTPYLILDSEMEHALQFAAERGVDVKIIMPGIPDKKVPFALAKRYF 360

Query: 431 QALTKSGVKIYEY 443
            AL  +GVKIYE+
Sbjct: 361 PALLDAGVKIYEF 373
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6203> which encodes the amino acid sequence <SEQ ID 6204>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood = -8.86   Transmembrane   84 - 100 (  81 - 104)
   INTEGRAL     Likelihood = -8.33   Transmembrane   28 -  44 (  23 -  49)
   INTEGRAL     Likelihood = -6.74   Transmembrane   56 -  72 (  53 -  74)

----- Final Results -----
               bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAA23240 GB:J02911 formyltetrahydrofolate synthetase (FTHFS)
          (ttg start codon) (EC 6.3.4.3) [Moorella thermoacetica]
 Identities = 350/557 (62%), Positives = 438/557 (77%), Gaps = 2/557 (0%)

Query: 2   VLSDIEIANSVTMEPISKVADQLGIDKEALCLYGKYKAKIDARQLVALKNKPDGKLILVT 61
           V SDIEIA +  M+P+ ++A  LGI ++ + LYGKYKAKI        LK+KPDGKLILVT
Sbjct: 4   VPSDIEIAQAAKMKPVMELARGLGIQEDEVELYGKYKAKISLDVYRRLKDKPDGKLILVT 63

Query: 62  AISPTPAGEGKTTTSVGLVDALSAIGKKAVIALREPSLGPVFGVKGGAAGGGHAQVVPME 121
```

```
            AI+PTPAGEGKTTTSVGL DAL+ +GK+ ++ LREPSLGP FG+KGGAAGGG+AQVVPME
Sbjct:  64  AITPTPAGEGKTTTSVGLTDALARLGKRVMVCLREPSLGPSFGIKGGAAGGGYAQVVPME 123

Query: 122 DINLHFTGDFHAIGVANNLLAALIDNHIHHGNSLGIDSRRITWKRVVDMNDRQLRHIVDG 181
            DINLHFTGD HA+  A+NLLAA++DNH+  GN L ID R ITW+RV+D+NDR LR+IV G
Sbjct: 124 DINLHFTGDIHAVTYAHNLLAAMVDNHLQQGNVLNIDPRTITWRRVIDLNDRALRNIVIG 183

Query: 182 LQGKVNGIPREDGYDITVASEIMAILCLSENISDLKARLEKIIIGYNYQGEPVTAKDLKA 241
            L GK NG+PRE G+DI+VASE+MA LCL+ ++ DLK R  +I++GY Y G+PVTA DL+A
Sbjct: 184 LGGKANGVPRETGFDISVASEVMACLCLASDLMDLKERFSRIVVGYTYDGKPVTAGDLEA 243

Query: 242 GGALAALLKDAIHPNLVQTLEHTPALIHGGPFANIAHGCNSVLATKLALKYGDYAVTEAG 301
             G++A L+KDAI PNLVQTLE+TPA IHGGPFANIAHGCNS++ATK ALK  DY VTEAG
Sbjct: 244 QGSMALLMKDAIKPNLVQTLENTPAFIHGGPFANIAHGCNSIIATKTALKLADYVVTEAG 303

Query: 302 FGADLGAEKFIDIKCRMSGLRPAAVVLVATIRALKMHGGVPKADLATENVQAVVDGLPNL 361
            FGADLGAEKF D+KCR +G +P A V+VAT+RALKMHGGVPK+DLATEN++A+ +G  NL
Sbjct: 304 FGADLGAEKFYDVKCRYAGFKPDATVIVATVRALKMHGGVPKSDLATENLEALREGFANL 363

Query: 362 DKHLANIQDVYGLPVVVAINKFPLDTDAELQAVYDACDKRGVDVVISDVWANGGAGGREL 421
            +KH+ NI   +G+P VVAIN FP DT+AEL  +Y+ C K G +V +S+VWA GG GG EL
Sbjct: 364 EKHIENI-GKFGVPAVVAINAFPTDTEAELNLLYELCAKAGAEVALSEVWAKGGEGGLEL 422

Query: 422 AEKVV-TLAEQDNQFRFVYEEDDSIETKLTKIVTKVYGGKGINLSSAAKRELADLERLGF 480
            A KV+ TL  + + F  +Y  D SI+ K+ KI T++YG  G+N ++ A + +  E LG+
Sbjct: 423 ARKVLQTLESRPSNFHVLYNLDLSIKDKIAKIATEIYGADGVNYTAEADKAIQRYESLGY 482

Query: 481 GNYPICMAKTQYSFSDDAKKLGAPTDFTVTISNLKVSAGAGFIVALTGAIMTMPGLPKVP 540
            GN P+ MAKTQYSFSDD  KLG P +FT+T+  +++SAG   IV +TGAIMTMPGLPK P
Sbjct: 483 GNLPVVMAKTQYSFSDDMTKLGRPRNFTITVREVRLSAGGRLIVPITGAIMTMPGLPKRP 542

Query: 541 ASETIDIDEEGNITGLF 557
            A+  IDID +G ITGLF
Sbjct: 543 AACNIDIDADGVITGLF 559
!GB:M37842 unknown protein [Streptococcus mutans] (v...    517  e-145

>GP:AAA88609 GB:M37842 unknown protein [Streptococcus mutans]
 Identities = 246/370 (66%), Positives = 303/370 (81%), Gaps = 1/370 (0%)

Query:  68  VLYLVNSDMDAISRMTWLILIMIAPLLGSLFLIYTKLDWGYRGLKQRINHLVDLSAPYLS 127
            VLYLVNS MD +S +TWL++I+  P+LG+LFLIYTK DWGYR LK  I        PY
Sbjct:   5  VLYLVNSQMDTLSIITWLLVILPFPILGTLFLIYTKQDWGYRELKSLIKKSTQAIKPYFQ  64

Query: 128 DDDAILEVLKDSTSTTYHLVQYLERSRGNFPIYNNTRVTYFPTGETFFDSLKEQLFLAKK 187
             D  IL  LK+S + TY+L QYL RS G FP+Y NT+VTYFP G++ F+ +K+QL  A+K
Sbjct:  65  YDQRILYKLKESHARTYNLAQYLHRS-GGFPVYKNTKVTYFPNGQSKFEEMKKQLLKAEK 123

Query: 188 YIFLEFFIIAEGQMWGEILSILEKKVSEGVEVRVLFDGMNELSTLSSDYAKRLEQIGIKA 247
            +IFLE+FIIAEG MWGEILSILE+KV EGVEVRV++DGM ELSTLS DYAKRLE+IGIKA
Sbjct: 124 FIFLEYFIIAEGLMWGEILSILEQKVQEGVEVRVMYDGMLELSTLSFDYAKRLEKIGIKA 183

Query: 248 KSFLPISPFISTYYNYRDHRKIVVIDGEVSFTGGINLADEYINEVERFGHWKDAGLMLEG 307
            K F PI+PF+STYYNYRDHRKI+VID +V+F GGINLADEYIN++ERFG+WKD  +MLEG
Sbjct: 184 KVFSPITPFVSTYYNYRDHRKILVIDNKVAFNGGINLADEYINQIERFGYWKDTAVMLEG 243

Query: 308 EATDSFLILFLQMWSITEKELIIDPYLSDHSLKLPSDGYVIPYGDSPLDTDKIGKNVYID 367
            E   SF ++FLQMWS T K+    PYL+ +  ++ ++GYVIPY DSPLD +K+G+NVYID
Sbjct: 244 EGVASFTLMFLQMWSTTNKDYEFAPYLTQNFHEIVANGYVIPYSDSPLDHEKVGENVYID 303

Query: 368 ILNHAKEYVYIMTPYLILDSEMEHALRFASERGVDIRIIMPGVPDKGVPYALAKTYYKAL 427
            ILN A++YVYIMTPYLILDSEMEHAL+FA+ERGVD++IIMPG+PDK VP+ALAK Y+ AL
Sbjct: 304 ILNQARDYVYIMTPYLILDSEMEHALQFAAERGVDVKIIMPGIPDKKVPFALAKRYFPAL 363

Query: 428 MSSGVKIYEY 437
            + +GVKIYE+
Sbjct: 364 LDAGVKIYEF 373
```

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

```
 Identities = 362/524 (69%), Positives = 437/524 (83%)

Query: 8   LISNKVKIVRLLNKSKKSLLRGIFSRTTVIAILLILQLLFLLASYSWLEQYRVWLATVEH 67
           +I  K K+  LL+K K   LRGIFSRTT+I +L+ILQL+FL  SY+W+EQYRVW+  +E
Sbjct: 2   IIKKKAKVKYLLHKGKHGFLRGIFSRTTIIVLLIILQLVFLFQSYAWMEQYRVWITILES 61

Query: 68  ILTIGAVLYLVNSEMDALSRVTWLILVMIAPLLGAMFLMYTKFDWGYRGLKQRLETLIDE 127
           +  I  VLYLVNS+MDA+SR+TWLIL+MIAPLLG++FL+YTK DWGYRGLKQR+  L+D
Sbjct: 62  VFAITIVLYLVNSDMDAISRMTWLILIMIAPLLGSLFLIYTKLDWGYRGLKQRINHLVDL 121

Query: 128 SQIYLEDDPETLNQLKSSTSTTYHLVQYFEKAHGNFPVYRNTDVTFLPTGEAFFEKMKEE 187
           S  YL DD   L  LK STSTTYHLVQY E++ GNFP+Y NT VT+ PTGE FF+ +KE+
Sbjct: 122 SAPYLSDDDAILEVLKDSTSTTYHLVQYLERSRGNFPIYNNTRVTYFPTGETFFDSLKEQ 181

Query: 188 LLKAKKYIFLEFFIIDEGIMWGEILSILEQKVEEGVEVRILYDGMIEITKLSFDYTKRLE 247
           L  AKKYIFLEFFII EG MWGEILSILE+KV EGVEVR+L+DGM E++ LS DY KRLE
Sbjct: 182 LFLAKKYIFLEFFIIAEGQMWGEILSILEKKVSEGVEVRVLFDGMNELSTLSSDYAKRLE 241

Query: 248 KIGIKAKAFSPISPFISTYYNYRDHRKIVVIDGVVGMTGGVNLADEYINHIELFGHWKDS 307
           +IGIKAK+F PISPFISTYYNYRDHRKIVVIDG V  TGG+NLADEYIN +E FGHWKD+
Sbjct: 242 QIGIKAKSFLPISPFISTYYNYRDHRKIVVIDGEVSFTGGINLADEYINEVERFGHWKDA 301

Query: 308 GIMLKGKAVDSFLLLFLQMWSITEEKMLVAPYLGVHDDLVENEGYVIPYGDSPLDTDKVG 367
           G+ML+G+A DSFL+LFLQMWSITE+++++ PYL  H   + ++GYVIPYGDSPLDTDK+G
Sbjct: 302 GLMLEGEATDSFLILFLQMWSITEKELIIDPYLSDHSLKLPSDGYVIPYGDSPLDTDKIG 361

Query: 368 ENVYIDILNHAREYVYIMTPYLILDSELEHAIQFAAERGVDVRIIMPGIPDKPIPYALAK 427
           +NVYIDILNHA+EYVYIMTPYLILDSE+EHA++FA+ERGVD+RIIMPG+PDK +PYALAK
Sbjct: 362 KNVYIDILNHAKEYVYIMTPYLILDSEMEHALRFASERGVDIRIIMPGVPDKGVPYALAK 421

Query: 428 TYYQALTKSGVKIYEYTLGFVHSKIFLSDNTKAVVGTINLDYRSLYHHFECAVYLYKVDA 487
           TYY+AL  SGVKIYEY  GFVHSK+F+SDNTKAVVGTINLDYRSLYHHFECA YLY+V
Sbjct: 422 TYYKALMSSGVKIYEYQPGFVHSKVFISDNTKAVVGTINLDYRSLYHHFECATYLYRVSV 481

Query: 488 IQDIYRDYMDTLNKSRLVSLKDINNIPKFQKVIGIVTKTIAPLL 531
           I DI  D+ +   +S L++   +  P +QK+IG++ + IAPLL
Sbjct: 482 IADIVNDFNEAQKQSLLMTSDHLTQRPWYQKLIGLLVRIIAPLL 525
```

A related GBS gene <SEQ ID 8953> and protein <SEQ ID 8954> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 6
McG: Discrim Score:      -8.80
GvH: Signal Score (-7.5): -1.94
     Possible site: 53
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 4 value: -10.03 threshold:  0.0
   INTEGRAL    Likelihood =-10.03   Transmembrane   34 -  50 (  29 -  56)
   INTEGRAL    Likelihood = -7.70   Transmembrane   90 - 106 (  84 - 110)
   INTEGRAL    Likelihood = -1.97   Transmembrane   62 -  78 (  62 -  78)
   PERIPHERAL  Likelihood =  1.22     199
 modified ALOM score:   2.51

*** Reasoning Step: 3

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

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

```
32.5/57.2% over 498aa

                                                                  Bacillus firmus
  SP|O66043| CARDIOLIPIN SYNTHETASE (EC 2.7.8.-) (CARDIOLIPIN SYNTHASE) (CL SYNTHASE).
Insert characterized
  GP|2952028|gb|AAC05444.1||U88888 cardiolipin synthase Insert characterized
```

```
ORF01572(409 - 1893 of 2193)
SP|O66043|CLS_BACFI(5 - 503 of 503) CARDIOLIPIN SYNTHETASE (EC 2.7.8.-) (CARDIOLIPIN
SYNTHASE) (CL SYNTHASE). GP|2952028|gb|AAC05444.1||U88888 cardiolipin synthase (Bacillus
firmus)
%Match = 17.9
%Identity = 32.5  %Similarity = 57.1
Matches = 162  Mismatches = 204  Conservative Sub.s = 123

153       183       213       243       273       303       333       363
NLQLSIWMF*KTVQPLDYFK**RGRACDASLFLLGIRF*LEII*NNRMLFK*QYAIIK*LIWRGEKLISNKVKIVRLLNK

393       423       447       477       507       528       558       588
SKKSLLRGIFSRTTVIAILLILQLLF--LLASYSWLEQYRVWLATVEHILT---IGAVLYLVNSEMDALSRVTWLILVMI
                MKNRLNVLAFFALLFAALYISRGFLQSWMVGTLSVVFTLSVIFIGIIIFFEN--RHPTKTLTWLLVLAA
                          10        20        30        40        50        60

618       648       678       705       735       765       789       819
APLLGAMFLMYTKFDWGYRGLKQRLETLIDESQIYLE-DDPETLNQLKSSTSTTYHLVQYFEKAH--GNFPVYRNTDVTF
FPVVG--FFFYLMFGQNHRKSKRFSKKAIEDERAFQKIEGQRQLNE-EQLKKMGGHQQLLFRLAHKLGKNPISFSSETKV
                80        90        100       110       120       130       140

849       879       909       939       969       999       1029      1059
LPTGEAFFEKMKEELLKAKKYIFLEFFIIDEGIMWGEILSILEQKVEEGVEVRILYDGMIEITKLSFDYTKRLEKIGIKA
LTDGKETYAHILQALKMAEHHIHLEYYIVRHDDLGNQIKDILISKAKEGVHVRFLYDG-VGSWKLSKSYVEELRDAGVEM
                160       170       180       190       200       210       220

1086      1116      1146      1176      1206      1236      1266      1293
KAFSPIS-PFISTYYNYRDHRKIVVIDGVVGMTGGVNLADEYINHIELFGHWKDSGIMLKGKAVDSFLLLFLQMWSI-TE
VSFSPVKLPFLTHTINYRNHRKIIVIDGVVGFVGGLNIGDEYLGKDAYFGYWRDTHLYVRGEAVRTLQLIFLQDWHYQTG
                240       250       260       270       280       290       300

1323      1353      1383      1413      1443      1473      1503      1533
EKMLVAPYLGVHDDLVENEGYVIPYGDSPLDTDKVGENVYIDILNHAREYVYIMTPYLILDSELEHAIQFAAERGVDVRI
ETILNQTYLSPSLSMTKGDGGVQMIASGPDTRWEVNKKLFFSMITSAKKSIWIASPYFIPDDDILSALKIAALSGIDVRI
                320       330       340       350       360       370       380

1563      1593      1623      1653      1683      1713      1743      1773
IMPGIPDKPIPYALAKTYYQALTKSGVKIYEYTLGFVHSKIFLSDNTKAVVGTINLDYRSLYHHFECAVYLYKVDAIQDI
LVPNRPDKRIVFHASRSYFPELLEAGVKVYEYNRGFMHSKIIIVDHEIASIGTSNMDMRSFHLNFEVNAYLYRTSSVTKL
                400       410       420       430       440       450       460

1803      1833      1863      1893      1923      1953      1983      2013
YRDYMDTLNKSRLVSLKDINNIPKFQKVIGIVTKTIAPLL*K*FIFNLILKVN*RI*LYLKSKGCILTKLC*TTVMR*VD
VSDYVYDLEHSNQINFSLFKNRPFFHRLIESTSRLLSPLL
                480       490       500
```

SEQ ID 8954 (GBS277d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 150 (lane 18; MW 51kDa), in Figure 151 (lane 17 & 18; MW 51kDa) and in Figure 182 (lane 12; MW 51kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 15 & 16; MW 76kDa) and in Figure 58 (lane 5; MW 87kDa).

GBS277d-His was purified as shown in Figure 235, lane 8.

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

Example 2006

A DNA sequence (GBSx2116) was identified in *S.agalactiae* <SEQ ID 6205> which encodes the amino acid sequence <SEQ ID 6206>. This protein is predicted to be aspartate-semialdehyde dehydrogenase. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have an uncleavable N-term signal seq

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

A related GBS nucleic acid sequence <SEQ ID 9831> which encodes amino acid sequence <SEQ ID 9832> was also identified.

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

```
>GP:AAA26850 GB:J02667 aspartate beta-semialdehyde dehydrogenase (EC
          1.2.1.11) [Streptococcus mutans]
 Identities = 261/357 (73%), Positives = 304/357 (85%), Gaps = 1/357 (0%)

Query: 1   MGYTVAIVGATGAVGTQMIRQLEQSNLPIEQVKLLSSSRSAGKILHFKDEAIRVEETTKE 60
           MGYTVAIVGATGAVGT+MI+QLEQS LP+++V+LLSSSRSAGK+L +KD+ + VE TTK+
Sbjct: 1   MGYTVAIVGATGAVGTRMIQQLEQSTLPVDKVRLLSSSRSAGKVLQYKDQDVTVELTTKD 60

Query: 61  SFYDVDIALFSAGGSISAKFAPYAVKSGAVVVDNTSYFRQNPDVPLVVPEVNAHAMIGHN 120
           SF  VDIALFSAGGS+SAKFAPYAVK+GAVVVDNTS+FRQNPDVPLVVPEVNA+AM  HN
Sbjct: 61  SFEAVDIALFSAGGSVSAKFAPYAVKAGAVVVDNTSHFRQNPDVPLVVPEVNAYAMDAHN 120

Query: 121 GIIACPNCSTIQMMIALEPIRQKWGIERVIVSTYQAVSGSGARAVEETKEQLRQVLNDNL 180
           GIIACPNCSTIQMM+ALEPIRQKWG+ RVIVSTYQAVSG+G  A+ ET  ++++V+ND +
Sbjct: 121 GIIACPNCSTIQMMVALEPIRQKWGLSRVIVSTYQAVSGAGQSAINETVREIKEVVNDGV 180

Query: 181 SPDQLIATVLPCSSDQKHYPIAFNALPQIDIFTDNDYTYEEMKMTLETKKIMEDATIKVS 240
            P  + A + P   D+KHYPIAFNAL QID+FTDNDYTYEEMKMT ETKKIME+  + VS
Sbjct: 181 DPKAVHADIFPSGGDKKHYPIAFNALAQIDVFTDNDYTYEEMKMTNETKKIMEEPELPVS 240

Query: 241 ATCVRIPVLSGHSESIYIETKELASISEIKKAIANFPGAVLQDLPSQQIYPQAINAVGHR 300
           A CVR+P+L  HSE++YIETK++A I E+K AIA FPGAVL+D     QIYPQA NAVG R
Sbjct: 241 AHCVRVPILFSHSEAVYIETKDVAPIEEVKAAIAAFPGAVLEDDIKHQIYPQAANAVGSR 300

Query: 301 ETFVGRIRKDLDQENGVHMWVVSDNLLKGAAWNSVQIAETLHKNGLVKPAKELKFEL 357
            TFVGRIRKDLD ENG+HMWVVSDNLLKGAAWNS+  A  LH+ GLV+   ELKFEL
Sbjct: 301 -TFVGRIRKDLDIENGIHMWVVSDNLLKGAAWNSIITANRLHERGLVRSTSELKFEL 356
```

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

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

Example 2007

A DNA sequence (GBSx2117) was identified in *S.agalactiae* <SEQ ID 6207> which encodes the amino acid sequence <SEQ ID 6208>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.66   Transmembrane   33 -  49 (  33 -  49)

----- Final Results -----
               bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 500.

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

Example 2008

A DNA sequence (GBSx2119) was identified in *S.agalactiae* <SEQ ID 6209> which encodes the amino acid sequence <SEQ ID 6210>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3853(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 2009

A DNA sequence (GBSx2120) was identified in *S.agalactiae* <SEQ ID 6211> which encodes the amino acid sequence <SEQ ID 6212>. This protein is predicted to be unnamed protein product (clpP). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3883(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 10061> which encodes amino acid sequence <SEQ ID 10062> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6213> which encodes the amino acid sequence <SEQ ID 6214>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2682(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 = 175/196 (89%), Positives = 187/196 (95%)
```

```
Query: 5   MIPVVIEQTSRGERSYDIYSRLLKDRIIMLTGQVEDNMANSIIAQLLFLDAQDNTKDIYL 64
           MIPVVIEQTSRGERSYDIYSRLLKDRIIMLTG VEDNMANS+IAQLLFLDAQDNTKDIYL
Sbjct: 1   MIPVVIEQTSRGERSYDIYSRLLKDRIIMLTGPVEDNMANSVIAQLLFLDAQDNTKDIYL 60

Query: 65  YVNTPGGSVSAGLAIVDTMNFIKSDVQTIVMGMAASMGTIIASSGAKGKRFMLPNAEYMI 124
           YVNTPGGSVSAGLAIVDTMNFIK+DVQTIVMGMAASMGT+IASSG KGKRFMLPNAEYMI
Sbjct: 61  YVNTPGGSVSAGLAIVDTMNFIKADVQTIVMGMAASMGTVIASSGTKGKRFMLPNAEYMI 120

Query: 125 HQPMGGTGGGTQQSDMAIAAEHLLKTRHTLEKILADNSGQSIEKVHDDAERDRWMSAQET 184
           HQPMGGTGGGTQQ+DMAIAAEHLLKTRH LEKILA N+G++I+++H DAERD WMSA+ET
Sbjct: 121 HQPMGGTGGGTQQTDMAIAAEHLLKTRHRLEKILAQNAGKTIKQIHKDAERDYWMSAEET 180

Query: 185 LDYGFIDAIMENNNLQ 200
           L YGFID IMENN L+
Sbjct: 181 LAYGFIDEIMENNELK 196
```

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

Example 2010

A DNA sequence (GBSx2121) was identified in *S.agalactiae* <SEQ ID 6215> which encodes the amino acid sequence <SEQ ID 6216>. This protein is predicted to be uracil phosphoribosyltransferase (upp). Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.43   Transmembrane  127 - 143 ( 127 - 144)
   INTEGRAL    Likelihood = -0.06   Transmembrane   72 -  88 (  72 -  89)
   INTEGRAL    Likelihood = -0.06   Transmembrane  154 - 170 ( 154 - 170)

----- Final Results -----
              bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10063> which encodes amino acid sequence <SEQ ID 10064> was also identified.

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

```
>GP:AAA26890 GB:L07793 uracil phosphoribosyltransferase
          [Streptococcus salivarius]
 Identities = 192/209 (91%), Positives = 202/209 (95%)

Query: 1   MGKFQVISHPLIQHKLSILRRTTTSTKDFRELVDEIAMLMGYEVSRDLPLEDVEIQTPVA 60
           MGKFQVISHPLIQHKLSILRR  TSTKDFRELV+EIAMLMGYEVSRDLPLE+VEIQTP+
Sbjct: 1   MGKFQVISHPLIQHKLSILRREDTSTKDFRELVNEIAMLMGYEVSRDLPLEEVEIQTPIT 60

Query: 61  TTVQKQLAGKKLAIVPILRAGIGMVDGFLSLVPAAKVGHIGMYRDEETFQPVEYLVKLPE 120
            TVQKQL+GKKLAIVPILRAGIGMVDGFLSLVPAAKVGHIGMYRDEET +PVEYLVKLPE
Sbjct: 61  KTVQKQLSGKKLAIVPILRAGIGMVDGFLSLVPAAKVGHIGMYRDEETLEPVEYLVKLPE 120

Query: 121 DIDQRQIFVVDPMLATGGSAILAVDSLKKRGAASIKFVCLVAAPEGVAALQEAHPDVDIY 180
           DIDQRQIFVVDPMLATGGSAILAVDSLKKRGAA+IKFVCLVAAPEGV  LQ+AHPD+DIY
Sbjct: 121 DIDQRQIFVVDPMLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQDAHPDIDIY 180

Query: 181 TAALDEKLNEHGYIVPGLGDAGDRLFGTK 209
           TA+LDEKLNE+GYIVPGLGDAGDRLFGTK
Sbjct: 181 TASLDEKLNENGYIVPGLGDAGDRLFGTK 209
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6217> which encodes the amino acid sequence <SEQ ID 6218>. Analysis of this protein sequence reveals the following:

```
     Possible site: 26
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.59   Transmembrane   72 -  88 (  72 -  89)
   INTEGRAL    Likelihood = -0.22   Transmembrane  127 - 143 ( 127 - 144)

----- 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 is similar to uracil phosphoribosyltransferase from *S.salivarius*:

```
>GP:AAA26890 GB:L07793 uracil phosphoribosyltransferase [Streptococcus salivarius]
 Identities = 191/209 (91%), Positives = 205/209 (97%)

Query: 1   MGKCQVISHPLIQHKLSILRRQTTSTKDFRELVNEIAMLMGYEVSRDLPLEDVDIQTPVS 60
            MGK QVISHPLIQHKLSILRR+ TSTKDFRELVNEIAMLMGYEVSRDLPLE+V+IQTP++
Sbjct: 1   MGKFQVISHPLIQHKLSILRREDTSTKDFRELVNEIAMLMGYEVSRDLPLEEVEIQTPIT 60

Query: 61  KTVQKQLAGKKLAIVPILRAGIGMVDGLLSLVPAAKVGHIGMYRNEETLEPVEYLVKLPE 120
            KTVQKQL+GKKLAIVPILRAGIGMVDG LSLVPAAKVGHIGMYR+EETLEPVEYLVKLPE
Sbjct: 61  KTVQKQLSGKKLAIVPILRAGIGMVDGFLSLVPAAKVGHIGMYRDEETLEPVEYLVKLPE 120

Query: 121 DINQRQIFLVDPMLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQEAHPDIDIF 180
            DI+QRQIF+VDPMLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQ+AHPDIDI+
Sbjct: 121 DIDQRQIFVVDPMLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQDAHPDIDIY 180

Query: 181 TAALDDHLNEHGYIVPGLGDAGDRLFGTK 209
            TA+LD+ LNE+GYIVPGLGDAGDRLFGTK
Sbjct: 181 TASLDEKLNENGYIVPGLGDAGDRLFGTK 209
```

/

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

```
 Identities = 190/209 (90%), Positives = 201/209 (95%)

Query: 1   MGKFQVISHPLIQHKLSILRRTTTSTKDFRELVDEIAMLMGYEVSRDLPLEDVEIQTPVA 60
            MGK QVISHPLIQHKLSILRR TTSTKDFRELV+EIAMLMGYEVSRDLPLEDV+IQTPV+
Sbjct: 1   MGKCQVISHPLIQHKLSILRRQTTSTKDFRELVNEIAMLMGYEVSRDLPLEDVDIQTPVS 60

Query: 61  TTVQKQLAGKKLAIVPILRAGIGMVDGFLSLVPAAKVGHIGMYRDEETFQPVEYLVKLPE 120
             VQKQLAGKKLAIVPILRAGIGMVDG LSLVPAAKVGHIGMYR+EET +PVEYLVKLPE
Sbjct: 61  KTVQKQLAGKKLAIVPILRAGIGMVDGLLSLVPAAKVGHIGMYRNEETLEPVEYLVKLPE 120

Query: 121 DIDQRQIFVVDPMLATGGSAILAVDSLKKRGAASIKFVCLVAAPEGVAALQEAHPDVDIY 180
            DI+QRQIF+VDPMLATGGSAILAVDSLKKRGAA+IKFVCLVAAPEGV  LQEAHPD+DI+
Sbjct: 121 DINQRQIFLVDPMLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQEAHPDIDIF 180

Query: 181 TAALDEKLNEHGYIVPGLGDAGDRLFGTK 209
            TAALD+ LNEHGYIVPGLGDAGDRLFGTK
Sbjct: 181 TAALDDHLNEHGYIVPGLGDAGDRLFGTK 209
```

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

Example 2011

A DNA sequence (GBSx2122) was identified in *S.agalactiae* <SEQ ID 6219> which encodes the amino acid sequence <SEQ ID 6220>. This protein is predicted to be hemolysin (patB). Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.29   Transmembrane   88 - 104 (  86 - 106)

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

```
bacterial membrane --- Certainty=0.2317(Affirmative) < succ>
 bacterial outside --- 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:CAB15133 GB:Z99120 aminotransferase [Bacillus subtilis]
 Identities = 130/381 (34%), Positives = 221/381 (57%), Gaps = 4/381 (1%)

Query: 5   DFTSLPERFSSNTIKWKAVQK---DQEILPLWIADMDFPIFPEMSEAIEDFSHQMVFGYD 61
            +F    ER  + ++KW    +     + LP+W+ADMDF      ++EA+++     +FGY
Sbjct: 2   NFDKREERLGTQSVKWDKTGELFGVTDALPMWVADMDFRAPEAITEALKERLDHGIFGYT 61

Query: 62  SPKDSLYQAISNWEVQEHGYQFDKKSLLLIDGVVPAISVAIQAFTKEGDAVLINTPVYPP 121
            +P       A+  W    HG++ + +S+    GVV A+S+A+QAFT+ GD V++  PVY P
Sbjct: 62  TPDQKTKDAVCGWMQNRHGWKVNPESITFSPGVVTALSMAVQAFTEPGDQVVVQPPVYTP 121

Query: 122 FARTIKYNNRHLVSNSLLNNNQYFEIDFKQLEKDIIENNVKLYIFCSPHNPGGRVWTKGE 181
            F   ++ N RH++ N LL  +  + IDF+ LE  + + +V L+I C+PHNP GR W++ +
Sbjct: 122 FYHMVEKNGRHILHNPLLEKDGAYAIDFEDLETKLSDPSVTLFILCNPHNPSGRSWSRED 181

Query: 182 IQKIGDICKRYNVILVSDEIHQDLVLFDNVHHSFNTVDSSFKELSVILSSATKTFNIAGT 241
            + K+G++C  + V +VSDEIH DL+L+ + H  F ++   F ++SV  ++ +KTFNIAG
Sbjct: 182 LLKLGELCLEHGVTVVSDEIHSDLMLYGHKHTPFASLSDDFADISVTCAAPSKTFNIAGL 241

Query: 242 KNSFAIIENEKLRSDFKKRQIANNQQEISSLGLLATEVAFTKEKQWLKALKMELEGSIEY 301
            + S  II +   R+ F      N    +++  + A E A++K   WL  L   +E ++
Sbjct: 242 QASAIIIPDRLKRAKFSASLQRNGLGGLNAFAVTAIEAAYSKGGPWLDELITYIEKNMNE 301

Query: 302 LYEQL-TQKTHIKVMKPEGTYLVWLDFSAYNLTHLEIQEKLRYDAKLILNDGLTFGKEGK 360
               L T+   +K+MKP+ +YL+WLDFSAY L+  E+Q+++     K+IL  G  +G  G+
Sbjct: 302 AEAFLSTELPKVKMMKPDASYLIWLDFSAYGLSDAELQQRMLKKGKVILEPGTKYGPGGE 361

Query: 361 KHARINVAAPRSVIEEAVLRL 381
               R+N      + +++ + R+
Sbjct: 362 GFMRLNAGCSLATLQDGLRRI 382
```

There is also homology to SEQ ID 1006.

SEQ ID 6220 (GBS392) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 2; MW 46.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 5; MW 71kDa).

GBS392-GST was purified as shown in Figure 217, lane 4.

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

Example 2012

A DNA sequence (GBSx2123) was identified in *S.agalactiae* <SEQ ID 6221> which encodes the amino acid sequence <SEQ ID 6222>. This protein is predicted to be rRNA methylase, SpoU family (cspR). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1436(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:AAB02738 GB:U58864 CspR [Bacillus subtilis]
 Identities = 84/155 (54%), Positives = 120/155 (77%), Gaps = 3/155 (1%)

Query: 19  HIVLFEPQIPANTGNIARTCAATNAPLHIIRPMGFPIDDKKMKRAGLDYWDKLDVSFYDG 78
           H+VL++P+IPANTGNIARTCAATN  LH+IRP+GF  DDK +KRAGLDYW+ ++V ++D
Sbjct: 4   HVVLYQPEIPANTGNIARTCAATNTTLHLIRPLGFSTDDKMLKRAGLDYWEFVNVVYHDS 63

Query: 79  LEE-FMLSCRGKVHLISKFADKVYSDENYND-DQDHYFMFGREDKGLPETFMREHAEKAL 136
           LEE F    +GK   I+KF  + ++  +Y D D+D++F+FGRE  GLP+  ++ + ++ L
Sbjct: 64  LEELFEAYKKGKFFFITKFGQQPHTSFDYTDLDEDYFFVFGRETSGLPKDLIQNNMDRCL 123

Query: 137 RIPMNDEHVRSLNVSNTVCMIVYEALRQQSFPNLE 171
           R+PM  EHVRSLN+SNT  ++VYEALRQQ++ +L+
Sbjct: 124 RLPMT-EHVRSLNLSNTAAILVYEALRQQNYRDLK 157
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6223> which encodes the amino acid sequence <SEQ ID 6224>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2236(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 = 135/182 (74%), Positives = 150/182 (82%)

Query: 1   MNIETLTQKNHRSDSGRNHIVLFEPQIPANTGNIARTCAATNAPLHIIRPMGFPIDDKKM 60
           M  + L  KN +    RNHIVLF+PQIP NTGNIARTCAATNAPLHII+PMGFPIDD+KM
Sbjct: 13  MTTKELINKNDKVKKARNHIVLFQPQIPQNTGNIARTCAATNAPLHIIKPMGFPIDDRKM 72

Query: 61  KRAGLDYWDKLDVSFYDGLEEFMLSCRGKVHLISKFADKVYSDENYNDDQDHYFMFGRED 120
           KRAGLDYWDKL++ FYD LE+F+  C G++HLISKFA   YS   Y D   HYF+FGRED
Sbjct: 73  KRAGLDYWDKLELHFYDHLEQFINQCHGQLHLISKFAVNNYSQATYADGDSHYFLFGRED 132

Query: 121 KGLPETFMREHAEKALRIPMNDEHVRSLNVSNTVCMIVYEALRQQSFPNLELSHTYENDK 180
            GLPE FMREHAEKALRIPMNDEHVRSLNVSNTVCM++YEALRQQ F  LEL HTYE+DK
Sbjct: 133 TGLPEDFMREHAEKALRIPMNDEHVRSLNVSNTVCMVIYEALRQQGFQGLELKHTYEHDK 192

Query: 181 LK 182
           LK
Sbjct: 193 LK 194
```

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

Example 2013

A DNA sequence (GBSx2124) was identified in *S.agalactiae* <SEQ ID 6225> which encodes the amino acid sequence <SEQ ID 6226>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.79   Transmembrane   82 -  98 (  69 - 100)
   INTEGRAL    Likelihood = -6.48   Transmembrane   27 -  43 (  24 -  47)
   INTEGRAL    Likelihood = -5.52   Transmembrane  132 - 148 ( 126 - 151)
   INTEGRAL    Likelihood = -5.10   Transmembrane  162 - 178 ( 161 - 185)

----- Final Results -----
            bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9411> which encodes amino acid sequence <SEQ ID 9412> was also identified.

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

```
>GP:CAB13143 GB:Z99110 similar to amino acid permease [Bacillus subtilis]
 Identities = 46/143 (32%), Positives = 81/143 (56%), Gaps = 1/143 (0%)

Query: 3   FAYDGWTIFVNIAPEVKNPKKNLPLAFVIGPALILLSYLAFFYGLTQILGASFIMTTGND 62
           FAYDGW +   +  E+KNP+K LP A   G  ++   Y+   + L  IL A+ I+T G +
Sbjct: 203 FAYDGWILLAALGGEMKNPEKLLPRAMTGGLLIVTAIYIFINFALLHILSANEIVTLGEN 262

Query: 63  AINYAANIIFGPSVGRLLSFIVILSVLGVANGLLLGTMRLPQAFAERGWIK-SERMANIN 121
           A + AA ++FG   G+L+S  +I+S+ G  NG +L   R+  A AER  +  +E++++++
Sbjct: 263 ATSTAATMLFGSIGGKLISVGIIVSIFGCLNGKVLSFPRVSFAMAERKQLPFAEKLSHVH 322

Query: 122 LKYQMSLPASLTVTAVAIFWLFV 144
             ++    A     A+A+  + +
Sbjct: 323 PSFRTPWIAISFQIALALIMMLI 345
```

There is also homology to SEQ ID 3114.

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

Example 2014

A DNA sequence (GBSx2125) was identified in *S.agalactiae* <SEQ ID 6227> which encodes the amino acid sequence <SEQ ID 6228>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1849(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 9439> which encodes amino acid sequence <SEQ ID 9440> was also identified.

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

```
>GP:AAD23454 GB:AF117741 cochaperonin GroES [Streptococcus pneumoniae]
 Identities = 31/52 (59%), Positives = 42/52 (80%)

Query: 2  GDGIRTLTGELVAPSVAEGDTVLVENGAGLEVKDGNEKVTVVRESDIVAVVK 53
          G G+RTL G+LVAPSV  GD VLVE  AGL+VKDG+EK  +V E++I+A+++
Sbjct: 42 GQGVRTLNGDLVAPSVKTGDRVLVEAHAGLDVKDGDEKYIIVGEANILAIIE 93
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6229> which encodes the amino acid sequence <SEQ ID 6230>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3290(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 = 29/49 (59%), Positives = 39/49 (79%)

Query: 4   GIRTLTGELVAPSVAEGDTVLVENGAGLEVKDGNEKVTVVRESDIVAVV 52
           G+RT+TG+ V PSV+ G  VLVENG  LEV   +EKV+++RESDI+A+V
Sbjct: 60  GVRTITGDSVLPSVSVGQEVLVENGHDLEVTVDDEKVSIIRESDIIAIV 108
```

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

Example 2015

A DNA sequence (GBSx2126) was identified in *S.agalactiae* <SEQ ID 6231> which encodes the amino acid sequence <SEQ ID 6232>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1272(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:AAD23455 GB:AF117741 chaperonin GroEL [Streptococcus pneumoniae]
 Identities = 472/539 (87%), Positives = 513/539 (94%), Gaps = 1/539 (0%)

Query: 1   MAKDIKFSADARSAMVRGVDILADTVKVTLGPKGRNVVLEKAFGSPLITNDGVTIAKEIE 60
           M+K+IKFS+DARSAMVRGVDILADTVKVTLGPK RNVVLEK+FGSPLITNDGVTIAKEIE
Sbjct: 1   MSKEIKFSSDARSAMVRGVDILADTVKVTLGPKDRNVVLEKSFGSPLITNDGVTIAKEIE 60

Query: 61  LEDHFENMGAKLVSEVASKTNDIAGDGTTTATVLTQAIVREGLKNVTAGANPIGIRRGIE 120
           LEDHFENMGAKLVSE+ASKTNDIAGDGTTTATVLTQAIVREG+KNVTAGANPIGIRRGIE
Sbjct: 61  LEDHFENMGAKLVSEIASKTNDIAGDGTTTATVLTQAIVREGIKNVTAGANPIGIRRGIE 120

Query: 121 TAVSAAVEELKEIAQPVSGKEAIAQVAAVSSRSEKVGEYISEAMERVGNDGVITIEESRG 180
           TAV+AAVE LK  A PV+ KEAI+QVAAVSSRSEKVGEYISEAME+VG DGVITIEESRG
Sbjct: 121 TAVAAAVEALKNNAIPVANKEAISQVAAVSSRSEKVGEYISEAMEKVGKDGVITIEESRG 180

Query: 181 METELEVVEGMQFDRGYLSQYMVTDNEKMVSELENPYILITDKKISNIQEILPLLEEVLK 240
           METELEVVEGMQFDRGYLSQYMVTD+EKMV++LENPYILITDKKISNIQEILPLLE +L+
Sbjct: 181 METELEVVEGMQFDRGYLSQYMVTDSEKMVADLENPYILITDKKISNIQEILPLLESILQ 240

Query: 241 TNRPLLIIADDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIAILTGGTVVT 300
           +NRPLLIIADDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIAILTGGTV+T
Sbjct: 241 SNRPLLIIADDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIAILTGGTVIT 300

Query: 301 EDLGLDLKDATMQVLGQSAKVTVDKDSTVIVEGAGDSSAIANRVAIIKSQMEATTSDFDR 360
           EDLGL+LKDAT++ LGQ+A+VTVDKDSTVIVEGAG+  AI++RVA+IKSQ+E TTS+FDR
Sbjct: 301 EDLGLELKDATIEALGQAARVTVDKDSTVIVEGAGNPEAISHRVAVIKSQIETTTSEFDR 360

Query: 361 EKLQERLAKLAGGVAVIKVGAATETELKEMKLRIEDALNATRAAVEEGIVSGGGTALVNV 420
           EKLQERLAKL+GGVAVIKVGAATETELKEMKLRIEDALNATRAAVEEGIV+GGGTAL NV
Sbjct: 361 EKLQERLAKLSGGVAVIKVGAATETELKEMKLRIEDALNATRAAVEEGIVAGGGTALANV 420

Query: 421 IEKVAALKLNGDEETGRNIVLRALEEPVRQIAYNAGYEGSVIIERLKQSEIGTGFNAANG 480
           I   A L+L GDE TGRNIVLRALEEPVRQIA+NAG+EGS++I+RLK +E+G GFNAA G
Sbjct: 421 IPAEATLELTGDEATGRNIVLRALEEPVRQIAHNAGFEGSIVIDRLKNAELGIGFNAATG 480

Query: 481 EWVDMVTTGIIDPVKVTRSALQNAASVASLILTTEAVVANKPEPEAPTAPAMDPSMMGG 539
           EWV+M+  GIIDPVKV+RSALQNAASVASLILTTEAVVANKPEP AP APAMDPSMMGG
Sbjct: 481 EWVNMIDQGIIDPVKVSRSALQNAASVASLILTTEAVVANKPEPVAP-APAMDPSMMGG 538
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6233> which encodes the amino acid sequence <SEQ ID 6234>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1070(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 = 491/543 (90%), Positives = 515/543 (94%), Gaps = 3/543 (0%)

Query: 1   MAKDIKFSADARSAMVRGVDILADTVKVTLGPKGRNVVLEKAFGSPLITNDGVTIAKEIE 60
           MAKDIKFSADAR+AMVRGVD+LADTVKVTLGPKGRNVVLEKAFGSPLITNDGVTIAKEIE
Sbjct: 3   MAKDIKFSADARAAMVRGVDMLADTVKVTLGPKGRNVVLEKAFGSPLITNDGVTIAKEIE 62

Query: 61  LEDHFENMGAKLVSEVASKTNDIAGDGTTTATVLTQAIVREGLKNVTAGANPIGIRRGIE 120
           LEDHFENMGAKLVSEVASKTNDIAGDGTTTATVLTQAIV EGLKNVTAGANPIGIRRGIE
Sbjct: 63  LEDHFENMGAKLVSEVASKTNDIAGDGTTTATVLTQAIVHEGLKNVTAGANPIGIRRGIE 122

Query: 121 TAVSAAVEELKEIAQPVSGKEAIAQVAAVSSRSEKVGEYISEAMERVGNDGVITIEESRG 180
           TA + AVE LK IAQPVSGKEAIAQVAAVSSRSEKVGEYISEAMERVGNDGVITIEESRG
Sbjct: 123 TATATAVEALKAIAQPVSGKEAIAQVAAVSSRSEKVGEYISEAMERVGNDGVITIEESRG 182

Query: 181 METELEVVEGMQFDRGYLSQYMVTDNEKMVSELENPYILITDKKISNIQEILPLLEEVLK 240
           METELEVVEGMQFDRGYLSQYMVTDNEKMV++LENP+ILITDKK+SNIQ+ILPLLEEVLK
Sbjct: 183 METELEVVEGMQFDRGYLSQYMVTDNEKMVADLENPFILITDKKVSNIQDILPLLEEVLK 242

Query: 241 TNRPLLIIADDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIAILTGGTVVT 300
           TNRPLLIIADDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIAILTGGTV+T
Sbjct: 243 TNRPLLIIADDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIAILTGGTVIT 302

Query: 301 EDLGLDLKDATMQVLGQSAKVTVDKDSTVIVEGAGDSSAIANRVAIIKSQMEATTSDFDR 360
           EDLGL+LKDATM  LGQ+AK+TVDKDSTVIVEG+G S AIANR+A+IKSQ+E TTSDFDR
Sbjct: 303 EDLGLELKDATMTALGQAAKITVDKDSTVIVEGSGSSEAIANRIALIKSQLETTTSDFDR 362

Query: 361 EKLQERLAKLAGGVAVIKVGAATETELKEMKLRIEDALNATRAAVEEGIVSGGGTALVNV 420
           EKLQERLAKLAGGVAVIKVGA TET LKEMKLRIEDALNATRAAVEEGIV+GGGTAL+ V
Sbjct: 363 EKLQERLAKLAGGVAVIKVGAPTETALKEMKLRIEDALNATRAAVEEGIVAGGGTALITV 422

Query: 421 IEKVAALKLNGDEETGRNIVLRALEEPVRQIAYNAGYEGSVIIERLKQSEIGTGFNAANG 480
           IEKVAAL+L GD+ TGRNIVLRALEEPVRQIA NAGYEGSV+I++LK S  GTGFNAA G
Sbjct: 423 IEKVAALELEGDDATGRNIVLRALEEPVRQIALNAGYEGSVVIDKLKNSPAGTGFNAATG 482

Query: 481 EWVDMVTTGIIDPVKVTRSALQNAASVASLILTTEAVVANKPEP--EAPTAPA-MDPSMM 537
           EWVDM+ TGIIDPVKVTRSALQNAASVASLILTTEAVVANKPEP   AP  PA MDP MM
Sbjct: 483 EWVDMIKTGIIDPVKVTRSALQNAASVASLILTTEAVVANKPEPATPAPAMPAGMDPGMM 542

Query: 538 GGF 540
           GGF
Sbjct: 543 GGF 545
```

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

Example 2016

A DNA sequence (GBSx2127) was identified in *S.agalactiae* <SEQ ID 6235> which encodes the amino acid sequence <SEQ ID 6236>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3216(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 10247> which encodes amino acid sequence <SEQ ID 10248> was also identified.

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

```
>GP:BAB06113 GB:AP001515 transcriptional regulator (GntR family)
          [Bacillus halodurans]
 Identities = 50/171 (29%), Positives = 86/171 (50%), Gaps = 17/171 (9%)

Query: 21  HVQVYNKIFNMIQDGTYSPGMQLPSEPELAGQLNVSRATLRKSLALLQEDHLVKNIRGKG 80
           ++QV +K+ + ++ G Y  G +LPSE EL+ QL VSRATLR++L LL+E+ +V    G G
Sbjct: 10  YLQVIDKLKHDMEAGVYEEGEKLPSEFELSKQLGVSRATLREALRLLEEEGVVVRRHGVG 69

Query: 81  NFIRENSSNLSENGYENRQHPIKTCLTSKITEVELE--------FRVEVPAEAITASLKQ 132
            F+  ++  L   G E          +T  I   ++E         +++E   +
Sbjct: 70  TFV--HTKPLFSAGIEELY-----SVTDMIRHADMEPGTIFLSSYQIEATDDDKRRFQTD 122

Query: 133 ETPVVVIADRWYHTDDGPLAYTLSFIPIELISDAEISLHDTKQLLNFIEEG 183
               +++ +R    D  P+ Y L  +P ELI   + S+H+   +L+ +E G
Sbjct: 123 NLDQLMMIERVRTADGVPIVYCLDKLPAELI--GQHSVHEINSILDHLESG 171
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6237> which encodes the amino acid sequence <SEQ ID 6238>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2297(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 154/244 (63%), Positives = 189/244 (77%)

Query: 7   MPKNELNNKLNKLKHVQVYNKIFNMIQDGTYSPGMQLPSEPELAGQLNVSRATLRKSLAL 66
           M  N+L  KL KLKHVQVYN IF +IQDGTYSPGMQLPSEPELA QLNVSR TLRKSLAL
Sbjct: 1   MSTNDLTKKLKKLKHVQVYNTIFQLIQDGTYSPGMQLPSEPELARQLNVSRMTLRKSLAL 60

Query: 67  LQEDHLVKNIRGKGNFIRENSSNLSENGYENRQHPIKTCLTSKITEVELEFRVEVPAEAI 126
           LQEDHL+KNIRGKGNFI +        G+E  QHPI    L+S IT+VELE+R+EVP  AI
Sbjct: 61  LQEDHLIKNIRGKGNFILKTPETKYHQGFEYLQHPIYASLSSDITKVELEYRIEVPTVAI 120

Query: 127 TASLKQETPVVVIADRWYHTDDGPLAYTLSFIPIELISDAEISLHDTKQLLNFIEEGIYQ 186
           TASLKQETPVV+I DRWYH+ +  +AY+LSFIPIE+IS   I+L+  + LL F+EE IY+
Sbjct: 121 TASLKQETPVVIIVDRWYHSQNKAIAYSLSFIPIEVISKYAINLNQEEPLLTFLEEKIYE 180

Query: 187 EGISSHSQSHLGYATSGNFSATKYTLSDHGQFILIQETIFKQEKILMCNKHYVPIEHFEL 246
           G +SHS + +GY  +GN++ATKYTLS++  FILIQET++  + IL+  KHYVP + F+L
Sbjct: 181 SGKASHSCNQIGYTKTGNYTATKYTLSENSAFILIQETLYNGKDILVSTKHYVPADLFDL 240

Query: 247 SITS 250
           + S
Sbjct: 241 KVQS 244
```

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

Example 2017

A DNA sequence (GBSx2128) was identified in *S.agalactiae* <SEQ ID 6239> which encodes the amino acid sequence <SEQ ID 6240>. This protein is predicted to be purine nucleoside phosphorylase (udp-1). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3910(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:AAC65977 GB:AE001270 uridine phosphorylase (udp) [Treponema
          pallidum]
 Identities = 145/246 (58%), Positives = 171/246 (68%)

Query: 11  QYHLQIRPGDVGRYVIMPGDPKRCAKIAEHFDNAVLVADSREYVTYTGTLNGEKVSVTST 70
           +YH+ ++  D+G YVI+PGDP R  KIA+HF +   V  +REYVTYTGTL    VSV ST
Sbjct: 10  EYHIGLKASDIGHYVILPGDPARSEKIAQHFSHPHKVGHNREYVTYTGTLCETPVSVMST 69

Query: 71  GIGGPSASIAMEELKLCGADTFIRVGTCGGIDLDVKGGDIVIATGAIRMEGTSKEYAPIE 130
           GIGGPS +I +EEL   GA TFIRVGT GG+  D+  G +VIATGAIR EGTSKEYAP+E
Sbjct: 70  GIGGPSTAIGVEELIHLGAHTFIRVGTSGGMQPDILAGTVVIATGAIRFEGTSKEYAPVE 129

Query: 131 FPAVADLEVTNALVNAAKKLGYTSHAGVVQCKDAFYGQHEPERMPVSYELLNKWEAWKRL 190
           FPAV D  VT AL +AA+ +      GVVQCKD FYGQH P  MPV  EL  KW AW
Sbjct: 130 FPAVPDFTVTAALKHAAEDVQVRHALGVVQCKDNFYGQHSPHTMPVHAELTQKWHAWIAC 189

Query: 191 GTKASEMESAALFVAASHLGVRCGSDFLVVGNQERNALGMDNPMAHDTEAAIQVAVEALR 250
           T ASEMESAALFV  S   VR G+  LV+GNQ R A G+++   HDTE AI+VAVEA++
Sbjct: 190 NTLASEMESAALFVLGSVRRVRTGAVLLVIGNQTRRAQGLEDIQVHDTENAIRVAVEAVK 249

Query: 251 TLIEND 256
           LI  D
Sbjct: 250 LLITQD 255
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6241> which encodes the amino acid sequence <SEQ ID 6242>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3910(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 259/259 (100%), Positives = 259/259 (100%)

Query: 1   MQNYSGEVGLQYHLQIRPGDVGRYVIMPGDPKRCAKIAEHFDNAVLVADSREYVTYTGTL 60
           MQNYSGEVGLQYHLQIRPGDVGRYVIMPGDPKRCAKIAEHFDNAVLVADSREYVTYTGTL
Sbjct: 1   MQNYSGEVGLQYHLQIRPGDVGRYVIMPGDPKRCAKIAEHFDNAVLVADSREYVTYTGTL 60

Query: 61  NGEKVSVTSTGIGGPSASIAMEELKLCGADTFIRVGTCGGIDLDVKGGDIVIATGAIRME 120
           NGEKVSVTSTGIGGPSASIAMEELKLCGADTFIRVGTCGGIDLDVKGGDIVIATGAIRME
Sbjct: 61  NGEKVSVTSTGIGGPSASIAMEELKLCGADTFIRVGTCGGIDLDVKGGDIVIATGAIRME 120

Query: 121 GTSKEYAPIEFPAVADLEVTNALVNAAKKLGYTSHAGVVQCKDAFYGQHEPERMPVSYEL 180
           GTSKEYAPIEFPAVADLEVTNALVNAAKKLGYTSHAGVVQCKDAFYGQHEPERMPVSYEL
Sbjct: 121 GTSKEYAPIEFPAVADLEVTNALVNAAKKLGYTSHAGVVQCKDAFYGQHEPERMPVSYEL 180
```

```
Query: 181 LNKWEAWKRLGTKASEMESAALFVAASHLGVRCGSDFLVVGNQERNALGMDNPMAHDTEA 240
           LNKWEAWKRLGTKASEMESAALFVAASHLGVRCGSDFLVVGNQERNALGMDNPMAHDTEA
Sbjct: 181 LNKWEAWKRLGTKASEMESAALFVAASHLGVRCGSDFLVVGNQERNALGMDNPMAHDTEA 240

Query: 241 AIQVAVEALRTLIENDKSQ 259
           AIQVAVEALRTLIENDKSQ
Sbjct: 241 AIQVAVEALRTLIENDKSQ 259
```

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

Example 2018

A DNA sequence (GBSx2129) was identified in *S.agalactiae* <SEQ ID 6243> which encodes the amino acid sequence <SEQ ID 6244>. This protein is predicted to be nucleoside transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.45   Transmembrane   35 -  51 (  30 -  57)
   INTEGRAL    Likelihood = -9.29   Transmembrane    8 -  24 (   1 -  28)
   INTEGRAL    Likelihood = -8.07   Transmembrane  388 - 404 ( 379 - 404)
   INTEGRAL    Likelihood = -7.27   Transmembrane  104 - 120 ( 100 - 127)
   INTEGRAL    Likelihood = -6.58   Transmembrane  259 - 275 ( 255 - 284)
   INTEGRAL    Likelihood = -4.35   Transmembrane  172 - 188 ( 171 - 190)
   INTEGRAL    Likelihood = -3.50   Transmembrane  200 - 216 ( 199 - 221)
   INTEGRAL    Likelihood = -2.18   Transmembrane  352 - 368 ( 352 - 371)

----- 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>
```

A related GBS nucleic acid sequence <SEQ ID 10245> which encodes amino acid sequence <SEQ ID 10246> was also identified.

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

```
>GP:BAB05165 GB:AP001512 nucleoside transporter [Bacillus halodurans]
 Identities = 160/405 (39%), Positives = 256/405 (62%), Gaps = 8/405 (1%)

Query: 5   MQFIYSIIGILLVLGIVYAISFNRKSVSLSLIGKALIVQFIIALILVRIPLGQQVVSVVS 64
           M  ++ ++GI++V  I +A S NR+++    I   L +Q + A+I+++IP GQ ++  ++
Sbjct: 1   MNILWGLLGIVVVFLIAFAFSTNRRAIKPRTILGGLAIQLLFAIIVLKIPAGQALLESLT 60

Query: 65  TGVTKVINCGQAGLNFVFGSLADSGAKTGFIFAIQTLGNIVFLSALVSLLYYVGILGFVV 124
             V  +I+    G++FVFG   + G+  GF+FAI  L  ++F SAL+S+LYY+GI+ FV+
Sbjct: 61  NVVLNIISYANEGIDFVFGGFFEEGSGVGFVFAINVLSVVIFFSALISILYYLGIMQFVI 120

Query: 125 KWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGMGSMS 184
           K IG  +  ++ +S+ ES  A AN+F+GQT++P++V  YL +MT SE+  V+  G+ S++
Sbjct: 121 KIIGGALSWLLGTSKAESMSAAANIFVGQTEAPLVVKPYLPKMTQSELFAVMTGGLASVA 180

Query: 185 VSILGGYIALGIPMEYLLIASTMVPIGSILIAKILLPQTEPVQKI-DDIKMDNKGNNANV 243
            S+L GY  LG+P++YLL AS M      +++AK+++P+TE      DD K+     + N+
Sbjct: 181 GSVLIGYSLLGVPLQYLLAASFMAAPAGLIMAKMIMPETEKTTDAEDDFKLAKDEESTNL 240

Query: 244 IDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLG-------IRLEQIFSYVFAP 296
           IDA A GASTG  +  +I A L+AFV L++LIN +L  +G        + LE I  YVFAP
Sbjct: 241 IDAAANGASTGLMLVLNIAAMLLAFVALIALINGILGWIGGLFGASQLSLELILGYVFAP 300

Query: 297 FGFLMGFDHKNILLEGNLLGSKLILNEFVSFQQLGDLIKSLDYRTALVATISLCGFANLS 356
            F++G      L  G+ +G KL++NEFV++      I++L  +  +V + +LCGFAN S
```

```
Sbjct: 301 LAFVIGIPWAEALQAGSYIGQKLVVNEFVAYLSFAPEIENLSDKAVMVISFALCGFANFS 360

Query: 357 SLGICVSGIAVLCPEKRGTLARLVFRAMIGGIAVSMLSAFIVGIV 401
           SLGI + G+  L P +R  +ARL  RA++ G   S+LSA I G++
Sbjct: 361 SLGILLGGLGKLAPSRRPDIARLGLRAILAGTLASLLSASIAGML 405
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6245> which encodes the amino acid sequence <SEQ ID 6246>. Analysis of this protein sequence reveals the following:

```
     Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.45   Transmembrane   35 -  51 (  30 -  57)
   INTEGRAL    Likelihood = -9.29   Transmembrane    8 -  24 (   1 -  28)
   INTEGRAL    Likelihood = -8.07   Transmembrane  388 - 404 ( 379 - 404)
   INTEGRAL    Likelihood = -7.27   Transmembrane  104 - 120 ( 100 - 127)
   INTEGRAL    Likelihood = -6.58   Transmembrane  259 - 275 ( 255 - 284)
   INTEGRAL    Likelihood = -4.35   Transmembrane  172 - 188 ( 171 - 190)
   INTEGRAL    Likelihood = -3.50   Transmembrane  200 - 216 ( 199 - 221)
   INTEGRAL    Likelihood = -2.18   Transmembrane  352 - 368 ( 352 - 371)

----- 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:

```
>GP:BAB05165 GB:AP001512 nucleoside transporter [Bacillus halodurans]
 Identities = 160/405 (39%), Positives = 257/405 (62%), Gaps = 8/405 (1%)

Query: 5   MQFIYSIIGILLVLGIVYAISFNRKSVSLSLIGKALIVQFIIALILVRIPLGQQIVSVVS 64
           M  ++ ++GI++V  I +A S NR+++    I   L +Q + A+I+++IP GQ ++  ++
Sbjct: 1   MNILWGLLGIVVVFLIAFAFSTNRRAIKPRTILGGLAIQLLFAIIVLKIPAGQALLESLT 60

Query: 65  TGVTSVINCGQAGLNFVFGSLADSGAKTGFIFAIQTLGNIVFLSALVSLLYYVGILGFVV 124
             V ++I+    G++FVFG   + G+  GF+FAI  L  ++F SAL+S+LYY+GI+ FV+
Sbjct: 61  NVVLNIISYANEGIDFVFGGFFEEGSGVGFVFAINVLSVVIFFSALISILYYLGIMQFVI 120

Query: 125 KWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGMGSMS 184
           K IG  +  ++ +S+ ES  A AN+F+GQT++P++V  YL +MT SE+  V+  G+ S++
Sbjct: 121 KIIGGALSWLLGTSKAESMSAAANIFVGQTEAPLVVKPYLPKMTQSELFAVMTGGLASVA 180

Query: 185 VSILGGYIALGIPMEYLLIASTMVPIGSILIAKILLPQTEPVQKI-DDIKMDNKGNNANV 243
            S+L GY  LG+P++YLL AS M      +++AK+++P+TE       DD K+     + N+
Sbjct: 181 GSVLIGYSLLGVPLQYLLAASFMAAPAGLIMAKMIMPETEKTTDAEDDFKLAKDEESTNL 240

Query: 244 IDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLG-------IRLEQIFSYVFAP 296
           IDA A GASTG  +  +I A L+AFV L++LIN +L  +G        + LE I  YVFAP
Sbjct: 241 IDAAANGASTGLMLVLNIAAMLLAFVALIALINGILGWIGGLFGASQLSLELILGYVFAP 300

Query: 297 FGFLMGFDHKNILLEGNLLGSKLILNEFVSFQQLGHLIKSLDYRTALVATISLCGFANLS 356
             F++G      L  G+ +G KL++NEFV++      I++L  +  +V + +LCGFAN S
Sbjct: 301 LAFVIGIPWAEALQAGSYIGQKLVVNEFVAYLSFAPEIENLSDKAVMVISFALCGFANFS 360

Query: 357 SLGICVSGIAVLCPEKRSTLARLVFRAMIGGIAVSMLSAFIVGIV 401
           SLGI + G+  L P +R  +ARL  RA++ G   S+LSA I G++
Sbjct: 361 SLGILLGGLGKLAPSRRPDIARLGLRAILAGTLASLLSASIAGML 405
```

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

```
 Identities = 399/404 (98%), Positives = 401/404 (98%)

Query: 1   MEVIMQFIYSIIGILLVLGIVYAISFNRKSVSLSLIGKALIVQFIIALILVRIPLGQQVV 60
           +EVIMQFIYSIIGILLVLGIVYAISFNRKSVSLSLIGKALIVQFIIALILVRIPLGQQ+V
Sbjct: 1   LEVIMQFIYSIIGILLVLGIVYAISFNRKSVSLSLIGKALIVQFIIALILVRIPLGQQIV 60

Query: 61  SVVSTGVTKVINCGQAGLNFVFGSLADSGAKTGFIFAIQTLGNIVFLSALVSLLYYVGIL 120
```

```
            SVVSTGVT VINCGQAGLNFVFGSLADSGAKTGFIFAIQTLGNIVFLSALVSLLYYVGIL
Sbjct: 61   SVVSTGVTSVINCGQAGLNFVFGSLADSGAKTGFIFAIQTLGNIVFLSALVSLLYYVGIL 120

Query: 121 GFVVKWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGM 180
            GFVVKWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGM
Sbjct: 121 GFVVKWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGM 180

Query: 181 GSMSVSILGGYIALGIPMEYLLIASTMVPIGSILIAKILLPQTEPVQKIDDIKMDNKGNN 240
            GSMSVSILGGYIALGIPMEYLLIASTMVPIGSILIAKILLPQTEPVQKIDDIKMDNKGNN
Sbjct: 181 GSMSVSILGGYIALGIPMEYLLIASTMVPIGSILIAKILLPQTEPVQKIDDIKMDNKGNN 240

Query: 241 ANVIDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLGIRLEQIFSYVFAPFGFL 300
            ANVIDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLGIRLEQIFSYVFAPFGFL
Sbjct: 241 ANVIDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLGIRLEQIFSYVFAPFGFL 300

Query: 301 MGFDHKNILLEGNLLGSKLILNEFVSFQQLGDLIKSLDYRTALVATISLCGFANLSSLGI 360
            MGFDHKNILLEGNLLGSKLILNEFVSFQQLG LIKSLDYRTALVATISLCGFANLSSLGI
Sbjct: 301 MGFDHKNILLEGNLLGSKLILNEFVSFQQLGHLIKSLDYRTALVATISLCGFANLSSLGI 360

Query: 361 CVSGIAVLCPEKRGTLARLVFRAMIGGIAVSMLSAFIVGIVTLF 404
            CVSGIAVLCPEKR TLARLVFRAMIGGIAVSMLSAFIVGIVTLF
Sbjct: 361 CVSGIAVLCPEKRSTLARLVFRAMIGGIAVSMLSAFIVGIVTLF 404
```

A related GBS gene <SEQ ID 8955> and protein <SEQ ID 8956> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 1
McG: Discrim Score:     13.83
GvH: Signal Score (-7.5): -2.63
     Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 8 value:  -9.45 threshold:  0.0
   INTEGRAL    Likelihood = -9.45   Transmembrane   35 -  51 (  30 -  57)
   INTEGRAL    Likelihood = -9.29   Transmembrane    8 -  24 (   1 -  28)
   INTEGRAL    Likelihood = -8.07   Transmembrane  388 - 404 ( 379 - 404)
   INTEGRAL    Likelihood = -7.27   Transmembrane  104 - 120 ( 100 - 127)
   INTEGRAL    Likelihood = -6.58   Transmembrane  259 - 275 ( 255 - 284)
   INTEGRAL    Likelihood = -4.35   Transmembrane  172 - 188 ( 171 - 190)
   INTEGRAL    Likelihood = -3.50   Transmembrane  200 - 216 ( 199 - 221)
   INTEGRAL    Likelihood = -2.18   Transmembrane  352 - 368 ( 352 - 371)
   PERIPHERAL  Likelihood =  3.82     286
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:

```
ORF01622(313 - 1512 of 1812)
GP|9656920|gb|AAF95495.1||AE004305(1 - 418 of 418) NupC family protein {Vibrio cholerae}
%Match = 24.0
%Identity = 39.5  %Similarity = 65.7
Matches = 160  Mismatches = 134  Conservative Sub.s = 106

276       306       336       366       396       426       456       486
C*STPHTY*K**ITISEVLEVIMQFIYSIIGILLVLGIVYAISFNRKSVSLSLIGKALIVQFIIALILVRIPLGQQVVSV
                     | :  |:||: ::|||   :| |||:::|  :| |: :|| :   :: :| ||:::
                    MSLFMSLIGMAVLLGIAVLLSSNRKAINLRTVGGAFAIQFSLGAFILYVPWGQELLRG
                             10        20        30        40        50

516       546                 591       621       651       681       711
VSTGVTKVINCGQAGLNFVFGSLADSG-----AKTGFIFAIQTLGNIVFLSALVSLLYYVGILGFVVKWIGKGVGKIMKS
 |  |: ||| |  | :|:|| |           |||||  : |  ::|:|||:|:|||:|:: :|:: :|  |: | : :
```

```
FSDAVSNVINYGNDGTSFLFGGLVSGKMFEVFGGGGFIFAFRVLPTLIFFSALISVLYYLGVMQWVIRILGGGLQKALGT
          70        80        90        100       110       120       130

741       771       801       831       861       891       921       951
SEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGMGSMSVSILGGYIALGIPMEYLLIASTMVPIGSILIAK
|  ||  | ||:|:|||::|::|  :: :|| ||:  |:  |: |::  :| || ::|: :|||: || |   | :| ||
SRAESMSAAANIFVGQTEAPLVVRPFVPKMTQSELFAVMCGGLASIAGGVLAGYASMGVKIEYLVAASFMAAPGGLLFAK
          150       160       170       180       190       200       210

981       1011      1038      1068      1098      1128                1167
ILLPQTEPVQKIDDIKMDNKGNN-ANVIDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLG-------IRLEQI
:::|:||  |  :|| :|   :  ||||||  | ||| | |:|:::|| ||||:||::||| || |:|       ::|| :
LMMPETEKPQDNEDITLDGGDDKPANVIDAAAGGASAGLQLALNVGAMLIAFIGLIALINGMLGGIGGWFGMPELKLEML
          230       240       250       260       270       280       290

1197      1227      1257      1287      1305      1332      1362      1392
FSYVFAPFGFLMGFDHKNILLEGNLLGSKLILNEFVSFQQ----LGDLIKS-LDYRTALVATISLCGFANLSSLGICVSG
: ::|||: ||:|       : | ::| | : ||||:: |    | :     | :|  : : :||||||||: | : |
LGWLFAPLAFLIGVPWNEATVAGEFIGLKTVANEFVAYSQFAPYLTEAAPVVLSEKTKAIISFALCGFANLSSIAILLGG
          310       320       330       340       350       360       370

1422      1452      1482      1512      1542      1572      1602      1632
IAVLCPEKRGTLARLVFRAMIGGIAVSMLSAFIVGIVTLF*KLTKERRIVTWK*KIF*KR*TILC*QQQQHGQKSKQF*M
:  | |::|| :||:  :|:| |   ::::| | |    |
LGSLAPKRRGDIARMGVKAVIAGTLSNLMAATIAGFFLSF
          390       400       410
```

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

Example 2019

A DNA sequence (GBSx2130) was identified in *S.agalactiae* <SEQ ID 6247> which encodes the amino acid sequence <SEQ ID 6248>. This protein is predicted to be deoxyribose-phosphate aldolase (deoC). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2196(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:CAA81646 GB:Z27121 deoxyribose aldolase [Mycoplasma hominis]
 Identities = 99/199 (49%), Positives = 140/199 (69%), Gaps = 1/199 (0%)

Query: 5   DILKTVDHTLLATTATWPEIQTILDDAMAYETASACIPASYVKKAAEYVSGK-LAICTVI 63
            ++ K +DHT L+ +AT  +I  ++ ++A+ Y+  S CI  SYVK A E +    + +CTVI
Sbjct: 3   ELNKYIDHTNLSPSATSKDIDKLIQEAIKYDFKSVCIAPSYVKYAKEALKNSDVLVCTVI 62

Query: 64  GFPNGYSTTAAKVFECQDAIKNGADEIDMVINLTDVKNGDFDTVEEEIRQIKAACQDHIL 123
            GFP GY+ T+ KV+E + A+++GADEIDMVIN+    K+G ++ V  EI+ IK AC    L
Sbjct: 63  GFPLGYNATSVKVYETKIAVEHGADEIDMVINVGRFKDGQYEYVLNEIKAIKEACNGKTL 122

Query: 124 KVIVETCQLTKEELIELCGVVTRSGADFIKTSTGFSTAGATFEDVEVMAKYVGEGVKIKA 183
            KVIVET  LTK ELI++  +V +SGADFIKTSTGFS  GA+FED++ M +  G+ + IKA
Sbjct: 123 KVIVETALLTKAELIKITELVMQSGADFIKTSTGFSYRGASFEDIQTMKETCGDKLLIKA 182

Query: 184 AGGISSLEDAEKFIALGAS 202
            +GGI +L DA++ I LGA+
Sbjct: 183 SGGIKNLADAQEMIRLGAN 201
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6249> which encodes the amino acid sequence <SEQ ID 6250>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2196(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 = 211/223 (94%), Positives = 217/223 (96%)

Query: 1   MEVKDILKTVDHTLLATTATWPEIQTILDDAMAYETASACIPASYVKKAAEYVSGKLAIC 60
           +EVKDILKTVDHTLLATTATWPEIQTILDDAMAYETASACIPASYVKKAAEYVSGKLAIC
Sbjct: 1   VEVKDILKTVDHTLLATTATWPEIQTILDDAMAYETASACIPASYVKKAAEYVSGKLAIC 60

Query: 61  TVIGFPNGYSTTAAKVFECQDAIKNGADEIDMVINLTDVKNGDFDTVEEEIRQIKAACQD 120
           TVIGFPNGYSTTAAKVFECQDAI+NGADEIDMVINLTDVKNGDFDTVEEEIRQIKA CQD
Sbjct: 61  TVIGFPNGYSTTAAKVFECQDAIQNGADEIDMVINLTDVKNGDFDTVEEEIRQIKAKCQD 120

Query: 121 HILKVIVETCQLTKEELIELCGVVTRSGADFIKTSTGFSTAGATFEDVEVMAKYVGEGVK 180
           HILKVIVETCQLTKEELIELCGVVTRSGADFIKTSTGFSTAGATFEDVEVMAKYVGEGVK
Sbjct: 121 HILKVIVETCQLTKEELIELCGVVTRSGADFIKTSTGFSTAGATFEDVEVMAKYVGEGVK 180

Query: 181 IKAAGGISSLEDAEKFIALGASRLGTSRIIKIVKNQKVEEGTY 223
           IKAAGGISSLEDA+ FIALGASRLGTSRIIKIVKN+  +  +Y
Sbjct: 181 IKAAGGISSLEDAKTFIALGASRLGTSRIIKIVKNEATKTDSY 223
```

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

Example 2020

A DNA sequence (GBSx2131) was identified in *S.agalactiae* <SEQ ID 6251> which encodes the amino acid sequence <SEQ ID 6252>. This protein is predicted to be phosphopentomutase (deoB). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0546(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:AAC45496 GB:U80410 phosphopentomutase [Lactococcus lactis subsp.
           cremoris]
 Identities = 275/408 (67%), Positives = 325/408 (79%), Gaps = 7/408 (1%)

Query: 3   QFDRIHLVVLDSVGIGAAPDANDFVNAGVP------DGASDTLGHISKTVGLAVPNMAKI 56
           +F RIHLVV+DSVGIGAAPDA+ F N  V        D  SDT+GHIS+  GL VPN+ K+
Sbjct: 4   KFGRIHLVVMDSVGIGAAPDADKFFNHDVETHEAINDVKSDTIGHISEIRGLDVPNLQKL 63

Query: 57  GLGNIPRPQALKTVPAEENPSGYATKLQEVSLGKDTMTGHWEIMGLNITEPFDTFWNGFP 116
           G GNIPR   LKT+PA + P+ Y TKL+E+S GKDTMTGHWEIMGLNI  PF T+  G+P
Sbjct: 64  GWGNIPRESPLKTIPAAQKPAAYVTKLEEISKGKDTMTGHWEIMGLNIQTPFPTYPEGYP 123

Query: 117 EDIITKIEDFSGRKVIREANKPYSGTAVIDDFGPRQMETGELIIYTSADPVLQIAAHEDI 176
           ED++ KIE+FSGRK+IREANKPYSGTAVI+DFGPRQ+ETGELIIYTSADPVLQIAAHED+
```

```
Sbjct: 124 EDLLEKIEEFSGRKIIREANKPYSGTAVIEDFGPRQLETGELIIYTSADPVLQIAAHEDV 183

Query: 177 IPLEELYRICEYARSITMERPALL-GRIIARPYVGEPGNFTRTANRHDYAVSPFEDTVLN 235
           I  EELY+ICEY RSIT+E   ++ GRIIARPYVGE GNF RT  R DYA+SPF +TVL
Sbjct: 184 ISREELYKICEYVRSITLEGSGIMIGRIIARPYVGEAGNFERTDGRRDYALSPFAETVLE 243

Query: 236 KLDQAGIDTYAVGKINDIFNGSGINHDMGHNKSNSHGIDTLIKTMGLSEFEKGFSFTNLV 295
           KL +AGIDTY+VGKI+DIFN  G+ +DMGHN ++  G+D L+K M  +EF +GFSFTNLV
Sbjct: 244 KLYKAGIDTYSVGKISDIFNTVGVKYDMGHNHNDMDGVDRLLKAMTKTEFTEGFSFTNLV 303

Query: 296 DFDALYGHRRDPHGYRDCLHEFDERLPEIISAMRDKDLLLITADHGNDPTYAGTDHTREY 355
           DFDA YGHRRD  GY   + +FD RLPEII AM++ DLL+ITADHGNDP+Y GTDHTREY
Sbjct: 304 DFDAKYGHRRDVEGYGKAIEDFDGRLPEIIDAMKEDDLLMITADHGNDPSYVGTDHTREY 363

Query: 356 IPLLAYSPSFTGNGLIPVGHFADISATVADNFGVDTAMIGESFLQDLV 403
           IPL+ +S SF    ++PVGHFADISAT+A+NF V  A  GESFL  LV
Sbjct: 364 IPLVIFSKSFKEPKVLPVGHFADISATIAENFSVKKAQTGESFLDALV 411
```

There is also homology to SEQ ID 2740:

```
Identities = 348/402 (86%), Positives = 374/402 (92%)

Query: 1   MSQFDRIHLVVLDSVGIGAAPDANDFVNAGVPDGASDTLGHISKTVGLAVPNMAKIGLGN 60
           MS+F+RIHLVVLDSVGIGAAPDA+ F NAGV D  SDTLGHIS+  GL+VPNMAKIGLGN
Sbjct: 1   MSKFNRIHLVVLDSVGIGAAPDADKFFNAGVADTDSDTLGHISEAAGLSVPNMAKIGLGN 60

Query: 61  IPRPQALKTVPAEENPSGYATKLQEVSLGKDTMTGHWEIMGLNITEPFDTFWNGFPEDII 120
           I RP  LKTVP E+NP+GY TKL+EVSLGKDTMTGHWEIMGLNITEPFDTFWNGFPE+I+
Sbjct: 61  ISRPIPLKTVPTEDNPTGYVTKLEEVSLGKDTMTGHWEIMGLNITEPFDTFWNGFPEEIL 120

Query: 121 TKIEDFSGRKVIREANKPYSGTAVIDDFGPRQMETGELIIYTSADPVLQIAAHEDIIPLE 180
           TKIE+FSGRK+IREANKPYSGTAVIDDFGPRQMETGELI+YTSADPVLQIAAHEDIIP+E
Sbjct: 121 TKIEEFSGRKIIREANKPYSGTAVIDDFGPRQMETGELIVYTSADPVLQIAAHEDIIPVE 180

Query: 181 ELYRICEYARSITMERPALLGRIIARPYVGEPGNFTRTANRHDYAVSPFEDTVLNKLDQA 240
           ELY+ICEYARSIT+ERPALLGRIIARPYVG+PGNFTRTANRHDYAVSPF+DTVLNKL  A
Sbjct: 181 ELYKICEYARSITLERPALLGRIIARPYVGDPGNFTRTANRHDYAVSPFQDTVLNKLADA 240

Query: 241 GIDTYAVGKINDIFNGSGINHDMGHNKSNSHGIDTLIKTMGLSEFEKGFSFTNLVDFDAL 300
           G+ TYAVGKINDIFNGSGI +DMGHNKSNSHGIDTLIKT+ L EF KGFSFTNLVDFDA
Sbjct: 241 GVPTYAVGKINDIFNGSGITNDMGHNKSNSHGIDTLIKTLQLPEFTKGFSFTNLVDFDAN 300

Query: 301 YGHRRDPHGYRDCLHEFDERLPEIISAMRDKDLLLITADHGNDPTYAGTDHTREYIPLLA 360
           +GHRRDP GYRDCLHEFD RLPEII+ M++ DLLLITADHGNDPTYAGTDHTREYIPLLA
Sbjct: 301 FGHRRDPEGYRDCLHEFDNRLPEIIANMKEDDLLLITADHGNDPTYAGTDHTREYIPLLA 360

Query: 361 YSPSFTGNGLIPVGHFADISATVADNFGVDTAMIGESFLQDL 402
           YS SFTGNGLIP GHFADISATVA+NFGVDTAMIGESFL  L
Sbjct: 361 YSVSFTGNGLIPQGHFADISATVAENFGVDTAMIGESFLSHL 402
```

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

Example 2021

A DNA sequence (GBSx2132) was identified in *S.agalactiae* <SEQ ID 6253> which encodes the amino acid sequence <SEQ ID 6254>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.05   Transmembrane    9 -  25 (   4 -  35)

----- 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 no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6255> which encodes the amino acid sequence <SEQ ID 6256>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.57   Transmembrane   41 -  57 (  38 -  60)

----- Final Results -----
               bacterial membrane --- Certainty=0.3230(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 9143> which encodes the amino acid sequence <SEQ ID 9144>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 49
>>> Seems to have an uncleavable N-term signal seq
     INTEGRAL    Likelihood = -5.57   Transmembrane   13 -  29 (  10 -  32)

----- Final Results -----
               bacterial membrane --- Certainty= 0.323(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 = 276/544 (50%), Positives = 368/544 (66%), Gaps = 5/544 (0%)

Query: 5   FKKKVVKVCLVIFGIVLVSLLSLGFFYFSKGQVLSRFVAARSRTSGQAFDNIKEYMVWSD 64
           F  K +K   +I    L   L  G FY+SK   ++ ++ ARS  SG  F+NIK Y+VW D
Sbjct: 33  FHHKKLKQITIIAATSLFLFLIGGAFYYSKNHCINAYLKARSAQSGPVFENIKAYLVWDD 92

Query: 65  TGESITNDEANYANFEPLSKSEARKLGQEIKEGNKNDSMYLKRVGSRLGIFPDYRIANKP 124
           T E ITNDEA Y  F   S+ E R+  Q++K  +++ ++ +K VG R  IFPDYRIA KP
Sbjct: 93  TNEQITNDEAMYTKFRRYSQKELRQKKQDLKAASQDSAVQVKSVGRRFWIFPDYRIAIKP 152

Query: 125 MSLTLKTNVPKLDVLLNQKKVATSNSDHFSVTVERLPRTHYTASLEGTSDGKEIKLKKDY 184
           M LT+KTNVP+ DVLLNQKKVA S+S+ FSV ++RLP   YTAS+ G  +G+ IK+ K Y
Sbjct: 153 MDLTIKTNVPQADVLLNQKKVAVSDSEQFSVKLDRLPTAEYTASIRGKHNGRNIKVNKSY 212

Query: 185 DGKNQTIDLSVAFKSFTVTSNLMDGNLYFGDNRIAKLKDGSHSVENYPVTDGSKAYIKKV 244
           DG N  +DLSV+F++F VTSN   G+LYF DN I  LKDG   VE+YPVT+ ++AY+K
Sbjct: 213 DGDNPVLDLSVSFRTFLVTSNAKQGDLYFDDNHIGTLKDGQLQVEDYPVTENAQAYMKTT 272

Query: 245 FNDGEITSHKQKLISIADNQTIKLDVDGLLNEKEAGQKLITAFNQLILYVSTGQDPQTLG 304
           F.DGE+ S K  L  + +  T+++ V  LL E +AG+ L++AF+QL+ Y+STGQD   L
Sbjct: 273 FPDGELRSQKYALADVEEGATLEILVTDLLEEDKAGELLVSAFDQLMHYLSTGQDSSNLR 332

Query: 305 TVFEKGAENDFYKGLKESIKAKFVTDNRKASHFTIPNIVLNKMTQVGKESYQVNFAADYD 364
           +VFE G+ N FY+GLKESIKAKF TD RKAS   IP+I+L  MTQVGK +Y ++F A Y+
Sbjct: 333 SVFEAGSSNAFYRGLKESIKAKFQTDTRKASRLNIPSILLTTMTQVGKTTYVLDFTATYE 392

Query: 365 FNYDKSTDPDKKTYGHIIQNLTGNFIMKKSGNSYLISNDGKKDITVAKETNKVKADPVSI 424
           F YDKSTDP++ T GHI Q+LTG   +KK G  YLIS  G K+ITV KE N++KA   S+
Sbjct: 393 FLYDKSTDPEQHTSGHINQDLTGKVTVKKVGQHYLISQSGSKNITVVKEDNQLKAP--SV 450

Query: 425 FPENLVGSWKGEVEDGTVIMTFDKDGKVTQK-KVYKDSKSKESNHSAKVTKLEDKGNGLY 483
           FPE+++G+W G+     ++ M+   DG +T K +  K ++SKE+  +AK++K+EDKGNG Y
Sbjct: 451 FPESILGTWTGQANGLSIHMSLASDGTITTKVEDQKGNRSKET-RTAKISKVEDKGNGFY 509

Query: 484 LYQYESGTDTTTFV-TGGIGGLKVKYAYGIKIEGNKIIPVIWQTSSDGEFDYHKPLLSKP 542
           LY  + G+D +  V  GG+GG  VKYAYG KI G     PV+WQ +   EFDY KPL
Sbjct: 510 LYTPDPGSDISALVPEGGLGGANVKYAYGFKISGKTASPVVWQAALTHEFDYTKPLSGVT 569
```

```
Query: 543 LTKQ 546
          L KQ
Sbjct: 570 LQKQ 573
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9065> which encodes amino acid sequence <SEQ ID 9066>. An alignment of the GAS and GBS sequences follows:

```
 Score = 47.3 bits (110), Expect = 4e-07
 Identities = 65/303 (21%), Positives = 119/303 (38%), Gaps = 18/303 (5%)

Query: 153 FYILGIGTSISIVVALTRFVKEISLNFKEIKKLANKMGIEVLSENENYSQII---EFDDI 209
           +YIL + T I+ +V     +  +S  F  +KKL  KM         +  +QI    EF D+
Sbjct: 37  YYILSV-TIIACIVGGIVNLFLLSSVFTSLKKLKQKMKDISQRCFDTKAQICSPQEFKDL 95

Query: 210 LRTLHIKGDNLKSLIEREILEKQDLSFQIAALSHDIKTPXXXXXXXXXXXXXXXXXXXQE 269
               +     L+S  +     +++ +  IA LSHDIKTP                    +
Sbjct: 96  ETAFNQMSSELESTFKSLNESEREKTMMIAQLSHDIKTPITSIQSTVEGILDGIISEEEV 155

Query: 270 GYIVSMNNSISVFEGYFNSLISYTRML--------SEDRSVKLILVEELLSELHFEVDDL 321
            Y +   N+IS      N L+     +        +E    + I +++LL ++  E   +
Sbjct: 156 NYYL---NTISRQTNRLNHLVEELSFITLETMSDTAEPHKEETIYLDKLLIDILSEFQLV 212

Query: 322 LNINNIEFSICNRLIITSFYGDEENLIRALSNLLVNAIRFMPVLDKKIEVILSESGEQIH 381
               N +  I     ++      + L R L NL+ NA ++       + +    + + I
Sbjct: 213 FEKENRQVMIDVAPDVSKLSSQYDKLSRILLNLISNAXKYSDP-GSPLTIKAYSNRQDIV 271

Query: 382 FEIWNNGERFSDSTLKKGDKLFYTEDYSRGNK--HYGIGLAFVKGVAIKHGGNLQLNNPA 439
            +I + G    D  L       Y  + SR  K   +G+GL  + +A +  G++ + +
Sbjct: 272 IDIIDQGYGIKDEDLASIFNRLYRVESSRNMKTGGHGLGLYIARQLAHQLNGDILVESQY 331

Query: 440 RGG 442
           + G
Sbjct: 332 QKG 334
```

A related sequence was also identified in GAS <SEQ ID 9135> which encodes the amino acid sequence <SEQ ID 9136>. 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  145 - 161 ( 145 - 164)

----- 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>
```

SEQ ID 6254 (GBS280) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 8; MW 63.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 7; MW 88.7kDa).

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

Example 2022

A DNA sequence (GBSx2133) was identified in *S.agalactiae* <SEQ ID 6257> which encodes the amino acid sequence <SEQ ID 6258>. This protein is predicted to be ribosomal large subunit pseudouridine synthase D (rluC). Analysis of this protein sequence reveals the following:

```
Possible site: 22
```

```
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -4.62   Transmembrane    2 -  18 (   1 -  19)

----- Final Results -----
               bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB12749 GB:Z99108 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 97/251 (38%), Positives = 147/251 (57%), Gaps = 15/251 (5%)

Query: 86  KHVLINNEFINWQTVVQENDTITLIFDDEDYPTKKIPLGRAELIDCLYEDEHLIIVNKPE 145
           + + +N+E +    +V++ D + +   + +  +     G    +D L+ED H++I+NKP
Sbjct: 43  QQIKVNHESVLNNMIVKKGDRVFIDLQESEASSVIPEYGE---LDILFEDNHMLIINKPA 99

Query: 146 GMKTHGNQPNEIALLNHVSAY----SGQTCYV--VHRLDMETSGAVLFAKNPFILPLINQ 199
           G+ TH N+  +   L ++ AY    +G+TC V  VHRLD +TSGA++FAK+      +++Q
Sbjct: 100 GIATHPNEDGQTGTLANLIAYHYQINGETCKVRHVHRLDQDTSGAIVFAKHRLAHAILDQ 159

Query: 200 RLERKEIWREYWALVEGKFSPKHQVLRDKIGRNR-HDRRKRIIDSKNGQHAMTIIDVL-- 256
           +LE+K + R Y A+ EGK   K   +   IGR+R H  R+R+  S  GQ A+T   V+
Sbjct: 160 QLEKKTLKRTYTAIAEGKLRTKKGTINPPIGRDRSHPTRRRV--SPGGQTAVTHFKVMAS 217

Query: 257 KYIQNSSLIKCRLETGRTHQIRVHLSHHGHPLIGDPLYNPSSN-NERLMLHAHRLTLSHP 315
              +  SL++  LETGRTHQIRVHL+  GHPL GD LY   S    R  LHA+++   HP
Sbjct: 218 NAKERLSLVELELETGRTHQIRVHLASLGHPLTGDSLYGGGSKLLNRQALHANKVQAVHP 277

Query: 316 LTCETISVEAP 326
           +T E I  EAP
Sbjct: 278 ITDELIVAEAP 288
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6259> which encodes the amino acid sequence <SEQ ID 6260>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4198(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 = 172/278 (61%), Positives = 212/278 (75%), Gaps = 2/278 (0%)

Query: 63  TVKELLEDYFLIPRKIRHFLRVKKHVLINNEFINWQTVVQENDTITLIFDDEDYPTKKIP 122
           TVK LLE+  LIPRKIRHFLR KKHVLIN   +NWQ+ V+  D + L FD EDYP K I
Sbjct: 2   TVKALLEEQLLIPRKIRHFLRTKKHVLINGHSVNWQSCVKYGDQVKLFFDHEDYPEKIIV 61

Query: 123 LGRAELIDCLYEDEHLIIVNKPEGMKTHGNQPNEIALLNHVSAYSGQTCYVVHRLDMETS 182
           +G+AE + CLYEDEH+IIVNKPEGMKTHGN P E+ALLNHVSAY+GQTCYVVHRLD ETS
Sbjct: 62  MGQAEKVTCLYEDEHIIIVNKPEGMKTHGNDPTELALLNHVSAYTGQTCYVVHRLDKETS 121

Query: 183 GAVLFAKNPFILPLINQRLERKEIWREYWALVEGKFSPKHQVLRDKIGRNRHDRRKRIID 242
           GA+LFAK PFILP++N+ LE+++I REY ALV G             IGR+RHDRRKR++D
Sbjct: 122 GAILFAKTPFILPILNRLLEKRDIHREYLALVHGSLDSPRVTYHHPIGRHRHDRRKRVVD 181

Query: 243 SKNGQHAMTIIDVLK-YIQNSSLIKCRLETGRTHQIRVHLSHHGHPLIGDPLY-NPSSNN 300
             NG+ A+T + ++K + + +SL+ C+L+TGRTHQIRVHL+H GH L GDPLY N   +
Sbjct: 182 PINGKKAITEVTLVKNFHKTASLLTCQLQTGRTHQIRVHLAHQGHVLFGDPLYSNGKKDC 241

Query: 301 ERLMLHAHRLTLSHPLTCETISVEAPSSTFEKILNNYK 338
           RLMLHA++L L HPLT E I V+A S+TF+ +LN  K
Sbjct: 242 ARLMLHAYQLRLKHPLTQEDICVQAKSATFDAVLNAQK 279
```

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

Example 2023

A DNA sequence (GBSx2134) was identified in *S.agalactiae* <SEQ ID 6261> which encodes the amino acid sequence <SEQ ID 6262>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.02   Transmembrane   98 - 114 (  93 - 119)

----- Final Results -----
               bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAF04735 GB:AF101780 penicillin-binding protein 2a
           [Streptococcus pneumoniae]
 Identities = 424/773 (54%), Positives = 555/773 (70%), Gaps = 47/773 (6%)

Query: 2   KLFDKFIDLFRVDEDNDEMTRKNEQETREETSNLDGEEVYDIDDITRPSKSQYQRGIRHQ 61
           KLF+KF+ LF+                  +ETS L+  +      I R S+S
Sbjct: 5   KLFEKFLSLFK-----------------KETSELEDSD----STILRRSRS--------- 34

Query: 62  KENAKSRPEWLQKVDRYLPSPKNPIRRFWRRYRIGKLLFIALMAFILIFGSYLFYLSKTA 121
                       DR   +   PIR+FWRRY + K++ I  ++  L+ G YLF ++K+
Sbjct: 35  --------------DRKKLAQVGPIRKFWRRYHLTKIILILGLSAGLLVGIYLFAVAKST 80

Query: 122 TVSDLQSALKTTTTIYDKNKEYAGKLSGQKGTYVELNAISDHLKNAVIATEDRTFYENNG 181
            V+DLQ+ALKT T I+D+ ++ AG LSGQKGTYVEL  IS +L+NAVIATEDR+FY+N+G
Sbjct: 81  NVNDLQNALKTRTLIFDREEKEAGALSGQKGTYVELTDISKNLQNAVIATEDRSFYKNDG 140

Query: 182 VNFKRFFLAVATLGKFGGGSTITQQLAKNAYLSQDQTIKRKAREFFLALELTKKYSKAEI 241
           +N+ RFFLA+ T G+ GGGSTITQQLAKNAYLSQDQT++RKA+EFFLALEL+KKYSK +I
Sbjct: 141 INYGRFFLAIVTAGRSGGGSTITQQLAKNAYLSQDQTVERKAKEFFLALELSKKYSKEQI 200

Query: 242 LTMYLNNSYFGNGVWGVEDASRKYFGTSAANLTVDEAATLAGMLKGPEVYNPYYSVENAT 301
           LTMYLNN+YFGNGVWGVEDAS+KYFG SA+ +++D+AATLAGMLKGPE+YNP  SVE++T
Sbjct: 201 LTMYLNNAYFGNGVWGVEDASKKYFGVSASEVSLDQAATLAGMLKGPELYNPLNSVEDST 260

Query: 302 NRRDTVLAAMVDAGKLTKSQAKEAASIGMKNRLADTYAGKINDYRYPSYFDAVVNEAIDT 361
           NRRDTVL  MV AG + K+Q  EAA + M ++L D Y GKI+DYRYPSYFDAVVNEA+
Sbjct: 261 NRRDTVLQNMVAAGYIDKNQETEAAEVDMTSQLHDKYEGKISDYRYPSYFDAVVNEAVSK 320

Query: 362 YGISEKDIVNNGYKIYTALDQNYQSGMQKTFDDTSLFPVSDYDGQSAQGASVALDPKTGG 421
           Y ++E++IVNNGY+IYT LDQNYQ+ MQ  +++TSLFP ++ DG  AQ  SVAL+PKTGG
Sbjct: 321 YNLTEEEIVNNGYRIYTELDQNYQANMQIVYENTSLFPRAE-DGTFAQSGSVALEPKTGG 379

Query: 422 VRGLVGRVQSTKDAQFRSFNYATQSKRSPASTIKPLVVYSPAIASGWSIDKELPNKVQDF 481
           VRG+VG+V        FR+FNYATQSKRSP STIKPLVVY+PA+ +GW+++K+L N    +
Sbjct: 380 VRGVVGQVADNDKTGFRNFNYATQSKRSPGSTIKPLVVYTPAVEAGWALNKQLDNHTMQY 439

Query: 482 HGYKPSNYGGIET-ESIPMYQALANSYNIPAVYTLDKLGINKAFTYGRKFGLNMSSANKE 540
             YK  NY GI+T   +PMYQ+LA S N+PAV T++ LG++KAF  G KFGLNM   ++
Sbjct: 440 DSYKVDNYAGIKTSREVPMYQSLAESLNLPAVATVNDLGVDKAFEAGEKFGLNMEKVDRV 499

Query: 541 LGVALGGSVTTNPLEMAQAYSTFANDGIMHRAHLITRIETANGKLVKQFTDKPKRVISRS 600
           LGVALG  V TNPL+MAQAY+ FAN+G+M  AH I+RIE A+G+++    +  KRVI +S
Sbjct: 500 LGVALGSGVETNPLQMAQAYAAFANEGLMPEAHFISRIENASGQVIASHKNSQKRVIDKS 559

Query: 601 VASKMTSMMLGTFSNGTAINANVYGYTMAGKTGTTETDFNPNLSGDQWVVGYTPDVVISQ 660
           VA KMTSMMLGTF+NGT I+++   Y MAGKTGTTE  FNP  + DQWV+GYTPDVVIS
```

```
Sbjct: 560 VADKMTSMMLGTFTNGTGISSSPADYVMAGKTGTTEAVFNPEYTSDQWVIGYTPDVVISH 619

Query: 661 WVGFKNTDKHHYLTDSSAGTASNIFSTQASYILPYTKGSSFTHIENAYFQNGIGSVYNAQ 720
           W+GF  TD++HYL  S++  A+++F   A+ ILPYT GS+FT +ENAY QNGI      +
Sbjct: 620 WLGFPTTDENHYLAGSTSNGAAHVFRNIANTILPYTPGSTFT-VENAYKQNGIAPANTKR 678

Query: 721 DASNTTNQESRSIINDLKDSASKAAQDISRAVEDSNFQEKVKDAWNSLKDYFR 773
                N ++   ++D++  A     + SRA+ D+  +EK +  W+S+ + FR
Sbjct: 679 QVQTNDNSQTDDNLSDIRGRAQSLVDEASRAISDAKIKEKAQTIWDSIVNLFR 731
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6263> which encodes the amino acid sequence <SEQ ID 6264>. Analysis of this protein sequence reveals the following:

```
      Possible site: 52

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.96   Transmembrane  104 - 120 (  99 - 124)

----- Final Results -----
            bacterial membrane --- Certainty=0.4185(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAF04735 GB:AF101780 penicillin-binding protein 2a [Streptococcus pneumoniae]
 Identities = 414/730 (56%), Positives = 539/730 (73%), Gaps = 17/730 (2%)

Query: 50  TKNSEQDPATALQRSRAYEGSPKSRPAWLQKLEAVLPSPQRPIRRFWRRYHIGKLLMILI 109
           T   E   +T L+RSR+            +KL  V      PIR+FWRRYH+ K+++IL
Sbjct: 18  TSELEDSDSTILRRSRSDR----------KKLAQV-----GPIRKFWRRYHLTKIILILG 62

Query: 110 GTLVLLLGSYLFYLSKTAKVSDLQDALKATTVIYDHKGEYAGSLSGQKGSYVELNAISDD 169
            +  LL+G YLF ++K+  V+DLQ+ALK  T+I+D + + AG+LSGQKG+YVEL  IS +
Sbjct: 63  LSAGLLVGIYLFAVAKSTNVNDLQNALKTRTLIFDREEKEAGALSGQKGTYVELTDISKN 122

Query: 170 LENAVIATEDRTFYSNSGINLKRFLLAVVTAGRFGGGSTITQQLAKNAYLSQDQTIKRKA 229
           L+NAVIATEDR+FY N GIN  RF LA+VTAGR GGGSTITQQLAKNAYLSQDQT++RKA
Sbjct: 123 LQNAVIATEDRSFYKNDGINYGRFFLAIVTAGRSGGGSTITQQLAKNAYLSQDQTVERKA 182

Query: 230 REFFLALELTKKYSKKDILTMYLNNSYFGNGVWGVEDASQKYFGTTAANLTLDEAATLAG 289
           +EFFLALEL+KKYSK+ ILTMYLNN+YFGNGVWGVEDAS+KYFG +A+ ++LD+AATLAG
Sbjct: 183 KEFFLALELSKKYSKEQILTMYLNNAYFGNGVWGVEDASKKYFGVSASEVSLDQAATLAG 242

Query: 290 MLKGPEIYNPYHSLKNATHRRDTVLGAMVDAKKITQTKAQQARAVGLKNRLADTYVGKTD 349
           MLKGPE+YNP +S++++T+RRDTVL  MV A  I + +  +A  V + ++L D Y GK
Sbjct: 243 MLKGPELYNPLNSVEDSTNRRDTVLQNMVAAGYIDKNQETEAAEVDMTSQLHDKYEGKIS 302

Query: 350 DYKYPSYFDAVISEAIATYGLSEKDIVNNGYKVYTELDQNYQTGMQTTFNNDELFPVSAY 409
           DY+YPSYFDAV++EA++ Y L+E++IVNNGY++YTELDQNYQ  MQ  + N  LFP  A
Sbjct: 303 DYRYPSYFDAVVNEAVSKYNLTEEEIVNNGYRIYTELDQNYQANMQIVYENTSLFP-RAE 361

Query: 410 DGSSAQAASVALDPKTGGVRGLIGRVNSSENPTFRSFNYATQAKRSPASTIKPLVVYAPA 469
           DG+ AQ+ SVAL+PKTGGVRG++G+V  ++   FR+FNYATQ+KRSP STIKPLVVY PA
Sbjct: 362 DGTFAQSGSVALEPKTGGVRGVVGQVADNDKTGFRNFNYATQSKRSPGSTIKPLVVYTPA 421

Query: 470 VASGWSIEKELPNTVQDFDGYQPHNY-GNYESEDVPMYQALANSYNIPAVSTLNDIGIDK 528
           V +GW++ K+L N     +D Y+  NY G   S +VPMYQ+LA S N+PAV+T+ND+G+DK
Sbjct: 422 VEAGWALNKQLDNHTMQYDSYKVDNYAGIKTSREVPMYQSLAESLNLPAVATVNDLGVDK 481

Query: 529 AFTYGKTFGLDMSSAKKELGVALGGSVTTNPLEMAQAYAAFANNGVIHPAHLINRIENAR 588
           AF  G+ FGL+M     + LGVALG  V TNPL+MAQAYAAFAN G++  AH I+RIENA
Sbjct: 482 AFEAGEKFGLNMEKVDRVLGVALGSGVETNPLQMAQAYAAFANEGLMPEAHFISRIENAS 541

Query: 589 GEVLKTFTDKAKRVVSQSVADKMTAMMLGTFSNGTAVNANVYGYTLAGKTGTTETNFNPD 648
           G+V+ +  +  KRV+ +SVADKMT+MMLGTF+NGT ++++   Y +AGKTGTTE  FNP+
Sbjct: 542 GQVIASHKNSQKRVIDKSVADKMTSMMLGTFTNGTGISSSPADYVMAGKTGTTEAVFNPE 601
```

```
Query: 649 LAGDQWVIGYTPDVVISQWVGFNQTDENHYLTDSSAGTASAIFSTQASYILPYTKGSQFH 708
           DQWVIGYTPDVVIS W+GF  TDENHYL  S++  A+ +F   A+ ILPYT GS F
Sbjct: 602 YTSDQWVIGYTPDVVISHWLGFPTTDENHYLAGSTSNGAAHVFRNIANTILPYTPGSTFT 661

Query: 709 VDNAYAQNGISAVYGVNETGNQSGVDTQSIIDGLRKSAQEASQSLSKAVDQSGLRDKAQS 768
           V+NAY QNGI+      +       T   +  +R  AQ      S+A+  + +++KAQ+
Sbjct: 662 VENAYKQNGIAPANTKRQVQTNDNSQTDDNLSDIRGRAQSLVDEASRAISDAKIKEKAQT 721

Query: 769 IWKEIVDYFR 778
           IW  IV+ FR
Sbjct: 722 IWDSIVNLFR 731
```

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

```
Identities = 530/715 (74%), Positives = 623/715 (87%), Gaps = 1/715 (0%)

Query: 59  RHQKENAKSRPEWLQKVDRYLPSPKNPIRRFWRRYRIGKLLFIALMAFILIFGSYLFYLS 118
           R  + + KSRP WLQK++  LPSP+ PIRRFWRRY IGKLL I +    +L+ GSYLFYLS
Sbjct: 65  RAYEGSPKSRPAWLQKLEAVLPSPQRPIRRFWRRYHIGKLLMILIGTLVLLLGSYLFYLS 124

Query: 119 KTATVSDLQSALKTTTTIYDKNKEYAGKLSGQKGTYVELNAISDHLKNAVIATEDRTFYE 178
           KTA VSDLQ ALK TT IYD   EYAG LSGQKG+YVELNAISD L+NAVIATEDRTFY
Sbjct: 125 KTAKVSDLQDALKATTVIYDHKGEYAGSLSGQKGSYVELNAISDDLENAVIATEDRTFYS 184

Query: 179 NNGVNFKRFFLAVATLGKFGGGSTITQQLAKNAYLSQDQTIKRKAREFFLALELTKKYSK 238
           N+G+N KRF LAV T G+FGGGSTITQQLAKNAYLSQDQTIKRKAREFFLALELTKKYSK
Sbjct: 185 NSGINLKRFLLAVVTAGRFGGGSTITQQLAKNAYLSQDQTIKRKAREFFLALELTKKYSK 244

Query: 239 AEILTMYLNNSYFGNGVWGVEDASRKYFGTSAANLTVDEAATLAGMLKGPEVYNPYYSVE 298
           +ILTMYLNNSYFGNGVWGVEDAS+KYFGT+AANLT+DEAATLAGMLKGPE+YNPY+S++
Sbjct: 245 KDILTMYLNNSYFGNGVWGVEDASQKYFGTTAANLTLDEAATLAGMLKGPEIYNPYHSLK 304

Query: 299 NATNRRDTVLAAMVDAGKLTKSQAKEAASIGMKNRLADTYAGKINDYRYPSYFDAVVNEA 358
           NAT+RRDTVL AMVDA K+T+++A++A ++G+KNRLADTY GK +DY+YPSYFDAV++EA
Sbjct: 305 NATHRRDTVLGAMVDAKKITQTKAQQARAVGLKNRLADTYVGKTDDYKYPSYFDAVISEA 364

Query: 359 IDTYGISEKDIVNNGYKIYTALDQNYQSGMQKTFDDTSLFPVSDYDGQSAQGASVALDPK 418
           I TYG+SEKDIVNNGYK+YT LDQNYQ+GMQ TF++  LFPVS YDG SAQ ASVALDPK
Sbjct: 365 IATYGLSEKDIVNNGYKVYTELDQNYQTGMQTTFNNDELFPVSAYDGSSAQAASVALDPK 424

Query: 419 TGGVRGLVGRVQSTKDAQFRSFNYATQSKRSPASTIKPLVVYSPAIASGWSIDKELPNKV 478
           TGGVRGL+GRV S+++  FRSFNYATQ+KRSPASTIKPLVVY+PA+ASGWSI+KELPN V
Sbjct: 425 TGGVRGLIGRVNSSENPTFRSFNYATQAKRSPASTIKPLVVYAPAVASGWSIEKELPNTV 484

Query: 479 QDFHGYKPSNYGGIETESIPMYQALANSYNIPAVYTLDKLGINKAFTYGRKFGLNMSSAN 538
           QDF GY+P NYG  E+E +PMYQALANSYNIPAV TL+ +GI+KAFTYG+ FGL+MSSA
Sbjct: 485 QDFDGYQPHNYGNYESEDVPMYQALANSYNIPAVSTLNDIGIDKAFTYGKTFGLDMSSAK 544

Query: 539 KELGVALGGSVTTNPLEMAQAYSTFANDGIMHRAHLITRIETANGKLVKQFTDKPKRVIS 598
           KELGVALGGSVTTNPLEMAQAY+ FAN+G++H AHLI RIE A G+++K FTDK KRV+S
Sbjct: 545 KELGVALGGSVTTNPLEMAQAYAAFANNGVIHPAHLINRIENARGEVLKTFTDKAKRVVS 604

Query: 599 RSVASKMTSMMLGTFSNGTAINANVVGYTMAGKTGTTETDFNPNLSGDQWVVGYTPDVVI 658
           +SVA KMT+MMLGTFSNGTA+NANVVGYT+AGKTGTTET+FNP+L+GDQWV+GYTPDVVI
Sbjct: 605 QSVADKMTAMMLGTFSNGTAVNANVVGYTLAGKTGTTETNFNPDLAGDQWVIGYTPDVVI 664

Query: 659 SQWVGFKNTDKHHYLTDSSAGTASNIFSTQASYILPYTKGSSFTHIENAYFQNGIGSVYN 718
           SQWVGF  TD++HYLTDSSAGTAS IFSTQASYILPYTKGS F H++NAY QNGI +VY
Sbjct: 665 SQWVGFNQTDENHYLTDSSAGTASAIFSTQASYILPYTKGSQF-HVDNAYAQNGISAVYG 723

Query: 719 AQDASNTTNQESRSIINDLKDSASKAAQDISRAVEDSNFQEKVKDAWNSLKDYFR 773
             +  N +  +++SII+ L+ SA +A+Q +S+AV+ S  ++K +  W  + DYFR
Sbjct: 724 VNETGNQSGVDTQSIIDGLRKSAQEASQSLSKAVDQSGLRDKAQSIWKEIVDYFR 778
```

SEQ ID 6262 (GBS397d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 153 (lane 13; MW 76kDa) and in Figure 184 (lane 9; MW 76kDa).

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

Example 2024

A DNA sequence (GBSx2135) was identified in *S.agalactiae* <SEQ ID 6265> which encodes the amino acid sequence <SEQ ID 6266>. This protein is predicted to be M-like protein. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.56   Transmembrane  609 - 625 ( 599 - 628)
   INTEGRAL    Likelihood = -0.00   Transmembrane   19 -  35 (  19 -  35)

----- Final Results -----
               bacterial membrane --- Certainty=0.5225(Affirmative) < succ>
                bacterial outside --- 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:CAB91647 GB:AJ130830 cell wall protein, putative [Zea mays]
 Identities = 106/182 (58%), Positives = 123/182 (67%), Gaps = 8/182 (4%)

Query: 396 KEDKKPDVKPEAKPEAK--PDVKPEAKPDVKPEAKPDVKPEAKPDV--KPEA 451
           K + KP+ KPE KPE K  P  KPE KP+ KPE KP+ KPE KP  KPE KP+   KPE
Sbjct: 116 KPEPKPEPKPEPKPEPKIKPKPKPEPKPEPKPEHKPEPKPEPKPKPKPEPKPEPQPKPEP 175

Query: 452 KPDVKPKAKPDVKPEA--KPDVKPDVKPDVKPEA--KPEDKPDVKPDVKPEAKPDVKPEA 507
           KP+ KP+ KP+ KPE   KP+ KP+ KP+ KPE   KPE KP+ KP+ KPE KP+ KPE
Sbjct: 176 KPEPKPEPKPEPKPEPQPKPEPKPEPKPEPKPEPQPKPEPKPEPKPEPKPEPKPEPKPEP 235

Query: 508 KPEAKPEAKPEAKPEAKPEAKPDVKPEAKPDVKPEAKPEAKPEAKSEAKPEAKLEAKPEA 567
           KPE KPE +PE KPE KPE KP   P+ +P  KPE KPE KPE K E KPE K E KPE
Sbjct: 236 KPEPKPEPRPEPKPEPKPEPKPKPDPKPEPQPKPEPKPEPKPEPKPEPKPEPKPEPKPEP 295

Query: 568 KP 569
           KP
Sbjct: 296 KP 297
```

There is also homology to SEQ ID 822.

A related GBS gene <SEQ ID 8957> and protein <SEQ ID 8958> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     -5.20
GvH: Signal Score (-7.5): 3.07
     Possible site: 27
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value: -10.56 threshold:  0.0
   INTEGRAL    Likelihood =-10.56   Transmembrane  609 - 625 ( 599 - 628)
   INTEGRAL    Likelihood = -0.00   Transmembrane   19 -  35 (  19 -  35)
   PERIPHERAL  Likelihood =  8.54     139
 modified ALOM score:   2.61

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5225(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 596-600
```

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

```
ORF00748(313 - 2190 of 2490)
GP|2462785|gb|AAB71985.1||U73163(3 - 374 of 374) M-like protein (Streptococcus equi)
%Match = 9.2
%Identity = 36.0  %Similarity = 55.4
Matches = 126  Mismatches = 147  Conservative Sub.s = 68

282       312       342       372       402       432       462       666
LS**IRIFN*LYKGANMNNNEKKVKYFLRKTAYGLASMSAAFAVCSGIVHADTSSGISASIPHKKQVNLGAVT~~~~THA
                    :|::|::|||:|:||||:|||: | :  | ||:
                  MAKKEMKFYLRKSAFGLASVSAALLVGAARVSADS--------------------------
                           10        20        30

696       726       756       786       813       843       870       900
KVSDQELGKQSRRSQDIIKSLGFLSSDQKDILVKSISSSK-DSQLILKFVTQATQLNNAESTKAK-QMAQNDVALIKNIS
                  ::| | ::    | |    :::| :: |:       |       : ||:   || |:|  :
-----------------VESAGPVAVAVTDSLDSEAAATKAEADLVAAKADLAAAEVAITAAKAEFDTAQADLATAEATI
                  40        50        60        70        80        90

921       951       981      1011      1041      1071      1101      1131
PEV---LEEYKEKIQRASTKSQVDEFVAEAKKVVNSNKETLVNQANGKKQEIAKLENLSNDEMLRYNTAIDNVVKQYNEG
 |:   : | ::||| |  |   :   :  |:| : :::  | :    ::|  | |      : :   |:|   : |
AELEQKIPELEKKIQEAQEKLNYENRPS-PKRVGSDDEDDTVARKLMSEKEALKAE------LQKTKEALDTAKRAYAGI
         110       120       130       140       150             160       170

1161      1191      1221      1251      1281      1311      1638      1668
KLNITAAMNALNSIKQAAQEVAQKNLQKQYAKKIERISSKGLALSKKAKEIYEKHKSILPTP~~~~AKPDVKPEAKPDVK
:     |   |::  :|   | :|: |   |      : ||
EERKQVAATKLDAANKAFAGVEEKHAQAMAAFGAAFAAYKGA-----------------------------------
               190       200       210

1698                                      1740      1770      1800      1830
PKAKPDVKPEAKPDVKPD-------------------------VKPDVKPEAKPEDKPDVKPDVKPEAKPDVKPEAKPE
      || | |     |                          ||:|:|||||| | : ||: ||| ||: ||| |||
------VKAELKAAGASDFYTKKIDSADTVDGVKTLREMILDSIAKPEVEPEAKPEPKLEPKPEPKPEPKPEPKPEPKPE
         220       230       240       250       260       270       280

1860      1890      1920      1950      1980      2010      2040      2070
AKPEAKPEAKPEAKPEAKPDVKPEAKPDVKPEAKPEAKPEAKSEAKPEAKLEAKPEAKPATKKSVNTSGNLAAKKAIENK
 ||| ||| ||| ||| ||: ||: :|   | |||| | |  |                                  |
PKPEPKPEPKPEPKPEPKPEPKPKPQPKPAPAPKPEAKKEEKKAAP--------------------------------K
         300       310       320       330

2100      2130      2160      2190      2220      2250      2280      2310
KYSKKLPSTGEAASPLLAIVSLIVMLSAGLITIVLKHKKN*IYF*T*TERSILSKS*GKPHQNFAFFI*ILE*FSRYFN*
: : |||||||| :|::   :| ||  ||:  :  : |:|
QDTNKLPSTGEATNPFFTAAALAVMAGAGVAAVSTRRKEN
             350       360       370
```

SEQ ID 6266 (GBS3) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 5; MW 65kDa). The GBS3-His fusion product was purified (Figure 189, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 261), 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 2025

A DNA sequence (GBSx2136) was identified in *S.agalactiae* <SEQ ID 6267> which encodes the amino acid sequence <SEQ ID 6268>. This protein is predicted to be transcription antitermination protein nusg (nusG). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3203(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:CAA53738 GB:X76134 nusG [Staphylococcus carnosus]
 Identities = 90/175 (51%), Positives = 118/175 (67%), Gaps = 2/175 (1%)

Query: 7   KGWFVLQTYSGYENKVKENLLQRAQTYNMLDNILRVEIPTQTVNVEKNGKTKEIEENRFP 66
           K W+ + TYSGYENKVK+NL +R ++ NM + I RV IP +     K+GK K++ +  FP
Sbjct: 8   KRWYAVHTYSGYENKVKKNLEKRVESMNMTEQIFRVVIPEEEETQVKDGKAKKLTKKTFP 67

Query: 67  GYVLVEMVMTDEAWFVVRNTPNVTGFVGSHGNRSKPTPLLEEEIRSILISMGQTVDVFDT 126
           GYVLVE+VMTDE+W+VVRNTP VTGFVGS G  SKP PLL +E+R IL  MG      D
Sbjct: 68  GYVLVELVMTDESWYVVRNTPGVTGFVGSAGAGSKPNPLLPDEVRFILKQMGMKEKTIDV 127

Query: 127 NIKEGDVVQIIDGAFIGQEGRVVEIENNKVKL--MINMFGSETQAELELYQVAEL 179
            ++ G+ V+I  G F  Q G V EIE +K KL  +++MFG ET  E+E  Q+ +L
Sbjct: 128 EVEVGEQVRIKSGPFANQVGEVQEIEADKFKLTVLVDMFGRETPVEVEFDQIEKL 182
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6269> which encodes the amino acid sequence <SEQ ID 6270>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3874(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 = 170/179 (94%), Positives = 178/179 (98%)

Query: 1   MLDSFDKGWFVLQTYSGYENKVKENLLQRAQTYNMLDNILRVEIPTQTVNVEKNGKTKEI 60
           MLDSFDKGWFVLQTYSGYENKVKENLLQRAQTYNMLDNILRVEIPTQTVNVEKNG+TKEI
Sbjct: 6   MLDSFDKGWFVLQTYSGYENKVKENLLQRAQTYNMLDNILRVEIPTQTVNVEKNGQTKEI 65

Query: 61  EENRFPGYVLVEMVMTDEAWFVVRNTPNVTGFVGSHGNRSKPTPLLEEEIRSILISMGQT 120
           EENRFPGYVLVEMVMTDEAWFVVRNTPNVTGFVGSHGNRSKPTPLLEEEIR+IL+SMGQT
Sbjct: 66  EENRFPGYVLVEMVMTDEAWFVVRNTPNVTGFVGSHGNRSKPTPLLEEEIRAILLSMGQT 125

Query: 121 VDVFDTNIKEGDVVQIIDGAFIGQEGRVVEIENNKVKLMINMFGSETQAELELYQVAEL 179
           +DVFDTNIKEGDVVQIIDGAF+GQEGRVVEIENNKVKLM+NMFGSET AE+ELYQ+AEL
Sbjct: 126 IDVFDTNIKEGDVVQIIDGAFMGQEGRVVEIENNKVKLMLNMFGSETVAEVELYQIAEL 184
```

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

Example 2026

A DNA sequence (GBSx2137) was identified in *S.agalactiae* <SEQ ID 6271> which encodes the amino acid sequence <SEQ ID 6272>. This protein is predicted to be a glycosyl transferase. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1558(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:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus
          ducreyi]
 Identities = 98/259 (37%), Positives = 155/259 (59%), Gaps = 10/259 (3%)

Query: 5   VALAVDSNYLDKALVTIKSICVYNRNITFYLFNQDTPVEWVRNINRKLEPLGSKLINVKI 64
           + LA + +Y +  L TIKSI ++N++I FYL N+D P EW   +N KL  L S++I++K+
Sbjct: 10  IVLAANQSYSEYILTTIKSIYLHNKHIRFYLLNRDYPTEWFDILNNKLRKLNSEIIDIKV 69

Query: 65  YNYDIAHLTTFLTVS---TWFRLFLADYIPSSRVLYLDSDIIVNTNLDYLFELDFKGYYL 121
            N  I +  T+  +S   T+FR F++D+I   +V+YLD+DI+VN +L  L++ D   Y+L
Sbjct: 70  TNDTIKNFKTYSHISSDTTFFRYFISDFIEQDKVIYLDADIVVNGSLTELYQTDISNYFL 129

Query: 122 AAVKDPHKNE----EGGFNAGMLLANLELWREDGLTKTLLKTAEELHRVVKTGDQSILNI 177
           AAVKD    +       FNAGMLL N + WRE  +T+  L  +E+    +   DQSILN+
Sbjct: 130 AAVKDIISEKIYVNNHIFNAGMLLINNKKWREHNITQFCLSLSEKYINSLPDADQSILNL 189

Query: 178 VCHNRWLSLNKTWNF--QTYDVVSRYNHRSYLYLNIENRTPNIIHFLTSDKPWNENSVAR 235
           +  ++WL LN+ +N+   T  +  +Y    YL  ++    P IIH+ T  KPW      R
Sbjct: 190 IFKDKWLKLNRGYNYLIGTDYLFFKYGKTRYLE-DLGETIPLIIHYNTEAKPWLNIFNTR 248

Query: 236 FRELWWYYFQLDFCQLTGK 254
           FR ++W+Y++L++  +  K
Sbjct: 249 FRNIYWFYYELNWQDIYAK 267
```

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 2027

A DNA sequence (GBSx2138) was identified in *S.agalactiae* <SEQ ID 6273> which encodes the amino acid sequence <SEQ ID 6274>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

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

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 2028

A DNA sequence (GBSx2139) was identified in *S.agalactiae* <SEQ ID 6275> which encodes the amino acid sequence <SEQ ID 6276>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60   Transmembrane  306 - 322 ( 306 - 322)

----- 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 GENPEPT database.

```
>GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus
          ducreyi]
 Identities = 88/259 (33%), Positives = 156/259 (59%), Gaps = 11/259 (4%)

Query: 7   VVLAGDYSYIRQIETTLKSLCVYHENLSIFIFNQDIPQEWFLAMKDRVGQTGNQIQDVKL 66
           +VLA + SY    I TT+KS+ ++++++  ++ N+D P EWF  + +++ +  ++I D+K+
Sbjct: 10  IVLAANQSYSEYILTTIKSIYLHNKHIRFYLLNRDYPTEWFDILNNKLRKLNSEIIDIKV 69

Query: 67  FHDHLSPKWENKKLNHINY-MTYARYFIPQYISADTVLYLDSDLVVTTNLDNLFQISLDN 125
            +D +       K  +HI+   T+ RYFI  +I  D V+YLD+D+VV  +L  L+Q  + N
Sbjct: 70  TNDTIK---NFKTYSHISSDTTFFRYFISDFIEQDKVIYLDADIVVNGSLTELYQTDISN 126

Query: 126 AYLAAVP-----ALFGLGYGFNAGVMVINNQRWRQENMTIKLIEKNQKEIENANEGDQTI 180
            +LAAV       ++   + FNAG+++INN++WR+ N+T   +  ++K I +  + DQ+I
Sbjct: 127 YFLAAVKDIISEKIYVNNHIFNAGMLLINNKKWREHNITQFCLSLSEKYINSLPDADQSI 186

Query: 181 LNRMFENQVIYLDDTYNFQIGFD-MGAAIDGHKFIFDIPITPLPKIIHYISGIKPWQTLS 239
           LN +F+++ + L+  YN+ IG D +       +++ D+  T +P IIHY +  KPW  +
Sbjct: 187 LNLIFKDKWLKLNRGYNYLIGTDYLFFKYGKTRYLEDLGET-IPLIIHYNTEAKPWLNIF 245

Query: 240 NMRLREVWWHYNLLEWSSI 258
           N R R ++W Y  L W  I
Sbjct: 246 NTRFRNIYWFYYELNWQDI 264
```

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

SEQ ID 6276 (GBS395) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 5; MW 47.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 8; MW 72kDa) and in Figure 177 (lane 5; MW 72kDa).

GBS395-GST was purified as shown in Figure 217, lane 7.

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

Example 2029

A DNA sequence (GBSx2140) was identified in *S.agalactiae* <SEQ ID 6277> which encodes the amino acid sequence <SEQ ID 6278>. Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1633(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 2030

A DNA sequence (GBSx2141) was identified in *S.agalactiae* <SEQ ID 6279> which encodes the amino acid sequence <SEQ ID 6280>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL   Likelihood = -0.16   Transmembrane   36 -  52 (  36 -  52)

----- Final Results -----
             bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10243> which encodes amino acid sequence <SEQ ID 10244> was also identified.

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

```
>GP:AAC77330 GB:AE000508 orf, hypothetical protein [Escherichia coli K12]
 Identities = 75/260 (28%), Positives = 123/260 (46%), Gaps = 22/260 (8%)

Query: 6   VGLVLEGGGMRGLYTAGVLDAFLDAGIK-IDGIVSVSAGALFGVNFVSRQRERALRYNKK 64
           + LV EGGG RG++TAGVLD F+ A     D  +  SAGA     F+  Q   A +   +
Sbjct: 25  IALVCEGGGQRGIFTAGVLDEFMRAQFNPFDLYLGTSAGAQNLSAFICNQPGYARKVIMR 84

Query: 65  YLSHPKYMSLRSWFRTGNFVNKDF----TYYEVPMKLD----VFDDEAFKKSSIDFYVVA 116
           Y +  ++      + R GN ++ D+    T  ++P+++D    +FD       S   FY+ A
Sbjct: 85  YTTKREFFDPLRFVRGGNLIDLDWLVEATASQMPLQMDTAARLFD------SGKSFYMCA 138

Query: 117 TEMTSGKPEYFKIDSVFEQMEILRASSALPVVSKM-VDWQGKKYLDGGLSDSIPVDFARG 175
                 P YF + +     ++++RASSA+P   +  V  +G  YLDGG+SD+IPV  A
Sbjct: 139 CRQDDYAPNYF-LPTKQNWLDVIRASSAIPGFYRSGVSLEGINYLDGGISDAIPVKEAAR 197

Query: 176 LGFDKLIVVMTRPLNYQKKPSSGR-----LYKTLYRKYPNFVKTASNRYQQYNNSLEKVM 230
            G   L+V+ T P       P   +     L  +  +  N V+      Y+      +EK
Sbjct: 198 QGAKTLVVIRTVPSQMYYTPQWFKRMERWLGDSSLQPLVNLVQHHETSYRDIQQFIEKPP 257

Query: 231 SLEKTGDLFAIRPSKSLVIG 250
              +  +++  +P  S+ +G
Sbjct: 258 GKLRIFEIYPPKPLHSIALG 277
```

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

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

A related GBS gene <SEQ ID 8959> and protein <SEQ ID 8960> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      -5.16
GvH: Signal Score (-7.5): -2.17
     Possible site: 44
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -0.16 threshold:  0.0
   INTEGRAL    Likelihood = -0.16   Transmembrane   36 -  52 (  36 -  52)
   PERIPHERAL  Likelihood =  4.14       18
 modified ALOM score:   0.53

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
ORF01611(316 - 1050 of 1449)
OMNI|NT01EC5264(37 - 289 of 369) hypothetical protein
%Match = 9.2
%Identity = 29.7  %Similarity = 49.8
Matches = 74  Mismatches = 118  Conservative Sub.s = 50

273       303       333       363       393       420       450       480
QKKQLYFAIL*SNINIRK*LPMLSVGLVLEGGGMRGLYTAGVLDAFLDAGIK-IDGIVSVSAGALFGVNFVSRQRERALR
                          : || |||| ||::|||||| |: |     | : ||||     |: |   | :
VGQRIPVTLGNIAPLSLRPFQPGRIALVCEGGGQRGIFTAGVLDEFMRAQFNPFDLYLGTSAGAQNLSAFICNQPGYARK
                 30        40        50        60        70        80        90

510       540                 588       618       648       678       708
YNKKYLSHPKYMSLRSWFRTGNFVNKDF-----TYYEVPMKLDVFDDEAFKKSSIDFYVVATEMTSGKPEYFKIDSVFEQM
   :| :  ::     : | ||::: |:    |  ::|:::|      :  |   ||: |       | || : :    :
VIMRYTTKREFFDPLRFVRGGNLIDLDWLVEATASQMPLQMDT--AARLFDSGKSFYMCACRQDDYAPNYF-LPTKQNWL
                110       120       130       140       150       160

738       765       795       825       855       885       912       930
EILRASSALPVVSKM-VDWQGKKYLDGGLSDSIPVDFARGLGFDKLIVVMTRPLNYQKKPSS-GRLYKTL----YRKYPN
:::|||||:|   :  |  :|  |||||:||:|||  |   |   |:|: | |      |    |: : |     :   |
DVIRASSAIPGFYRSGVSLEGINYLDGGISDAIPVKEAARQGAKTLVVIRTVPSQMYYTPQWFKRMERWLGDSSLQPLVN
        180       190       200       210       220       230       240

960       990       1020      1050      1080      1110      1140      1170
FVKTASNRYQQYNNSLEKVMSLEKTGDLFAIRPSKSLVIGRLEKNPDKLDSIYQLGMKDAKSVMPELNSYLMK*RKQYFS
:|:     |:     :||     :  :::  :|  |: :|
LVQHHETSYRDIQQFIEKPPGKLRIFEIYPPKPLHSIALGSRIPALREDYKLGRLCGRYFLATVGKLLTEKAPLTRHLVP
        260       270       280       290       300       310       320
```

SEQ ID 8960 (GBS394) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 4; MW 34.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 7; MW 60kDa).

GBS394-GST was purified as shown in Figure 217, lane 6.

Example 2031

A DNA sequence (GBSx2142) was identified in *S.agalactiae* <SEQ ID 6281> which encodes the amino acid sequence <SEQ ID 6282>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3004(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 2032

A DNA sequence (GBSx2143) was identified in *S.agalactiae* <SEQ ID 6283> which encodes the amino acid sequence <SEQ ID 6284>. This protein is predicted to be transporter protein. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL     Likelihood = -6.85   Transmembrane  373 - 389 ( 370 - 395)
   INTEGRAL     Likelihood = -6.74   Transmembrane  168 - 184 ( 162 - 187)
   INTEGRAL     Likelihood = -6.32   Transmembrane  259 - 275 ( 257 - 280)
   INTEGRAL     Likelihood = -4.78   Transmembrane  286 - 302 ( 285 - 306)
   INTEGRAL     Likelihood = -3.19   Transmembrane   55 -  71 (  54 -  71)
   INTEGRAL     Likelihood = -2.97   Transmembrane   84 - 100 (  79 - 101)
   INTEGRAL     Likelihood = -2.87   Transmembrane  311 - 327 ( 310 - 328)
   INTEGRAL     Likelihood = -1.44   Transmembrane  355 - 371 ( 355 - 371)
   INTEGRAL     Likelihood = -0.64   Transmembrane  108 - 124 ( 108 - 125)

----- Final Results -----
               bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAC22759 GB:U32790 transporter protein [Haemophilus influenzae
          Rd]
 Identities = 139/391 (35%), Positives = 221/391 (55%), Gaps = 4/391 (1%)

Query: 6   INKNNWRALIAAIVASGTDDLNIMFLAFSMSTIITDLHLSAAQAGWIGTITNLGMLVGGL 65
           +N   W+ALI + V  G D  +++ L F +S I  DL+L+ AQ G + T T +G + GG+
Sbjct: 5   VNSYGWKALIGSAVGYGMDGFDLLILGFMLSAISADLNLTPAQGGSLVTWTLIGAVFGGI 64

Query: 66  IFGLLADRYNKFKVFKWTILIFSIATGLVFFTTNLSYLYIMRFIAGIGVGGEYGIAIAIM 125
           +FG L+D+Y + +V  WTIL+F++ TGL          L I R IAGIG+GGE+GI +A+
Sbjct: 65  LFGALSDKYGRVRVLTWTILLFAVFTGLCAIAQGYWDLLIYRTIAGIGLGGEFGIGMALA 124

Query: 126 AGIVPTNKMGRISSLNGIAGQVGSISSALLAGWLAPALGWRGLFLFGLLPIVLVLWMQFA 185
           A   P     + +S   +  QVG + +ALL   L P +GWRG+FL G+ P  +  +++
Sbjct: 125 AEAWPARHRAKAASYVALGWQVGVLGAALLTPLLLPHIGWRGMFLVGIFPAFVAWFLRSH 184

Query: 186 VDDKDILDQYNTDADDEPLDI----SIKALFDTPVLATQSLALMVMTTVQIAGYFGMMNW 241
           + + +I  Q  T    +        S + L      + SL ++V+T+VQ  GY+G+M W
Sbjct: 185 LHEPEIFTQKQTALSTQSSFTDKLRSFQLLIKDKATSKISLGIVVLTSVQNFGYYGIMIW 244

Query: 242 LPTIIQTNLNVSVKNSSLWMIATILGMCLGMLVFGQLLDKFGPRLVYGCFLLSSAICVYL 301
           LP  +   L  S+  S LW   T+ GM  G+ +FGQL D+ G +  +  F L + I + +
Sbjct: 245 LPNFLSKQLGFSLTKSGLWTAVTVCGMMAGIWIFGQLADRIGRKPSFLLFQLGAVISIVV 304

Query: 302 FQFATTMPSMIIGGAVVGFFVNGMFAGYGAMITRLYPHHIRSTANNLILNVGRAIGGFSS 361
           +  T    M++ GA +G FVNGM  GYGA++   YP   R+TA N++ N+GRA+GGF
Sbjct: 305 YSQLTDPDIMLLAGAFLGMFVNGMLGGYGALMAEAYPTEARATAQNVLFNIGRAVGGFGP 364

Query: 362 VIIGMILDVSNVSMVMLFLASLYIVSFLSML 392
           V++G ++   +  + LA +Y++  L+ +
Sbjct: 365 VVVGSVVLAYSFQTAIALLAIIYVIDMLATI 395
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2377> which encodes the amino acid sequence <SEQ ID 2378>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.92   Transmembrane  168 - 184 ( 162 - 188)
   INTEGRAL    Likelihood = -5.41   Transmembrane  286 - 302 ( 285 - 306)
   INTEGRAL    Likelihood = -5.15   Transmembrane  372 - 388 ( 362 - 394)
   INTEGRAL    Likelihood = -3.45   Transmembrane  259 - 275 ( 257 - 276)
   INTEGRAL    Likelihood = -2.87   Transmembrane  311 - 327 ( 306 - 328)
   INTEGRAL    Likelihood = -2.81   Transmembrane   55 -  71 (  51 -  71)
   INTEGRAL    Likelihood = -0.48   Transmembrane  108 - 124 ( 108 - 125)
   INTEGRAL    Likelihood = -0.37   Transmembrane   84 - 100 (  84 - 100)

----- 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>
```

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

```
Identities = 306/402 (76%), Positives = 354/402 (87%)

Query: 1   MSPLNINKNNWRALIAAIVASGTDDLNIMFLAFSMSTIITDLHLSAAQAGWIGTITNLGM 60
           MS L+++  N RAL+AAI ASGTDDLN+MFLAFSMS+I+TDL LS  Q GWI TITNLGM
Sbjct: 1   MSTLSLDTTNKRALVAAIAASGTDDLNVMFLAFSMSSIMTDLGLSGTQGGWIATITNLGM 60

Query: 61  LVGGLIFGLLADRYNKFKVFKWTILIFSIATGLVFFTTNLSYLYIMRFIAGIGVGGEYGI 120
           LVGGL+FGLLADR++KFKVFKWTIL+FS+ATGL++FT +L YLY+MRFIAGIGVGGEYG+
Sbjct: 61  LVGGLLFGLLADRHHKFKVFKWTILLFSVATGLIYFTQSLPYLYLMRFIAGIGVGGEYGV 120

Query: 121 AIAIMAGIVPTNKMGRISSLNGIAGQVGSISSALLAGWLAPALGWRGLFLFGLLPIVLVL 180
           AIAIMAGIVP  KMGR+SSLNGIAGQ+GSISSALLAGWLAP+LGWRGLFLFGLLPI+LV+
Sbjct: 121 AIAIMAGIVPPEKMGRMSSLNGIAGQLGSISSALLAGWLAPSLGWRGLFLFGLLPILLVI 180

Query: 181 WMQFAVDDKDILDQYNTDADDEPLDISIKALFDTPVLATQSLALMVMTTVQIAGYFGMMN 240
           WM  A+DD+ I D Y  + ++    I I  LF T  L  Q+LALMVMTTVQIAGYFGMMN
Sbjct: 181 WMTLAIDDQKIWDHYGQEEEECSQPIKINELFKTKSLTAQTLALMVMTTVQIAGYFGMMN 240

Query: 241 WLPTIIQTNLNVSVKNSSLWMIATILGMCLGMLVFGQLLDKFGPRLVYGCFLLSSAICVY 300
           WLPTIIQT+LN+SVK+SSLWM+ATI+GMCLGML FGQLLD FGPRL+Y  FLL+S+ICVY
Sbjct: 241 WLPTIIQTSLNLSVKSSSLWMVATIVGMCLGMLYFGQLLDCFGPRLIYSLFLLASSICVY 300

Query: 301 LFQFATTMPSMIIGGAVVGFFVNGMFAGYGAMITRLYPHHIRSTANNLILNVGRAIGGFS 360
           LFQFA +M SM+IGGA+VGFFVNGMFAGYGAMITRLYPHHIRSTANN+ILNVGRA+GGFS
Sbjct: 301 LFQFANSMASMVIGGAIVGFFVNGMFAGYGAMITRLYPHHIRSTANNVILNVGRALGGFS 360

Query: 361 SVIIGMILDVSNVSMVMLFLASLYIVSFLSMLSIKQLKRQKY 402
           SV IG ILD S +SMVM+FLASLY++SF +M SI QLK ++Y
Sbjct: 361 SVAIGSILDASGISMVMIFLASLYVISFGAMWSIGQLKAERY 402
```

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

Example 2033

A DNA sequence (GBSx2144) was identified in *S.agalactiae* <SEQ ID 6285> which encodes the amino acid sequence <SEQ ID 6286>. This protein is predicted to be leucyl-tRNA synthetase (leuS). Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3481(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 10241> which encodes amino acid sequence <SEQ ID 10242> was also identified.

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

```
>GP:AAC00259 GB:AF008220 leucine tRNA synthetase [Bacillus subtilis]
 Identities = 569/835 (68%), Positives = 666/835 (79%), Gaps = 42/835 (5%)

Query: 10  YNHKEIEPKWQAFWADNHTFKTGTDASKPKFYALDMFPYPSGAGLHVGHPEGYTATDILS 69
           + HKEIE KWQ +W +N TF T  +  K KFYALDMFPYPSGAGLHVGHPEGYTATDILS
Sbjct: 3   FQHKEIEKKWQTYWLENKTFATLDNNEKQKFYALDMFPYPSGAGLHVGHPEGYTATDILS 62

Query: 70  RFKRAQGHNVLHPMGWDAFGLPAEQYAMDTGNDPAEFTAENIANFKRQINALGFSYDWDR 129
           R KR QG++VLHPMGWDAFGLPAEQYA+DTGNDPA FT +NI NF+RQI ALGFSYDWDR
Sbjct: 63  RMKRMQGYDVLHPMGWDAFGLPAEQYALDTGNDPAVFTKQNIDNFRRQIQALGFSYDWDR 122

Query: 130 EVNTTDPNYYKWTQWIFTKLYEKGLAYEAEVPVNWVEELGTAIANEEVLPDGTSERGGYP 189
           E+NTTDP YYKWTQWIF KLYEKGLAY  EVPVNW   LGT +ANEEV+ DG SERGG+P
Sbjct: 123 EINTTDPEYYKWTQWIFLKLYEKGLAYVDEVPVNWCPALGTVLANEEVI-DGKSERGGHP 181

Query: 190 VVRKPMRQWMLKITAYAERLLEDLEEVDWPESIKDMQRNWIGKSTGANVTFKVKDTDKDF 249
           V R+PM+QWMLKITAYA+RLLEDLEE+DWPESIKDMQRNWIG+S GA+V F +   D  F
Sbjct: 182 VERRPMKQWMLKITAYADRLLEDLEELDWPESIKDMQRNWIGRSEGAHVHFAIDGHDDSF 241

Query: 250 TVFTTRPDTLFGATYAVLAPEHALVDAITTADQAEAVAEYKRQASLKSDLARTDLAKEKT 309
           TVFTTRPDTLFGATY VLAPEHALV+ ITTA+Q EAV  Y ++   KSDL RTDLAK KT
Sbjct: 242 TVFTTRPDTLFGATYTVLAPEHALVENITTAEQKEAVEAYIKEIQSKSDLERTDLAKTKT 301

Query: 310 GVWTGAYAINPVNGKEIPVWIADYVLASYGTGAIMAVPAHDERDWEFAKQFNLDIIPVLE 369
           GV+TGAYAINPVNG+++P+WIADYVLASYGTGA+MAVP HDERD+EFAK F L +  V++
Sbjct: 302 GVFTGAYAINPVNGEKLPIWIADYVLASYGTGAVMAVPGHDERDFEFAKTFGLPVKEVVK 361

Query: 370 GGNVEEAAFTEDGLHINSDFLDGLDKAAAIAKMVEWLEAEGVGNEKVTYRLRDWLFSRQR 429
           GGNVEEAA+T DG H+NSDFL+GL K  AI K++ WLE     G +KVTYRLRDWLFSRQR
Sbjct: 362 GGNVEEAAYTGDGEHVNSDFLNGLHKQEAIEKVIAWLEETKNGEKKVTYRLRDWLFSRQR 421

Query: 430 YWGEPIPIIHWEDGTSTAVPESELPLVLPVTKDIRPSGTGESPLANLTDWLEVT-REDGV 488
           YWGEPIP+IHWEDGTSTAVPE ELPL+LP T +I+PSGTGESPLAN+ +W+EVT  E G
Sbjct: 422 YWGEPIPVIHWEDGTSTAVPEEELPLILPKTDEIKPSGTGESPLANIKEWVEVTDPETGK 481

Query: 489 KGRRETNTMPQWAGSSWYYLRYIDPHNTEKLADEELLKQWLPVDIYVGGAEHAVLHLLYA 548
           KGRRETNTMPQWAGS WY+LRYIDPHN ++LA  E L++WLPVD+Y+GGAEHAVLHLLYA
Sbjct: 482 KGRRETNTMPQWAGSCWYFLRYIDPHNPDQLASPEKLEKWLPVDMYIGGAEHAVLHLLYA 541

Query: 549 RFWHKVLYDLGVVPTKEPFQKLFNQGMILGTSYRDSRGALVATDKVEKRDGSFFHVETGE 608
           RFWHK LYD+GVVPTKEPFQKL+NQGMILG                          E  E
Sbjct: 542 RFWHKFLYDIGVVPTKEPFQKLYNQGMILG--------------------------ENNE 575

Query: 609 ELEQAPAKMSKSLKNVVNPDDVVEQYGADTLRVYEMFMGPLDASIAWSEEGLEGSRKFLD 668
                 KMSKS  NVVNPD++V  +GADTLR+YEMFMGPLDASIAWSE GL+G+R+FLD
Sbjct: 576 -------KMSKSKGNVVNPDEIVASHGADTLRLYEMFMGPLDASIAWSESGLDGARRFLD 628

Query: 669 RVYRLI------TTKEITEENSGALDKVYNETVKAVTEQVDQMKFNTAIAQLMVFVNAAN 722
           RV+RL           +I E      L++VY+ETV  VT+  + ++FNT I+QLMVF+N A
Sbjct: 629 RVWRLFIEDSGELNGKIVEGAGETLERVYHETVMKVTDHYEGLRFNTGISQLMVFINEAY 688

Query: 723 KEDKLFSDYAKGFVQLIAPFAPHLGEELWQVLTASGQSISYVPWPSYDESKLVENEIEIV 782
           K  +L  +Y +GFV+L++P APHL EELW+ L  SG +I+Y  WP YDE+KLV++E+EIV
Sbjct: 689 KATELPKEYMEGFVKLLSPVAPHLAEELWEKLGHSG-TIAYEAWPVYDETKLVDDEVEIV 747

Query: 783 VQIKGKVKAKLVVAKDLSREELQDLALANEKVQAEIAGKDIIKVIAVPNKLVNIV 837
           VQ+ GKVKAKL V  D ++E+L+ LA A+EKV+ ++ GK I K+IAVP KLVNIV
```

```
Sbjct: 748 VQLNGKVKAKLQVPADATKEQLEQLAQADEKVKEQLEGKTIRKIIAVPGKLVNIV 802
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6287> which encodes the amino acid sequence <SEQ ID 6288>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4358(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 = 815/833 (97%), Positives = 827/833 (98%)

Query: 7   MTFYNHKEIEPKWQAFWADNHTFKTGTDASKPKFYALDMFPYPSGAGLHVGHPEGYTATD 66
           MTFY+H  IEPKWQAFWADNHTFKTGTDASKPKFYALDMFPYPSGAGLHVGHPEGYTATD
Sbjct: 1   MTFYDHTAIEPKWQAFWADNHTFKTGTDASKPKFYALDMFPYPSGAGLHVGHPEGYTATD 60

Query: 67  ILSRFKRAQGHNVLHPMGWDAFGLPAEQYAMDTGNDPAEFTAENIANFKRQINALGFSYD 126
           ILSRFKRAQGHN+LHPMGWDAFGLPAEQYAMDTGNDPAEFTAENIANFKRQINALGFSYD
Sbjct: 61  ILSRFKRAQGHNILHPMGWDAFGLPAEQYAMDTGNDPAEFTAENIANFKRQINALGFSYD 120

Query: 127 WDREVNTTDPNYYKWTQWIFTKLYEKGLAYEAEVPVNWVEELGTAIANEEVLPDGTSERG 186
           WDREVNTTDPNYYKWTQWIFTKLYEKGLAYEAEVPVNWVEELGTAIANEEVLPDGTSERG
Sbjct: 121 WDREVNTTDPNYYKWTQWIFTKLYEKGLAYEAEVPVNWVEELGTAIANEEVLPDGTSERG 180

Query: 187 GYPVVRKPMRQWMLKITAYAERLLEDLEEVDWPESIKDMQRNWIGKSTGANVTFKVKDTD 246
           GYPVVRKPMRQWMLKITAYAERLLEDLEEVDWPESIKDMQRNWIGKSTGANVTFKVKDTD
Sbjct: 181 GYPVVRKPMRQWMLKITAYAERLLEDLEEVDWPESIKDMQRNWIGKSTGANVTFKVKDTD 240

Query: 247 KDFTVFTTRPDTLFGATYAVLAPEHALVDAITTADQAEAVAEYKRQASLKSDLARTDLAK 306
           KDFTVFTTRPDTLFGATYAVLAPEHALVDAITTADQAEAVA+YKRQASLKSDLARTDLAK
Sbjct: 241 KDFTVFTTRPDTLFGATYAVLAPEHALVDAITTADQAEAVAKYKRQASLKSDLARTDLAK 300

Query: 307 EKTGVWTGAYAINPVNGKEIPVWIADYVLASYGTGAIMAVPAHDERDWEFAKQFNLDIIP 366
           EKTGVWTGAYAINPVNG E+PVWIADYVLASYGTGAIMAVPAHDERDWEFAKQF LDIIP
Sbjct: 301 EKTGVWTGAYAINPVNGNEMPVWIADYVLASYGTGAIMAVPAHDERDWEFAKQFKLDIIP 360

Query: 367 VLEGGNVEEAAFTEDGLHINSDFLDGLDKAAAIAKMVEWLEAEGVGNEKVTYRLRDWLFS 426
           VLEGGNVEEAAFTEDGLHINS FLDGLDKA+AIAKMVEWLEAEGVGNEKVTYRLRDWLFS
Sbjct: 361 VLEGGNVEEAAFTEDGLHINSGFLDGLDKASAIAKMVEWLEAEGVGNEKVTYRLRDWLFS 420

Query: 427 RQRYWGEPIPIIHWEDGTSTAVPESELPLVLPVTKDIRPSGTGESPLANLTDWLEVTRED 486
           RQRYWGEPIPIIHWEDGTSTAVPESELPLVLPVTKDIRPSGTGESPLAN+TDWLEVTRED
Sbjct: 421 RQRYWGEPIPIIHWEDGTSTAVPESELPLVLPVTKDIRPSGTGESPLANVTDWLEVTRED 480

Query: 487 GVKGRRETNTMPQWAGSSWYYLRYIDPHNTEKLADEELLKQWLPVDIYVGGAEHAVLHLL 546
           GVKGRRETNTMPQWAGSSWYYLRYIDPHNTEKLADEELLKQWLPVDIYVGGAEHAVLHLL
Sbjct: 481 GVKGRRETNTMPQWAGSSWYYLRYIDPHNTEKLADEELLKQWLPVDIYVGGAEHAVLHLL 540

Query: 547 YARFWHKVLYDLGVVPTKEPFQKLFNQGMILGTSYRDSRGALVATDKVEKRDGSFFHVET 606
           YARFWHKVLYDLGVVPTKEPFQKLFNQGMILGTSYRDSRGALVATDKVEKRDGSFFHVET
Sbjct: 541 YARFWHKVLYDLGVVPTKEPFQKLFNQGMILGTSYRDSRGALVATDKVEKRDGSFFHVET 600

Query: 607 GEELEQAPAKMSKSLKNVVNPDDVVEQYGADTLRVYEMFMGPLDASIAWSEEGLEGSRKF 666
           GEELEQAPAKMSKSLKNVVNPDDVVEQYGADTLRVYEMFMGPLDASIAWSEEGLEGSRKF
Sbjct: 601 GEELEQAPAKMSKSLKNVVNPDDVVEQYGADTLRVYEMFMGPLDASIAWSEEGLEGSRKF 660

Query: 667 LDRVYRLITTKEITEENSGALDKVYNETVKAVTEQVDQMKFNTAIAQLMVFVNAANKEDK 726
           LDRVYRLITTKEITEENSGALDKVYNETVKAVTEQVDQMKFNTAIAQLMVFVNAANKEDK
Sbjct: 661 LDRVYRLITTKEITEENSGALDKVYNETVKAVTEQVDQMKFNTAIAQLMVFVNAANKEDK 720

Query: 727 LFSDYAKGFVQLIAPFAPHLGEELWQVLTASGQSISYVPWPSYDESKLVENEIEIVVQIK 786
           LFSDYAKGFVQLIAPFAPHLGEELWQ LTASG+SISYVPWPSYDESKLVEN++EIVVQIK
```

```
Sbjct: 721 LFSDYAKGFVQLIAPFAPHLGEELWQALTASGESISYVPWPSYDESKLVENDVEIVVQIK 780

Query: 787 GKVKAKLVVAKDLSREELQDLALANEKVQAEIAGKDIIKVIAVPNKLVNIVVK 839
           GKVKAKLVVAKDLSREELQ++ALANEKVQAEIAGKDIIKVIAVPNKLVNIV+K
Sbjct: 781 GKVKAKLVVAKDLSREELQEVALANEKVQAEIAGKDIIKVIAVPNKLVNIVIK 833
```

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

Example 2034

A DNA sequence (GBSx2145) was identified in *S.agalactiae* <SEQ ID 6289> which encodes the amino acid sequence <SEQ ID 6290>. This protein is predicted to be KIAA1074 protein. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have an uncleavable N-term signal seq

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

A related GBS nucleic acid sequence <SEQ ID 8961> which encodes amino acid sequence <SEQ ID 8962> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
SRCFLG: 0
McG: Length of UR:   19
     Peak Value of UR:   2.86
     Net Charge of CR: 4
McG: Discrim Score:     10.27
GvH: Signal Score (-7.5): -3.61
     Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 0 value:   2.12 threshold:  0.0
   PERIPHERAL  Likelihood =  2.12       7
 modified ALOM score:  -0.92

*** 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 no significant homology with any sequences in the GENPEPT database.

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

SEQ ID 8962 (GBS117) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 8; MW 22.5kDa).

GBS117-His was purified as shown in Figure 200, lane 7.

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

Example 2035

A DNA sequence (GBSx2146) was identified in *S.agalactiae* <SEQ ID 6291> which encodes the amino acid sequence <SEQ ID 6292>. This protein is predicted to be YirC (resE). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL     Likelihood =-10.88   Transmembrane  177 - 193 ( 173 - 196)
   INTEGRAL     Likelihood = -4.09   Transmembrane   10 -  26 (   5 -  29)

----- Final Results -----
               bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB15292 GB:Z99120 similar to two-component sensor histidine
          kinase [YvqA] [Bacillus subtilis]
 Identities = 108/379 (28%), Positives = 193/379 (50%), Gaps = 33/379 (8%)

Query: 92  DNHKKESHDIIRYLTQKRLWQISKEKDGMFVTIKKKTYYVMTKDYSGILVDGSIKKVPKA 151
           +N +  S   +  L+   + ++ K  D      KKK Y   + D +G  V   IKK
Sbjct: 86  ENEEASSDKDLSILSSSFIHKVYKLADKQ--EAKKKRY---SADVNGEKVFFVIKKGLSV 140

Query: 152 QSQLFHVINFS------DITYTQHLITKINHFLIVILVLTYIPMLFIMRKTFTGIRESIQ 205
             Q   +++++      D+ YT  L  ++   + V+++L++IP +++ +     +   +
Sbjct: 141 NGQSAMMLSYALDSYRDDLAYT--LFKQLLFIIAVVILLSWIPAIWLAKY----LSRPLV 194

Query: 206 SVQTYISSLWKNQGNHQSSQKEIVFSDFDPLLLESQEMANRIYQAEESQRNFFQNASHEL 265
           S + ++  +  ++ +     K     +   L    +EM  ++ Q +E++R   QN SH+L
Sbjct: 195 SFEKHVKRI--SEQDWDDPVKVDRKDEIGKLGHTIEEMRQKLVQKDETERTLLQNISHDL 252

Query: 266 RTPLMSIQGYTEGVQEGII---DAELAHSVILQESKKMKQLVDDIILLSKLD--SNLSDQ 320
           +TP+M I+GYT+ +++GI    D E    VI  E+ K+++ + D++ L+KLD  +    Q
Sbjct: 253 KTPVMVIRGYTQSIKDGIFPKGDLENTVDVIECEALKLEKKIKDLLYLTKLDYLAKQKVQ 312

Query: 321 KDEFSLNELLNSIIAYFKPLANKQKISITYRPDKHEKLLK-GNEELIQRAINNILSNALR 379
            D FS+ E+   +I   K  A K+   +++  D  E +L  G+ E   + + NIL N +R
Sbjct: 313 HDMFSIVEVTEEVIERLK-WARKE---LSWEIDVEEDILMPGDPEQWNKLLENILENQIR 368

Query: 380 YAVSHIEISYT----NQKLTISNDGPAISKEDLPYIFDRFYKGHGGQTGIGLAMTKEIIK 435
           YA + IEIS      N  +TI NDGP I  E L  +++ F KG  G+ GIGL++ K I+
Sbjct: 369 YAETKIEISMKQDDRNIVITIKNDGPHIEDEMLSSLYEPFNKGKKGEFGIGLSIVKRILT 428

Query: 436 QHHGNIIAESDSTSTTFTI 454
            H  +I  E+D T  ++ I
Sbjct: 429 LHKASISIENDKTGVSYRI 447
```

There is also homology to SEQ ID 1178.

SEQ ID 6292 (GBS279) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 7; MW 54.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 6; MW 79.4kDa).

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

Example 2036

A DNA sequence (GBSx2147) was identified in *S.agalactiae* <SEQ ID 6293> which encodes the amino acid sequence <SEQ ID 6294>. This protein is predicted to be two-component response regulator (mtrA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1706(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 10239> which encodes amino acid sequence <SEQ ID 10240> was also identified.

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

```
>GP:BAB05663 GB:AP001513 two-component response regulator [Bacillus halodurans]
 Identities = 87/220 (39%), Positives = 124/220 (55%), Gaps = 4/220 (1%)

Query: 11  IYFADDEKNIRDLVVPFLEHDGFTVRAFETGDLLLEAYKNQKPDLVILDIMMPGTNGLDV 70
           I   DDE ++R+LV  +L  +GF V   ETGD  ++  + +  DLV+LD+MM   +G
Sbjct: 7   ILIVDDELDLRELVTSYLRKEGFAVYTAETGDEAIKRLEQEPMDLVVLDVMMDEMDGFTA 66

Query: 71  MKSIRQYDNIPIIMLTARDSDVDFITAFNLGTDDYFTKPFSPIKLSLHVKALFKRLDEKA 130
            K IR +  IPIIMLTAR  + D +    +G DDY  KPFSP +L   ++   +R
Sbjct: 67  CKEIRAFSQIPIIMLTARGGEDDKVMGLQIGADDYIVKPFSPRELVARIEVALRRTQGIQ 126

Query: 131 IKNDTQYQFLDLTLDTEKRIALLSNEEMPLTKTEFDFLLVLIEKPETAFSRETLLNRIWG 190
             +DT Y+F +L +    R   ++ +E+ LTK E+D L+ L+E     F+RE L +R+WG
Sbjct: 127 QVDDTGYRFNELRIQPSGRKVFVNGQEISLTKKEYDLLVFLLEHRGRVFTREHLHDRLWG 186

Query: 191 FDDIES--RAVDDTIKRLRKKFKQYHSQVSIKTVWGYGFK 228
            D  +   R VD  IK LR K K   +   IKTVWG G+K
Sbjct: 187 MDTQQGTLRTVDTHIKTLRLKLKP--ADRFIKTVWGVGYK 224
```

There is also homology to SEQ ID 3260.

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

Example 2037

A DNA sequence (GBSx2148) was identified in *S.agalactiae* <SEQ ID 6295> which encodes the amino acid sequence <SEQ ID 6296>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -2.18   Transmembrane 1568 -1584 (1568 -1585)
   INTEGRAL    Likelihood = -0.16   Transmembrane  338 - 354 ( 338 - 354)

----- Final Results -----
             bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10237> which encodes amino acid sequence <SEQ ID 10238> was also identified.

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

>GP:AAG09771 GB:AF243528 cell envelope proteinase [Streptococcus thermophilus]
Identities = 797/1594 (50%), Positives = 1056/1594 (66%), Gaps = 39/1594 (2%)

```
Query: 21   MNTKQRFSIRKYKLGAVSVLLGTLFFLGGITNVAADSVINKPSDIAVEQQVKDSPTS-IA 79
            M  K+ FS+RKYK+G VSVLLG +F   G  +VAAD + +   + VE  V D+  S  A
Sbjct: 1    MKKKETFSLRKYKIGTVSVLLGAVFLFAGAPSVAADELTSLV-ETKVEATVPDAIVSESA 59

Query: 80   NETPTNN--TSSALASTAQDNLVTKANNSPTETQPVAESHSQATETFSPVANQPVESTQE 137
            +E+P       +++ +T+ D   T       ++ +  S +   ET  P   P  S ++
Sbjct: 60   SESPVVEELVDTSVEATSTDVTTTDNEEETPGSEALENSANTEVETTQPAVETPAISEKK 119

Query: 138  VSKTPLTKQNLAVKSTPAISKET--PQNIDSNKIITVPKVWNTGYKGEGTVVAIIDSGLD 195
            V +    K ++A ++T   ++E   PQNIDSN IITVPKVW +GYKGEGTVVAIIDSGLD
Sbjct: 120  VEEEE--KLSVADETTAITNQEEAKPQNIDSNTIITVPKVWYSGYKGEGTVVAIIDSGLD 177

Query: 196  INHDALQLNDSTKAKYQNEQQMNAAKAKAGINYGKWYNNKVIFGHNYVDVNTELKEVKST 255
            ++HD L ++D + AKY++E+++ AAK  AGI YG+W+N+KV+FG+NYVDVNT LKE
Sbjct: 178  VDHDVLHISDLSTAKYKSEKEIEAAKEAAGITYGEWFNDKVVFGYNYVDVNTVLKEEDKR 237

Query: 256  SHGMHVTSIATANPSKKDTNELIYGVAPEAQVMFMRVFSDEKRGTGPALYVKAIEDAVKL 315
            SHGMHVTSIAT NP++    +L+YGVAPEAQVMFMRVFSD K  TG ALYVKAIEDAVKL
Sbjct: 238  SHGMHVTSIATGNPTQPVAGQLMYGVAPEAQVMFMRVFSDLKATTGAALYVKAIEDAVKL 297

Query: 316  GADSINLSLGGANGSLVNADDRLIKALEMARLAGVSVVIAAGNDGTFGSGASKPSALYPD 375
            GADSINLSLGGANGS+VN ++ +  A+E AR AGVSVVIAAGNDGTFGSG S PSA YPD
Sbjct: 298  GADSINLSLGGANGSVVNMNENVTAAIEAARRAGVSVVIAAGNDGTFGSGHSNPSADYPD 357

Query: 376  YGLVGSPSTAREAISVASYNNTTLVNKVFNIIGLENNRNLNNGLAAYADPKVSDKTFEVG 435
            YGLVG+PSTA +AISVASYNNTT+ +KV NIIGLENN +LN G +++ +P+ S   FE+G
Sbjct: 358  YGLVGAPSTAHDAISVASYNNTTVGSKVINIIGLENNADLNYGKSSFDNPEKSPVPFEIG 417

Query: 436  KQYDYVFVGKGNDNDYKDKTLNGKIALIERGDITFTKKVVNAINHGAVGAIIFNNKAGEA 495
            K+Y+YV+ G G  +D+   L GK+ALI+RG ITF++K+ NA   GAVG +IFN++ GEA
Sbjct: 418  KEYEYVYAGIGQASDFDGLDLTGKLALIKRGTITFSEKIANATAAGAVGVVIFNSRPGEA 477

Query: 496  NLTMSLDPEASAIPAIFTQKEFGDVLAKNNYKIVFNNIKNKQANPNAGVLSDFSSWGLTA 555
            N++M LD  A AIP++F   EFG+ LA N+YKI FNN  + + NP AG+LSDFSSWGL+A
Sbjct: 478  NVSMQLDDTAIAIPSVFIPLEFGEALAANSYKIAFNNETDIRPNPEAGLLSDFSSWGLSA 537

Query: 556  DGQLKPDLSAPGGSIYAAINDNEYDMMSGTSMASPHVAGATALVKQYLLKEHPELKKGDI 615
            DG+LKPDL+APGG+IYAAINDN+Y  M GTSMASPHVAGA  LVKQYLL  +P     +I
Sbjct: 538  DGELKPDLAAPGGAIYAAINDNDYANMQGTSMASPHVAGAAVLVKQYLLATYPTKSPQEI 597

Query: 616  ERTVKYLLMSTAKAHLNKDTGAYTSPRQQGAGIIDVAAAVQTGLYLTGGENNYGSVTLGN 675
            E  VK+LLMSTAKAH+NK+T AYTSPRQQGAGIID AAA+ TGLYLT GE+ YGS+TLGN
Sbjct: 598  EALVKHLLMSTAKAHVNKETTAYTSPRQQGAGIIDTAAAISTGLYLT-GEDGYGSITLGN 656

Query: 676  IKDKISFDVTVHNINKVAKDLHYTTYLNTDQVKDGFVTLAPQQLGTFTGKTIRIEPGQTQ 735
            ++D  SF VT+HNI     K L+Y+T L TD  +     L    +  + + ++  +
Sbjct: 657  VEDTFSFTVTLHNITNEDKTLNYSTQLTTDTAQKRIDHLGSTSISRDSWRKVTVKANSST 716

Query: 736  TITIDIDVSKYHDMLKKVMPNGYFLEGYVRFTDPVDGGEVLSIPYVGFKGEFQNLEVLEK 795
            T+TI++D S + + L  +M NGY+LEG+VRFTD  D G+++SIPYVGF+GEFQNL VLE+
Sbjct: 717  TVTINVDASSFAEELTGLMKNGYYLEGFVRFTDVADDGDIVSIPYVGFRGEFQNLAVLEE 776

Query: 796  SIYKLVANKEKGFYFQP--KQTNEVPGSEDYTALMTTSSEPIYSTDGTSPIQLKALGSYK 853
             IY L+A+ + GFYF+P    Q N V  S  YT L+T S+E IYSTD  S   +K LG++K
Sbjct: 777  PIYNLIADGKGGFYFEPVTAQPNTVDISHHYTGLVTGSTELIYSTDKRSDSAIKTLGTFK 836

Query: 854  SIDGKWILQLDQKGQPHLAISPNDDQNQDAVAVKGVFLRNFNNLRAKVYRADDVNLQKPL 913
            +  G ++L+LD+ G+PHLAISPN D NQD++   KGVFLRN+ +L A VY ADD      PL
Sbjct: 837  NKAGYFVLELDESGKPHLAISPNGDDNQDSLVFKGVFLRNYTDLVASVYAADDTERTNPL 896

Query: 914  WVSAPQAGDKNYYSGNTENPKSTFLYDTEWKGTTTDGIPLEDGKYKYVLTYYSDVPGSKP 973
            W S PQ+GDKN YSGN +NPKS+ +Y TEW GT +DG  L DGKY+YVLTY S VPG+
Sbjct: 897  WESQPQSGDKNIYSGNPKNPKSSIIYPTEWNGTDSDGNALADGKYQYVLTYSSKVPGAAV 956

Query: 974  QQMVFDITLDRQAPTLTTATYDKDRRIFKARPAVEHGESGIFREQVFYLKKDKDGHYNSV 1033
```

```
            Q M+FD+ +DR++P +TTATYD+    F  RPA+E GESG++REQVFYL  D  G   ++
Sbjct: 957  QTMIFDVIIDRESPVITTATYDETNFTFNPRPAIEKGESGLYREQVFYLVADASG-VTTI 1015

Query: 1034 LRQQGEDGILVEDNKVFIKQEKDGSFILPKEVNDFSHVYYTVEDYAGNLVSAKLEDLINI 1093
                 + V DNKVF+ Q  DGSF LP ++ D S  YYTVEDYAGN+   K+E+LI+I
Sbjct: 1016 PSLLKNGDVTVSDNKVFVAQNDDGSFTLPLDLADISKFYYTVEDYAGNISYEKVENLISI 1075

Query: 1094 GNKNGLVNVKVFSPELNSNVDIDFSYSVKDDKGNIIKK-QHHGKDLNLLKLPFGTYTFDL 1152
            GN+ GLV V +   + NS V I FSYSV D+ G I+ +   + D ++LKLPFGTYTFDL
Sbjct: 1076 GNEKGLVTVNILDKDTNSPVPILFSYSVTDETGKIVAELPRYAGDTSVLKLPFGTYTFDL 1135

Query: 1153 FLYDEERANLISPKSVTVTISEKDSLKDVLFKVNLLKKAALLVEFDKLLPKGATVQLVTK 1212
            FLYD E ++L       VTI E +S  +V F V L  KA LL++ D LLP G+T+QLVT
Sbjct: 1136 FLYDTEWSSLAGETKAVVTILEDNSTAEVNFYVTLKDKANLLIDIDALLPSGSTIQLVTA 1195

Query: 1213 TNTVVDLPKATYSPTDYGKNIPVGDYRLNVTLPSGYSTLENLDDLLVSVKEDQVNLTKLT 1272
                + LP A YS TDYGK +PVG Y +  TLP GY  LE LD   V+V  +Q N+ KLT
Sbjct: 1196 DGQAIQLPNAKYSKTDYGKFVPVGTYTILPTLPEGYEFLEELD---VAVLANQSNVKKLT 1252

Query: 1273 LINKAPLINALAEQTDIITQPVFYNAGTHLKNNYLANLEKAQTLIKNRVEQTSIDNAIAA 1332
            LINK  L   +AE   +     +YNA   L+  Y   LE A  + N+  Q  +D+A+A+A
Sbjct: 1253 LINKVALKELIAELAGLEETARYYNASPELQTAYAKALEDANAVYANKHNQAQVDSALAS 1312

Query: 1333 LRESRQALNGKETDTSLLAKAILAETEIKGNYQFVNASPLSQSTYINQVQLAKNLLQKPN 1392
            L  +R+ LNG+ TD    L   +   T  + N+ + NA    Q  Y   V+ A+ +L + N
Sbjct: 1313 LVAAREQLNGQATDKEKLIAEVSNYTPTQANFIYYNAENTKQIAYDTAVRSAQLVLNQEN 1372

Query: 1393 VTQSEVDKALENLDIAKNQLNGHETDYSGLHHMIIKANVLKQTSSKYQNASQFAKENYNN 1452
            VTQ+ V++AL +L  AK  L+G +TD S L   +  ++VLK T +KY NAS+  K+ Y+
Sbjct: 1373 VTQAVVNQALADLLAAKANLDGQKTDISALRSAVSVSSVLKATDAKYLNASENVKQAYDQ 1432

Query: 1453 LIKKAELLLSNRQATQAQVEELLNQIKATEQELDG----RDRVSSAENYSQSLNDNDSLN 1508
             ++ A+ +L +  A+QA V++ L  + + + ELDG     +      N +    D  ++
Sbjct: 1433 AVEAAKAILVDESASQASVDQALAVLTSAQAELDGVATSTNDAKEPANTATDKKDEGTVT 1492

Query: 1509 TTPIN--------PP-----NQPQALIFKKGMTKESEVAQKRVLGVTSQTDNQKVKTNKL 1555
              PI+        PP     N     I +K    +  + +   L  +  + NQ+ +  +L
Sbjct: 1493 PPPIDSEIVDVQAPPVKDTGNSEHVPIGQK-PNPQPTLPRPVTLQASLSSPNQEKQVTQL 1551

Query: 1556 PKTGESTPKITYTILLFSLSMLGLATIKLKSIKR 1589
            P TGE+  K      L    ++GL T+ L SI+R
Sbjct: 1552 PNTGENDTK----YYLVPGVIIGLGTL-LVSIRR 1580
```

A related GBS gene <SEQ ID 8963> and protein <SEQ ID 8964> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
SRCFLG: 0
McG: Length of UR:   1
     Peak Value of UR:   2.55
     Net Charge of CR: 4
McG: Discrim Score:      2.60
GvH: Signal Score (-7.5): -0.78
     Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 36
ALOM program   count: 1 value:  -0.16 threshold:  0.0
   INTEGRAL    Likelihood = -0.16   Transmembrane  318 - 334 ( 318 - 334)
   PERIPHERAL  Likelihood =  2.54    1161
 modified ALOM score:   0.53
icm1 HYPID: 7  CFP: 0.106

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

LPXTG motif: 1535-1539

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

50.5/67.5% over 1583aa

Streptococcus thermophilus
GP|9963932| cell envelope proteinase Insert characterized

ORF01603(361 - 5070 of 5370)
GP|9963932|gb|AAG09771.1|AF243528_1|AF243528(1 - 1584 of 1585) cell envelope proteinase {Streptococcus thermophilus}
%Match = 41.2
%Identity = 50.4 %Similarity = 67.4
Matches = 794 Mismatches = 498 Conservative Sub.s = 267

```
255       285       315       345       375       405       435       465
KNALGTVLNLPQNNL**KFRKL*KILIFYVLIVFVIIMLQEKEIFMNTKQRFSIRKYKLGAVSVLLGTLFFLGGITNVAA
                                                 |  |: ||:|||||:| |||||| :|:: |  :|||
                                                MKKKETFSLRKYKIGTVSVLLGAVFLFAGAPSVAA
                                                          10        20        30

495       525       552       576       606       636       666       696
DSVINKPSDIAVEQQVKDSPTS-IANETPT--NNTSSALASTAQDNLVTKANNSPTETQPVAESHSQATETFSPVANQPV
|  :    :  ||  | |:  |  |:|:|       ::: :|: |   |        :: :  | :   ||  |    |
DE-LTSLVETKVEATVPDAIVSESASESPVVEELVDTSVEATSTDVTTTDNEEETPGSEALENSANTEVETTQPAVETPA
               50        60        70        80        90       100       110

726       756       780       810       840       870       900       930
ESTQEVSKTPLTKQNLAVKSTPAISKE--TPQNIDSNKIITVPKVWNTGYKGEGTVVAIIDSGLDINHDALQLNDSTKAK
 | ::| :    | ::| ::|   ::|   |||||||| |||||||| :|||||||||||||||||::|| |:::| : ||
ISEKKVEEEE--KLSVADETTAITNQEEAKPQNIDSNTIITVPKVWYSGYKGEGTVVAIIDSGLDVDHDVLHISDLSTAK
               130       140       150       160       170       180       190

960       990       1020      1050      1080      1110      1140      1170
YQNEQQMNAAKAKAGINYGKWYNNKVIFGHNYVDVNTELKEVKSTSHGMHVTSIATANPSKKDTNELIYGVAPEAQVMFM
|::|::: |||  ||| ||:|:|:||:||:|||||| |||    |||||||||| ||::    :|:||||||||||||
YKSEKEIEAAKEAAGITYGEWFNDKVVFGYNYVDVNTVLKEEDKRSHGMHVTSIATGNPTQPVAGQLMYGVAPEAQVMFM
               210       220       230       240       250       260       270

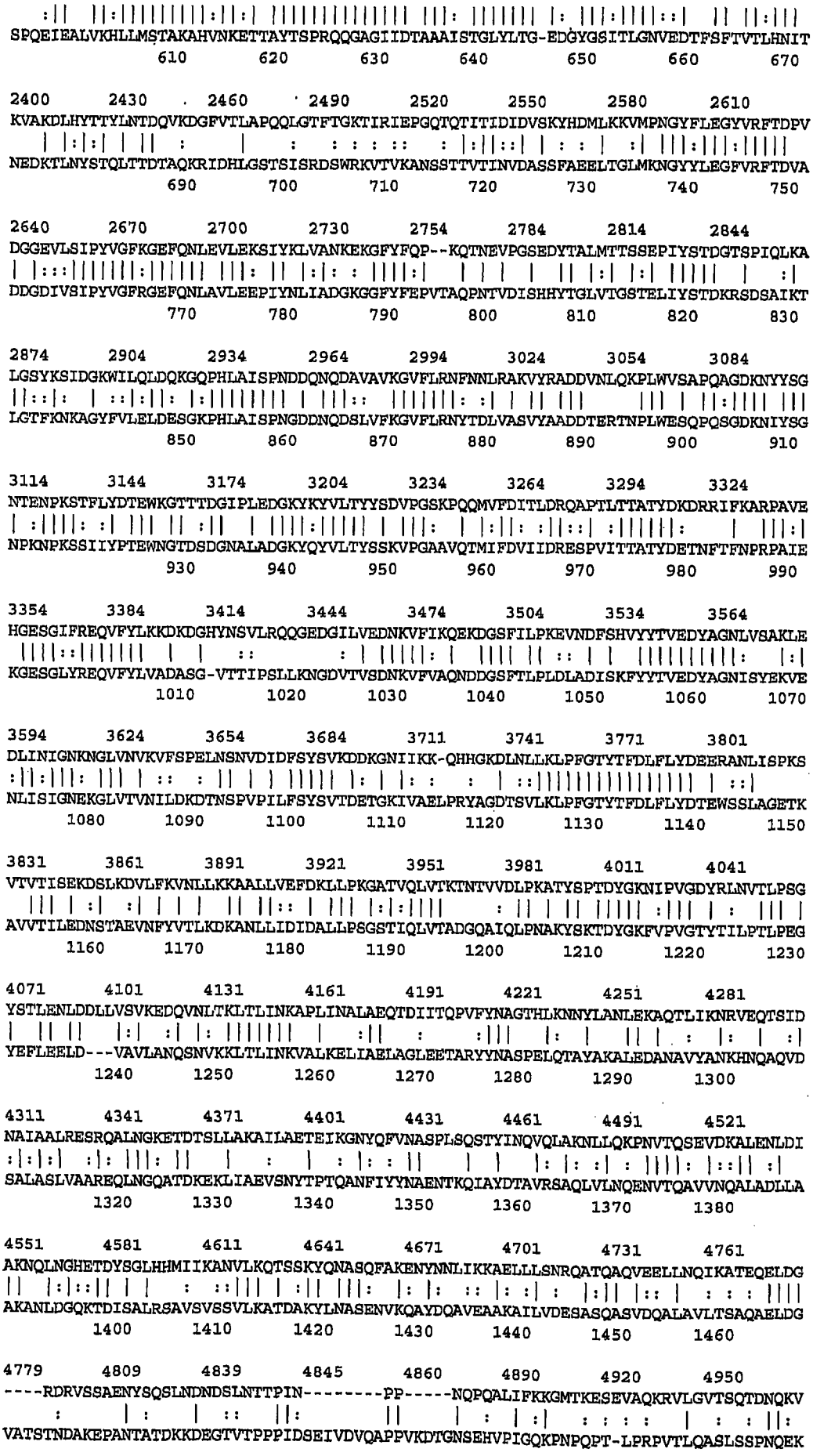
1200      1230      1260      1290      1320      1350      1380      1410
RVFSDEKRGTGPALYVKAIEDAVKLGADSINLSLGGANGSLVNADDRLIKALEMARLAGVSVVIAAGNDGTFGSGASKPS
|||||  ||  |||||||||||||||||||||||||||:|| :: :  |:| || |||||||||||||||| | ||
RVFSDLKATTGAALYVKAIEDAVKLGADSINLSLGGANGSVVNMNENVTAAIEAARRAGVSVVIAAGNDGTFGSGHSNPS
               290       300       310       320       330       340       350

1440      1470      1500      1530      1560      1590      1620      1650
ALYPDYGLVGSPSTAREAISVASYNNTTLVNKVFNIIGLENNRNLNNGLAAYADPKVSDKTFEVGKQYDYVFVGKGNDND
| ||||||||:|||| :|||||||||||: :|| |||||||| :|| | ::: :|: |   ||:||:|:||: | |  :|
ADYPDYGLVGAPSTAHDAISVASYNNTTVGSKVINIIGLENNADLNYGKSSFDNPEKSPVPFEIGKEYEYVYAGIGQASD
               370       380       390       400       410       420       430

1680      1710      1740      1770      1800      1830      1860      1890
YKDKTLNGKIALIERGDITFTKKVVNAINHGAVGAIIFNNKAGEANLTMSLDPEASAIPAIFTQKEFGDVLAKNNYKIVF
:   | ||:|||:|| |||::|: ||   |||| :|||:: ||||::| ||  | |||::|   |||: || |:||| |
FDGLDLTGKLALIKRGTITFSEKIANATAAGAVGVVIFNSRPGEANVSMQLDDTAIAIPSVFIPLEFGEALAANSYKIAF
               450       460       470       480       490       500       510

1920      1950      1980      2010      2040      2070      2100      2130
NNIKNKQANPNAGVLSDFSSWGLTADGQLKPDLSAPGGSIYAAINDNEYDMMSGTSMASPHVAGATALVKQYLLKEHPEL
||  : : || ||:||||||||:|||:||||||:||||:|||||||:|  | ||||||||||| |||||||  :|
NNETDIRPNPEAGLLSDFSSWGLSADGELKPDLAAPGGAIYAAINDNDYANMQGTSMASPHVAGAAVLVKQYLLATYPTK
               530       540       550       560       570       580       590

2160      2190      2220      2250      2280      2310      2340      2370
KKGDIERTVKYLLMSTAKAHLNKDTGAYTSPRQQGAGIIDVAAAVQTGLYLTGGENNYGSVTLGNIKDKISFDVTVHNIN
```



```
          1480      1490      1500      1510      1520      1530      1540

4980      5010      5040      5070      5100      5130      5160      5190
KTNKLPKTGESTPKITYTILLFSLSMLGLATIKLKSIKRE*NTLKNRARHQLLAINS**LVPF*GA*NDVPKDLFSAVSW
:  :|| |||:  |  | ::   :  ||   : ::  | |
QVTQLPNTGENDTK--YYLVPGVIIGLGTLLVSIRRHKEEV
          1560      1570      1580
```

A related GBS nucleic acid sequence <SEQ ID 10965> which encodes amino acid sequence <SEQ ID 10966> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6297> which encodes the amino acid sequence <SEQ ID 6298>. Analysis of this protein sequence reveals the following:

```
LPXTG motif: 1614-1619

     Possible site: 33

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -4.46   Transmembrane 1623 -1639 (1621 -1641)

----- 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:

```
>GP:AAG09771 GB:AF243528 cell envelope proteinase [Streptococcus thermophilus]
 Identities = 465/1125 (41%), Positives = 668/1125 (59%), Gaps = 61/1125 (5%)

Query: 1    VEKKQRFSLRKYKSGTFSVLIGSVFLVM-TTTVAADELSTMSEPTITNHAQQQAQHLTNT 59
             ++KK+ FSLRKYK GT SVL+G+VFL      +VAADEL+++ E  +             T
Sbjct: 1    MKKKETFSLRKYKIGTVSVLLGAVFLFAGAPSVAADELTSLVETKVEA-----------T 49

Query: 60   ELSSAESKSQDTSQITLKTNREKEQSQDLVSEPTTTELADTDAASMANTGSDATQKSASL 119
               +  S+S   S +         E+  D   E T+T++  TD       GS+A + SA
Sbjct: 50   VPDAIVSESASESPVV-------EELVDTSVEATSTDVTTTDNEE-ETPGSEALENSA-- 99

Query: 120  PPVNTDVHDWVKTKGAWDKGYKGQGKVVAVIDTGIDPAHQSMRISDVSTAKVKSKEDMLA 179
             NT+V     T+ A +     + KV             + + ++D +TA     +E
Sbjct: 100  ---NTEVET---TQPAVETPAISEKKV---------EEEEKLSVADETTAITNQEE---- 140

Query: 180  RQKAAGINYGSWINDKVVFAHNYVENSDNIKE-NQFEDFDEDWENFEFDAEAEPKAIKKH 238
               K   I+  + I    V+   Y       +   +  D D  D  +    + A+ K+ K+
Sbjct: 141  -AKPQNIDSNTIITVPKVWYSGYKGEGTVVAIIDSGLDVDHDVLHISDLSTAKYKSEKEI 199

Query: 239  KIYRPQSTQAPKETVIKTEETDGSHDIDWTQTDDDTKYESHGMHVTGIVAGNSKEAAATG 298
             +  +     E     G + +D       +    SHGMHVT I  GN  +  A G
Sbjct: 200  EAAKEAAGITYGEW-FNDKVVFGYNYVDVNTVLKEEDKRSHGMHVTSIATGNPTQPVA-G 257

Query: 299  ERFLGIAPEAQVMFMRVFANDIMGSAESLFIKAIEDAVALGADVINLSLGTANGAQLSGS 358
             +   G+APEAQVMFMRVF++     +  +L++KAIEDAV LGAD INLSLG ANG+ ++ +
Sbjct: 258  QLMYGVAPEAQVMFMRVFSDLKATTGAALYVKAIEDAVKLGADSINLSLGGANGSVVNMN 317

Query: 359  KPLMEAIEKAKKAGVSVVVAAGNERVYGSDHDDPLATNPDYGLVGSPSTGRTPTSVAAIN 418
             + +  AIE A++AGVSVV+AAGN+  +GS H +P A  PDYGLVG+PST     SVA+ N
Sbjct: 318  ENVTAAIEAARRAGVSVVIAAGNDGTFGSGHSNPSADYPDYGLVGAPSTAHDAISVASYN 377

Query: 419  SKWVIQRLMTVKELENRADLNHGKAIYSESVDFKDIKDSLGYDKSHQFAYVKESTDAGYN 478
             + V  +++ +  LEN ADLN+GK+ + ++ +    +   +G +  + +A + +++D  ++
Sbjct: 378  NTTVGSKVINIIGLENNADLNYGKSSF-DNPEKSPVPFEIGKEYEYVYAGIGQASD--FD 434

Query: 479  AQDVKGKIALIERDPNKTYDEMIALAKKHGALGVLIFNNKPGQSNRSMRLTANGMGIPSA 538
              D+ GK+ALI+R     T+ E IA A   GA+GV+IFN++PG++N SM+L     + IPS
Sbjct: 435  GLDLTGKLALIKRG-TITFSEKIANATAAGAVGVVIFNSRPGEANVSMQLDDTAIAIPSV 493
```

```
Query: 539  FISHEFGKAMSQLNGNGTGSLEFDSVVSKAPSQKGNEMNHFSNWGLTSDGYLKPDITAPG 598
            FI  EFG+A++       +  + F++      P+ +    ++ FS+WGL++DG LKPD+ APG
Sbjct: 494  FIPLEFGEALAA----NSYKIAFNNETDIRPNPEAGLLSDFSSWGLSADGELKPDLAAPG 549

Query: 599  GDIYSTYNDNHYGSQTGTSMASPQIAGASLLVKQYLEKTQPNLPKEKIADIVKNLLMSNA 658
            G IY+  NDN Y +  GTSMASP +AGA++LVKQYL  T P     ++I  +VK+LLMS A
Sbjct: 550  GAIYAAINDNDYANMQGTSMASPHVAGAAVLVKQYLLATYPTKSPQEIEALVKHLLMSTA 609

Query: 659  QIHVNPETKTTTSPRQQGAGLLNIDGAVTSGLYVTGKDNYGSISLGNITDTMTFDVTVHN 718
            + HVN ET   TSPRQQGAG+++    A+++GLY+TG+D YGSI+LGN+ DT +F VT+HN
Sbjct: 610  KAHVNKETTAYTSPRQQGAGIIDTAAAISTGLYLTGEDGYGSITLGNVEDTFSFTVTLHN 669

Query: 719  LSNKDKTLRYDTELLTDHVDPQKGRFTLTSHSLKTYQGGEVTVPANGKVTVRVTMDVSQF 778
            ++N+DKTL Y T+L TD     +      TS S  +++  +VTV AN   TV + +D S F
Sbjct: 670  ITNEDKTLNYSTQLTTDTAQKRIDHLGSTSISRDSWR--KVTVKANSSTTVTINVDASSF 727

Query: 779  TKELTKQMPNGYYLEGFVRFRDSQDDQLNRVNIPFVGFKGQFENLAVAEESIYRLKSQGK 838
             +ELT  M NGYYLEGFVRF D  DD  + V+IP+VGF+G+F+NLAV EE IY L + GK
Sbjct: 728  AEELTGLMKNGYYLEGFVRFTDVADDG-DIVSIPYVGFRGEFQNLAVLEEPIYNLIADGK 786

Query: 839  TGFYFDE-SGPKDDIYVGKHFTGLVTLGSETNVSTKTISDNGLHTLGTFKNADGKFILEK 897
             GFYF+  +   + + +  H+TGLVT  +E   ST   SD+ + TLGTFKN  G F+LE
Sbjct: 787  GGFYFEPVTAQPNTVDISHHYTGLVTGSTELIYSTDKRSDSAIKTLGTFKNKAGYFVLEL 846

Query: 898  NAQGNPVLAISPNGDNNQDFAAFKGVFLRKYQGLKASVYHASDKEHKNPLWVS-PESFKG 956
            +  G P LAISPNGD+NQD   FKGVFLR Y  L ASVY A D E  NPLW S P+S  G
Sbjct: 847  DESGKPHLAISPNGDDNQDSLVFKGVFLRNYTDLVASVYAADDTERTNPLWESQPQS--G 904

Query: 957  DKN-FNSDIRFAKSTTLLGTAFSGKSLTGAELPDGHYHYVVSYYPDVVGAKRQEMTFDMI 1015
            DKN ++ + +  KS+ +  T ++G    G  L DG Y YV++Y    V GA  Q M FD+I
Sbjct: 905  DKNIYSGNPKNPKSSIIYPTEWNGTDSDGNALADGKYQYVLTYSSKVPGAAVQTMIFDVI 964

Query: 1016 LDRQKPVLSQATFDPETNRFKPEPLKDRGLAGVRKDSVFYLERKDNKPYTVTINDSYKYV 1075
            +DR+ PV++ AT+D      F P P  ++G +G+ ++ VFYL     +    T+       V
Sbjct: 965  IDRESPVITTATYDETNFTFNPRPAIEKGESGLYREQVFYLVADASGVTTIPSLLKNGDV 1024

Query: 1076 SVEDNKTFVERQADGSFILPLDKAKLGDFYYMVEDFAGNVAIAKL 1120
            +V DNK FV +  DGSF LPLD A +  FYY VED+AGN++  K+
Sbjct: 1025 TVSDNKVFVAQNDDGSFTLPLDLADISKFYYTVEDYAGNISYEKV 1069
```

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

```
Identities = 543/1676 (32%), Positives = 821/1676 (48%), Gaps = 158/1676 (9%)

Query: 24   KQRFSIRKYKLGAVSVLLGTLFFLGGITNVAAD--SVINKPSDIAVEQQVKDSPTSI--- 78
            KQRFS+RKYK G  SVL+G++F +   T VAAD  S +++P+     QQ     T+
Sbjct: 4    KQRFSLRKYKSGTFSVLIGSVFLVM-TTTVAADELSTMSEPTITNHAQQQAQHLTNTELS 62

Query: 79   ANETPTNNTSSALASTAQD-----NLVTKANNSPTETQPVAESHSQATETFSPVANQPVE 133
            + E+ + +TS     T ++     +LV++   +      A   +  ++    A+ P
Sbjct: 63   SAESKSQDTSQITLKTNREKEQSQDLVSEPTTTELADTDAASMANTGSDATQKSASLPPV 122

Query: 134  STQEVSKTPLTKQ--NLAVKSTPAISKETPQNID-SNKIITVPKVWNTGYKGEGTVVAI- 189
            +T +V      TK   +   K    +      ID +++ + +  V     K +  ++A
Sbjct: 123  NT-DVHDWVKTKGAWDKGYKGQGKVVAVIDTGIDPAHQSMRISDVSTAKVKSKEDMLARQ 181

Query: 190  ----IDSGLDIN------HDALQLNDSTKAK--------YQNEQQMNAAKAKAGINYGKW 231
                I+ G  IN      H+ ++ +D+ K          ++N +    A+ KA I   K
Sbjct: 182  KAAGINYGSWINDKVVFAHNYVENSDNIKENQFEDFDEDWENFEFDAEAEPKA-IKKHKI 240

Query: 232  YN----------NKVIFGHNYVDVNTELKEVKSTSHGMHVTSIATANPSKKD-TNEL 277
            Y               +  G + +D      + K  SHGMHVT I   N  +  T E
Sbjct: 241  YRPQSTQAPKETVIKTEETDGSHDIDWTQTDDDTKYESHGMHVTGIVAGNSKEAAATGER 300

Query: 278  IYGVAPEAQVMFMRVFSDEKRGTGPALYVKAIEDAVKLGADSINLSLGGANGSSLVNADDR 337
              G+APEAQVMFMRVF+++  G+  +L++KAIEDAV LGAD INLSLG ANG+ ++
Sbjct: 301  FLGIAPEAQVMFMRVFANDIMGSAESLFIKAIEDAVALGADVINLSLGTANGAQLSGSKP 360
```

```
Query: 338  LIKALEMARLAGVSVVIAAGNDGTFGSGASKPSALYPDYGLVGSPSTAREAISVASYNNT 397
            L++A+E A+ AGVSVV+AAGN+  +GS    P A  PDYGLVGSPST R   SVA+ N+
Sbjct: 361  LMEAIEKAKKAGVSVVVAAGNERVYGSDHDDPLATNPDYGLVGSPSTGRTPTSVAAINSK 420

Query: 398  TLVNKVFNIIGLENNRNLNNGLAAYA---DPKVSDKTFEVGKQYDYVFVGKGNDNDYKDK 454
             ++ ++  +  LEN  +LN+G A Y+   D K    +    K + + +V +  D  Y  +
Sbjct: 421  WVIQRLMTVKELENRADLNHGKAIYSESVDFKDIKDSLGYDKSHQFAYVKESTDAGYNAQ 480

Query: 455  TLNGKIALIERG-DITFTKKVVNAINHGAVGAIIFNNKAGEANLTMSLDPEASAIPAIFT 513
             + GKIALIER  + T+ + +  A  HGA+G +IFNNK G++N +M L      IP+ F
Sbjct: 481  DVKGKIALIERDPNKTYDEMIALAKKHGALGVLIFNNKPGQSNRSMRLTANGMGIPSAFI 540

Query: 514  QKEFGDVLAKNNYK----IVFNNIKNKQANPNAGVLSDFSSWGLTADGQLKPDLSAPGGS 569
              EFG  +++ N       + F+++ +K  +      ++ FS+WGLT+DG LKPD++APGG
Sbjct: 541  SHEFGKAMSQLNGNGTGSLEFDSVVSKAPSQKGNEMNHFSNWGLTSDGYLKPDITAPGGD 600

Query: 570  IYAAINDNEYDMMSGTSMASPHVAGATALVKQYLLKEHPELKKGDIERTVKYLLMSTAKA 629
            IY+  NDN Y   +GTSMASP +AGA+ LVKQYL K  P L K  I   VK LLMS A+
Sbjct: 601  IYSTYNDNHYGSQTGTSMASPQIAGASLLVKQYLEKTQPNLPKEKIADIVKNLLMSNAQI 660

Query: 630  HLNKDTGAYTSPRQQGAGIIDVAAAVQTGLYLTGGENNYGSVTLGNIKDKISFDVTVHNI 689
            H+N +T   TSPRQQGAG++++  AV +GLY+TG ++NYGS++LGNI D ++FDVTVHN+
Sbjct: 661  HVNPETKTTTSPRQQGAGLLNIDGAVTSGLYVTG-KDNYGSISLGNITDTMTFDVTVHNL 719

Query: 690  NKVAKDLHYTTYLNTDQV--KDGFVTLAPQQLGTFTGKTIRIEPGQTQTITIDIDVSKYH 747
            +  K L Y T L TD V  + G  TL    L T+ G  + +      T+ + +DVS++
Sbjct: 720  SNKDKTLRYDTELLTDHVDPQKGRFTLTSHSLKTYQGGEVTVPANGKVTVRVTMDVSQFT 779

Query: 748  DMLKKVMPNGYFLEGYVRFTDPVDGG-EVLSIPYVGFKGEFQNLEVLEKSIYKLVANKEK 806
              L K MPNGY+LEG+VRF D  D      ++IP+VGFKG+F+NL V E+SIY+L +  +
Sbjct: 780  KELTKQMPNGYYLEGFVRFRDSQDDQLNRVNIPFVGFKGQFENLAVAEESIYRLKSQGKT 839

Query: 807  GFYFQPK-QTNEVPGSEDYTALMTTSSEPIYSTDGTSPIQLKALGSYKSIDGKWILQLDQ 865
            GFYF      +++   + +T L+T  SE   ST   S   L  LG++K+ DGK+IL+ +
Sbjct: 840  GFYFDESGPKDDIYVGKHFTGLVTLGSETNVSTKTISDNGLHTLGTFKNADGKFILEKNA 899

Query: 866  KGQPHLAISPNDDQNQDAVAVKGVFLRNFNNLRAKVYRADDVNLQKPLWVSAPQ-AGDKN 924
            +G P LAISPN D NQD  A KGVFLR +  L+A VY A D   + PLWVS     GDKN
Sbjct: 900  QGNPVLAISPNGDNNQDFAAFKGVFLRKYQGLKASVYHASDKEHKNPLWVSPESFKGDKN 959

Query: 925  YYSGNTENPKSTFLYDTEWKGTTTDGIPLEDGKYKYVLTYYSDVPGSKPQQMVFDITLDR 984
            + S +    KST L  T + G +  G  L DG Y YV++YY DV G+K Q+M FD+ LDR
Sbjct: 960  FNS-DIRFAKSTTLLGTAFSGKSLTGAELPDGHYHYVVSYYPDVVGAKRQEMTFDMILDR 1018

Query: 985  QAPTLTTATYDKDRRIFKARPAVEHGESGIFREQVFYLKKDKDGHYNSVLRQQGEDGILV 1044
            Q P L+ AT+D +   FK  P  + G +G+ ++ VFYL++ KD    +V       + V
Sbjct: 1019 QKPVLSQATFDPETNRFKPEPLKDRGLAGVRKDSVFYLER-KDNKPYTVTINDSYKYVSV 1077

Query: 1045 EDNKVFIKQEKDGSFILPKEVNDFSHVYYTVEDYAGNLVSAKLEDLINIGNKNGLVNVKV 1104
            EDNK F++++ DGSFILP +       YY VED+AGN+  AKL D +       + +K+
Sbjct: 1078 EDNKTFVERQADGSFILPLDKAKLGDFYYMVEDFAGNVAIAKLGDHLPQTLGKTPIKLKL 1137

Query: 1105 FSPELNSNVDIDFSYSVKDDKGNIIKKQ------HHGKDLNLLKLPFGTYTFDLFLYDEE 1158
                +   +  +      ++ Q      H  + + L          D F+   E
Sbjct: 1138 TDGNYQTKETLKDNLEMTQSDTGLVTNQAQLAVVHRNQPQSQLT----KMNQDFFISPNE 1193

Query: 1159 RANLISPKSVTVTISEKDSLKDVLFKVNLLKKAALLVEFDKLLP-----KGATVQLVTKT 1213
              N    K        K+++ + L  VN+  K     +  K  P      GA+V  + T
Sbjct: 1194 DGN----KDFVAFKGLKNNVYNDL-TVNVYAKD----DHQKQTPIWSSQAGASVSAIEST 1244

Query: 1214 NTVVDLPKATYSPTDYGKNIPVGDYRLNVTLPSGYSTLENLDDLLVSVKEDQVNLT--KL 1271
                  A Y  T  G  +  GDY+  VT    +  E+     +SV + +  +T  +
Sbjct: 1245 --------AWYGITARGSKVMPGDYQYVVTYRDEHGK-EHQKQYTISVNDKKPMITQGRF 1295

Query: 1272 TLINK----APLINALAEQTDIITQPVFYNAGTHLKNNYLANLEKAQTLIKNRVEQTSID 1327
              IN      P      + + I+ + VFY A    KN    ++ + +  I      T  D
Sbjct: 1296 DTINGVDHFTPDKTKALDSSGIVREEVFYLA---KKNGRKFDVTEGKDGI------TVSD 1346
```

```
Query: 1328 NAIAALRESRQALNGKETDTSLLAKAILAETEIKGNYQFVNASPL----SQSTYIN---- 1379
            N +   +    +    + D   L+       +  GN  F     L          +N
Sbjct: 1347 NKVYIPKNPDGSYTISKRDGVTLSDYYYLVEDRAGNVSFATLRDLKAVGKDKAVVNFGLD 1406

Query: 1380 -QVQLAKNLLQKPNVTQSEVDKALENLDIAKNQLNGHETDYS--GLHHMIIKANVLKQTS 1436
              V   K ++    + +    K +ENL+   N  N     Y    +  +    N  K  S
Sbjct: 1407 LPVPEDKQIVNFTYLVRDADGKPIENLEYYNNSGNSLILPYGKYTVELLTYDTNAAKLES 1466

Query: 1437 SKYQNASQFAKENYNNLIKKAELLLSNR----------QATQAQVEELLNQIKATEQEL- 1485
            K  + +  A  N+  +  K  +L +++           + ++  ++    +Q+   EQ L
Sbjct: 1467 DKIVSFTLSADNNFQQVTFKITMLATSQITAHFDHLLPEGSRVSLKTAQDQLIPLEQSLY 1526

Query: 1486 ----------DGRDRVSSAENYSQSLNDNDSLNTTPINPPNQPQALIFKKGMTKES---- 1531
                      +G   V  +      +  N  +NT P N  ++    + K G   +S
Sbjct: 1527 VPKAYGKTVQEGTYEVVVSLPKGYRIEGNTKVNTLP-NEVHELSLRLVKVGDASDSTGDH 1585

Query: 1532 -----EVAQKRVLGVTSQTDNQKVKTNKLPKTGESTPKITYTILLFSLSMLGLATI 1582
                 +Q      T            LP TGE   K+   + +  L +LGL  +
Sbjct: 1586 KVMSKNNSQALTASATPTKSTTSATAKALPSTGE---KMGLKLRIVGLVLLGLTCV 1638
```

SEQ ID 8964 (GBS92) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 2; MW 48kDa).

GBS92-His was purified as shown in Figure 199, lane 9 .

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

Example 2038

A DNA sequence (GBSx2149) was identified in *S.agalactiae* <SEQ ID 6299> which encodes the amino acid sequence <SEQ ID 6300>. This protein is predicted to be AzlC family protein. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.80   Transmembrane  212 - 228 ( 196 - 230)
   INTEGRAL    Likelihood = -7.27   Transmembrane  167 - 183 ( 159 - 185)
   INTEGRAL    Likelihood = -5.68   Transmembrane  189 - 205 ( 188 - 210)
   INTEGRAL    Likelihood = -2.28   Transmembrane   17 -  33 (  13 -  34)
   INTEGRAL    Likelihood = -1.06   Transmembrane  135 - 151 ( 135 - 151)
   INTEGRAL    Likelihood = -1.01   Transmembrane   61 -  77 (  60 -  77)

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

A related GBS nucleic acid sequence <SEQ ID 10235> which encodes amino acid sequence <SEQ ID 10236> was also identified.

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

```
>GP:AAF10212 GB:AE001921 AzlC family protein [Deinococcus radiodurans]
 Identities = 72/224 (32%), Positives = 117/224 (52%), Gaps = 8/224 (3%)

Query: 6   FKEGVKDALPTALGYISIGLAFGIVASASDLSAIEVGLMSALVYGGSAQFAMCALLLAKA 65
            F +G +  +P  LG +   LA+ + A A+ LS  +  LMS   + G++QFA   L  A A
Sbjct: 7   FWQGFRALVPLWLGTVPFALAYAVTARAAGLSVGDTCLMSLTTFAGASQFAAAGLFGAHA 66

Query: 66  DLMTITMTVFLVNLRNMLMSLHATTIFKSAHLMNQLAIGTLITDESYGV-LLGEALHHKV 124
              ++I +T FL+N R++L  L      +   L  ++      +TDE+YGV ++  A
```

```
Sbjct: 67  GGLSIVLTTFLLNARHLLYGLSLARELRLT-LPQRVVAAQFLTDEAYGVAVVSGARLPGG 125

Query: 125 VSPSWMHGNNVMSYLTWVISTIIGTLLGSTIPNPEMFGLDFALVAMFIGLFVFQLFGMLS 184
           ++ +++ G  +  YL+W +ST++G L GS +P PE  G+       F+GL V     ++
Sbjct: 126 LTFAFLLGAELSLYLSWNVSTLLGALAGSVLPPPEQLGVGVVFPLAFLGLLV----PLVV 181

Query: 185 DGKRLVVYVLASVGLSYFLLATFLSGALSVLLATVVGCSVGVVL 228
           D  RL + V  + GL  + L+  L G L +LLA V G  +G  L
Sbjct: 182 D--RLSLLVALAAGLGGWALSRVLPGGLVILLAGVGGALLGAAL 223
```

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 2039

A DNA sequence (GBSx2150) was identified in *S.agalactiae* <SEQ ID 6301> which encodes the amino acid sequence <SEQ ID 6302>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3794(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 2040

A DNA sequence (GBSx2151) was identified in *S.agalactiae* <SEQ ID 6303> which encodes the amino acid sequence <SEQ ID 6304>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5087(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 10233> which encodes amino acid sequence <SEQ ID 10234> was also identified.

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

```
>GP:BAB04157 GB:AP001508 homosystein methyl transferase [Bacillus halodurans]
 Identities = 397/751 (52%), Positives = 519/751 (68%), Gaps = 14/751 (1%)

Query: 10  SNLGYPRLGEQREWKQAIEAFWAGNLEQKDLEKQLKQLRINHLKKQKEAGIDLIPVGDFS 69
           SNLGYPR+GE REWK+A+E+FWA +  ++ L   +K+LR+NHL+ Q+E  +DLIPVGDF+
Sbjct: 4   SNLGYPRIGENREWKKALESFWANDTTEEQLLATMKELRLNHLRVQQEQEVDLIPVGDFT 63

Query: 70  CYDHVLDLSFQFNVIPKRFDEY--ERNLDLYFAIARGDKDNVASSMKKWFNTNYHYIVPE 127
            YDHVLD++  F +IPKRF +      L  YFA+ARG K+  A  M KW+NTNYHYIVPE
```

```
Sbjct: 64  LYDHVLDMAVMFGIIPKRFLQQGDTPTLSTYFAMARGSKNAQACEMTKWYNTNYHYIVPE 123

Query: 128 WEVETKPHLQNNYLLDLYLEAREVVGDKAKPVITGPITYVSLSSGIVD--FEATVQRLLP 185
           +  P L  N  L+ YLEA+  +G   KPVI GP ++V L+ G  +   + T+Q LLP
Sbjct: 124 LH-DAAPRLTKNAPLEAYLEAKNELGIDGKPVILGPYSFVKLAKGYEEDKLQETIQSLLP 182

Query: 186 LYKQVFQDLIDAGATYIQIDEPIFVTDEGELLVDIAKSVYDFFAREVPQAHFIFQTYFES 245
           LY QV Q+L+DAGA  IQ+DEP  VT      + +   +Y+     +  A    QTYF++
Sbjct: 183 LYIQVIQELVDAGARSIQVDEPSLVTSISAREMALVTRIYEQINEAIADAPLFLQTYFDA 242

Query: 246 AVCLDKLSKLPVTGFGLDFIHGRAENLAAVKQ-GLFREKELFAGIVNGRNIWAVNLEETL 304
               +++  LPV G GLDF+HG A+NL A++  G   +K L AGI++GRNIW  NL E
Sbjct: 243 VTFYEEVVSLPVKGIGLDFVHGGAKNLEALRTFGFPEDKVLAAGIIDGRNIWISNLRERH 302

Query: 305 ALLEEIGPFVK--RLTLQPSSSLLHVPVTTKYETHLDPVLKNGLSFADEKLKELELLASA 362
            L+ ++   V   RL LQPS SLLHVPVTTK E  LDP L   L+FA+EKL EL  L
Sbjct: 303 ELVHQLEQHVAKDRLVLQPSCSLLHVPVTTKREEKLDPTLLGVLAFANEKLTELHTLKQL 362

Query: 363 FDGNKTKGYHEALSR----FSALQAADFRHVALESL-AEVKLERSPYKLRQALQAEKLQL 417
             GN+ +   EAL       +AL+ + +R  A  S   E K    +  R+ LQ EK QL
Sbjct: 363 AAGNEAE-VKEALEANDDALAALEKSGWRSGAATSHNLENKKRPQSFNERRPLQEEKWQL 421

Query: 418 PILPTTTIGSFPQSPEIRKKRLAWKRGNLSDSDYKDFIKTEIRRWIAIQEDLDLDVLVHG 477
           P+LPTTTIGSFPQ+ ++R+ R  W++G LS  +Y+  +K+ I +WI IQE+L LDVLVHG
Sbjct: 422 PLLPTTTIGSFPQTKDVRRTRSLWRKGELSTVEYERTMKSYIEKWINIQEELGLDVLVHG 481

Query: 478 EFERVDMVEFFGQKLAGFTTTKLGWVQSYGSRAVKPPIIYGDVKHIQPLSLEETVYAQSL 537
           EFER DMVEFFG+KL GF  T  GWVQSYGSR VKPPIIYG+V   +P+++ ETVYAQSL
Sbjct: 482 EFERNDMVEFFGEKLDGFAFTANGWVQSYGSRCVKPPIIYGNVSFTEPMTVAETVYAQSL 541

Query: 538 TKKPVKGMLTGPITITNWSFERDDISRSDLFNQIALAIKDEIQLLEQSGIAIIQVDEAAL 597
           T KPVKGMLTGP+TI NWSF RDD+  + + +QIA A+  E+  LE++GI +IQ+DE A+
Sbjct: 542 TDKPVKGMLTGPVTILNWSFVRDDLPLTVIAHQIAEALTHEVTALEEAGIEMIQIDEPAI 601

Query: 598 REGLPLRQQKQQAYLDDAVAAFKIATSSVKDETQIHTHMCYSKFDEIIDSIRALDADVIS 657
           REGLPL+ + QQ YLD AV+AF+ + + VK  TQIHTHMCYS+F E+I++I  LDADVIS
Sbjct: 602 REGLPLKAEDQQEYLDWAVSAFRASCAHVKATTQIHTHMCYSEFHEMIEAIDDLDADVIS 661

Query: 658 IETSRSHGDIIESFETAVYPLGIGLGVYDIHSPRIPTKEEIIVNIQRSLKCLSKEQFWVN 717
           IETSRSHG++I +FE   Y  GIGLGVYDIHSPR+P++EE++  I+R+L  L    FWVN
Sbjct: 662 IETSRSHGEMISAFEKTTYEKGIGLGVYDIHSPRVPSEEEMLNVIRRALTVLPASLFWVN 721

Query: 718 PDCGLKTRREAETIAALEVLVSATKEVRQQL 748
           PDCGLKTR E ET+AAL+ +V+A +  R++L
Sbjct: 722 PDCGLKTRAEKETVAALKNMVAAARAAREEL 752
```

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 2041

A DNA sequence (GBSx2152) was identified in *S.agalactiae* <SEQ ID 6305> which encodes the amino acid sequence <SEQ ID 6306>. This protein is predicted to be metH. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0753(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:BAB05348 GB:AP001512 unknown conserved protein [Bacillus halodurans]
 Identities = 301/610 (49%), Positives = 437/610 (71%), Gaps = 9/610 (1%)

Query: 1   MSKFLEKLKTDILVADGAMGTLLYTYGLDTCHESYNVTHPEKVLAIHQAYIEAGADVIQT 60
           M+  +E LKT+ILV DGAMGTLLY  G+D C E  NVT PEK++A H AY+EAGADVIQT
Sbjct: 1   MTNLVEALKTNILVGDGAMGTLLYEQGIDRCFEELNVTDPEKIVAAHVAYVEAGADVIQT 60

Query: 61  NTYGAQRHRLKNYGLEDQVVSINQAAVNIAHQATLGKETFILGTVGGFRSQRQCDLTLDN 120
           NTY A R +L  Y L+DQV+ IN+AAV +A +A   +ETF+LGT+GG RS +  ++ +
Sbjct: 61  NTYAANRMKLAKYQLDDQVLEINRAAVRLARKAAK-QETFVLGTIGGIRSVQFEEVEIQE 119

Query: 121 IVEETLEQVEALLATGQLDGLLFETYYDIEEITTVLKIVREMTDLPIITNISLHEAGVTS 180
           + +  LEQ++AL++ G +DGLL ET+YD+EE    + + R +TDLP+I ++S+ E GV
Sbjct: 120 VQDVFLEQMKALVSEG-VDGLLLETFYDLEEAKLAVSLARSLTDLPVIAHLSIAEIGVLQ 178

Query: 181 NGKPIVEALSQLVMLGADVIGLNCHLGPYHMIQSLKQVPLFAQSYLSVYPNASQLSLDGE 240
            GK + EA ++L  LGAD++G+NC +GPY M++SL+ V L  ++Y S YPNAS    D
Sbjct: 179 GGKLLEEAFAELEGLGADLVGINCRMGPYQMLRSLETVQLLDRAYYSAYPNASLP--DYR 236

Query: 241 NSQYQFSQNSEYFGKSAELLVAEGVRLIGGCCGTTPDHIRAVKRSIRGLKPIERKVVTPI 300
           + +  +  N EYF +  +  V +GVRL+GGCCGTTP+H+RA  + ++GLKP+  K V
Sbjct: 237 DGRLYYHSNPEYFYEMGKRFVQQGVRLLGGCCGTTPEHVRAFAKVVKGLKPVVSKPVR-- 294

Query: 301 IPVKDFVRRIRRT---DTLVDKVKKEVTIIAELDPPKHLDIVQFQKAIRAIDQKGIAAIT 357
           + +K+ +     +  + L +KVKK +II ELDPPK+L I +F +   A+   G+ A+T
Sbjct: 295 LEIKETLSSTGQKTAREPLAEKVKKQPSIIVELDPPKNLAIDRFVEGAAALKNAGVDAVT 354

Query: 358 LADNSLSNTRICNLSIASLLKDEISTPFLLHIACRDHNLIGLQSRLLGMELLGFNHILAI 417
           +ADNSL++ R+ NL++ ++++ ++    L+H+ CRD NLIGLQS L+G+  LG   +LAI
Sbjct: 355 MADNSLASPRVDNLALGAIIQQQVGARPLVHVTCRDRNLIGLQSHLMGLHALGMTDLLAI 414

Query: 418 TGDPTKLGDFPGATSVYDVTSFKLLSLIKQLNQGLSYSGASLRRPTDFTVAAAFNPNVKN 477
           TGDPTK+GDFPGATSVYDVTSF+L+SLIKQLN+G+S+SG  L +  +F+V AAFNPNV++
Sbjct: 415 TGDPTKVGDFPGATSVYDVTSFQLISLIKQLNEGISFSGKELGQKANFSVGAAFNPNVRH 474

Query: 478 LTRTVKLIEKKVASGADYFMTQPIFDHSVLKELADLTKTVEQPFFIGIMPITSYNNAVFL 537
           L R V+ +EKK+ +GADYFMTQPI++   ++++ + TK +E+P +IGIMP+ +  NA FL
Sbjct: 475 LERAVQRMEKKIEAGADYFMTQPIYNEKQIEDIYEATKHIEKPIYIGIMPLINGRNAEFL 534

Query: 538 HNEVPGIKLSESFLSALEKVKDDKEACLTLALNESKSLIDEALNYFNGIYLITPFLRYDL 597
           HNEVPGIKL++     + +  +D++      L  +KSL+D A +YFNGIYLITPFLRY +
Sbjct: 535 HNEVPGIKLTDQIRERMARAGEDRQKGEREGLAIAKSLLDVATHYFNGIYLITPFLRYGM 594

Query: 598 TLELIDYIQK 607
           T++L  Y+++
Sbjct: 595 TVDLTHYVKE 604
```

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 2042

A DNA sequence (GBSx2153) was identified in *S.agalactiae* <SEQ ID 6307> which encodes the amino acid sequence <SEQ ID 6308>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.55   Transmembrane  127 - 143 ( 121 - 147)
   INTEGRAL    Likelihood = -1.44   Transmembrane  157 - 173 ( 155 - 175)

----- Final Results -----
               bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
```

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

A related GBS nucleic acid sequence <SEQ ID 10231> which encodes amino acid sequence <SEQ ID 10232> was also identified.

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

```
>GP:CAC01354 GB:AL390975 putative integral membrane protein
          [Streptomyces coelicolor A3(2)]
 Identities = 38/98 (38%), Positives = 59/98 (59%)

Query: 113 RIADDVARFGGSWTFIIVFVSIMAIWMLVNIMKPFGIQFDPYPFILLNLALSTIAAIQAP 172
           R+++ VARF G+  FI+    ++ +W++ N+  P G++FD YPFI L L LS  A+  AP
Sbjct: 47  RLSERVARFLGTGRFIVWMTVVIILWVVWNVSAPSGLRFDEYPFIFLTLMLSLQASYAAP 106

Query: 173 LIMMSQNRAADYDRLQARNDFNVNKTSELEIRLLHEKI 210
           LI+++QNR  D DR+    D   N+ S  +   L  +I
Sbjct: 107 LILLAQNRQDDRDRVNLEQDRKQNERSIADTEYLTREI 144
```

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

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

A related GBS gene <SEQ ID 8965> and protein <SEQ ID 8966> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:      -3.84
GvH: Signal Score (-7.5): -5.05
     Possible site: 53
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value:  -9.55 threshold:  0.0
   INTEGRAL    Likelihood = -9.55   Transmembrane  127 - 143 ( 121 - 147)
   INTEGRAL    Likelihood = -1.44   Transmembrane  157 - 173 ( 155 - 175)
   PERIPHERAL  Likelihood =  5.46       27
 modified ALOM score:   2.41

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
ORF01598(637 - 930 of 1341)
GP|9714438|emb|CAC01354.1||AL390975(47 - 144 of 198) putative integral membrane protein
(Streptomyces coelicolor A3(2))
%Match = 8.2
%Identity = 38.8  %Similarity = 61.2
Matches = 38  Mismatches = 38  Conservative Sub.s = 22

600       630       660       690       720       750       780       810
MKEEEKPFNVEERLNKQATIGQRIADDVARFGGSWTFIIVFVSIMAIWMLVNIMKPFGIQFDPYPFILLNLALSTIAAIQ
                        |::: |||| |:  ||:    :: :|:: |: | |::|| ||||:| | ||  |:
RLDQPRPPRRRLLPEWDPESFGRLSERVARFLGTGRFIVWMTVVIILWVVWNVSAPSGLRFDEYPFIFLTLMLSLQASYA
              40        50        60        70        80        90        100

840       870       900       930       960       990      1020      1050
APLIMMSQNRAADYDRLQARNDFNVNKTSELEIRLLHEKIDHMVQQDQFELLEIQKLQTEMLVSLGNQLAQLKQLQK*SF
||||:::|||  | ||:    |   |: |  :   |  :|
APLILLAQNRQDDRDRVNLEQDRKQNERSIADTEYLTREIAALRIGLGEVATRDWIRSELQDLVRDLEERQNGHHPDRGV
```

```
          120       130       140       150       160       170       180
```

SEQ ID 8966 (GBS393) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 3; MW 30.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 177 (lane 4; MW 56kDa) and in Figure 83 (lane 6; MW 56kDa).

GBS393-GST was purified as shown in Figure 217, lane 5.

Example 2043

A DNA sequence (GBSx2154) was identified in *S.agalactiae* <SEQ ID 6309> which encodes the amino acid sequence <SEQ ID 6310>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.29   Transmembrane  274 - 290 ( 271 - 291)

----- Final Results -----
               bacterial membrane --- Certainty=0.2317(Affirmative) < succ>
                bacterial outside --- 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:AAD35508 GB:AE001721 glycerol dehydrogenase [Thermotoga maritima]
 Identities = 94/307 (30%), Positives = 157/307 (50%), Gaps = 21/307 (6%)

Query: 63  VYGTDSTQSNIDKLVANPQVQAADAILGFGGGKALDTAKMVAKELGKNSFTIPTICSNCS 122
           ++G + +   I++L    + +  D ++G GGGK LDTAK VA +L K    +PTI S  +
Sbjct: 62  IFGGECSDEEIERLSGLVE-EETDVVVGIGGGKTLDTAKAVAYKLKKPVVIVPTIASTDA 120

Query: 123 AGTAIAVVYNDDHSFLRYGY-PESPLHIFINTRIIAQAPSKYFWAGIGDGISKAPEVERA 181
             +A++V+Y  +  F RY + P +P  + ++T I+A+AP+++  AG+GD ++   E E
Sbjct: 121 PCSALSVIYTPNGEFKRYLFLPRNPDVVLVDTEIVAKAPARFLVAGMGDALATWFEAESC 180

Query: 182 TLEAKTNKLPHT-AVLGQAVALSSKEAFYQFGEQGLKDVEANLASRAVEEI--ALDILIS 238
             +  N      ++   A+A    E   ++G    + VE    + A+E+I  A  +L
Sbjct: 181 KQKYAPNMTGRLGSMTAYALARLCYETLLEYGVLAKRSVEEKSVTPALEKIVEANTLLSG 240

Query: 239 TGYASNLVNQPDFYYNSCHAHAFYYGTTAIQRQGEFLHGVVVAFGVLV-LHAYFNELEEL 297
            G+ S              AHA + G T ++   ++LHG  VA GVL  L       + +
Sbjct: 241 LGFESG---------GLAAAHAIHNGLTVLENTHKYLHGEKVAIGVLASLFLTDKPRKMI 291

Query: 298 EKVARFNKSLGLPTTLADVSL---SEKDIPKIVEIAMTTNE---YKNTPFDPKMFAQAIL 351
           E+V  F + +GLPTTLA++ L   S++D+ K+ E A   NE    +  P   K    A+
Sbjct: 292 EEVYSFCEEVGLPTTLAEIGLDGVSDEDLMKVAEKACDKNETIHNEPQPVTSKDVFFALK 351

Query: 352 AADAFGQ 358
           AAD +G+
Sbjct: 352 AADRYGR 358
```

There is also homology to SEQ ID 3078.

SEQ ID 6310 (GBS123) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 7; MW 43.3kDa).

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

Example 2044

A DNA sequence (GBSx2155) was identified in *S.agalactiae* <SEQ ID 6311> which encodes the amino acid sequence <SEQ ID 6312>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0974(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 6313> which encodes the amino acid sequence <SEQ ID 6314>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2368(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 92/167 (55%), Positives = 121/167 (72%)

Query: 1   MKIAIIGYSGSGKSTLARKLGNYYNCNVLHLDSIHFAPNWEERKYDDMIDDVSNMLEKRT 60
           +KIAIIG+SGSGKSTLAR LG +Y+C V HLD +HF+ NW+ER   DMI D+S  L K+
Sbjct: 1   LKIAIIGHSGSGKSTLARFLGQHYHCEVFHLDQLHFSSNWQERSDHDMIADLSTCLLKQD 60

Query: 61  WIIEGNYKKLLYQERLADADEIIFFDFNRFNCLWRAFKRYCKFRGKTRPDMANGCPEKLD 120
            IIEGNY   LY+ER+++AD II+ +F+RF+C++RAFKRY  +RGKTRPDMA+ C EK D
Sbjct: 61  LIIEGNYANCLYEERMSEADYIIYVNFSRFHCVYRAFKRYLNYRGKTRPDMADNCQEKFD 120

Query: 121 FEFISWILKDGRSDKQKSNYKQVVEDYPQKIKILKHQRDLDQYLKEL 167
             F+ WIL DGRS  Q   Y+ VV+ Y  K  +L +Q+ L  Y+  +
Sbjct: 121 VAFVKWILLDGRSRNQLKKYQSVVQKYSHKTIVLTNQKQLSHYMNTI 167
```

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

Example 2045

A DNA sequence (GBSx2156) was identified in *S.agalactiae* <SEQ ID 6315> which encodes the amino acid sequence <SEQ ID 6316>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3874(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:CAA41941 GB:X59250 initiation factor IF-1 [Lactococcus lactis]
 Identities = 62/72 (86%), Positives = 70/72 (97%)

Query: 1  MAKEDVIEIEGKVVETMPNAMFTVELENGHQILATVSGKIRKNYIRILVGDRVTVEMSPY 60
          MAK+DVIE++GKVV+TMPNAMFTVELENGHQ+LAT+SGKIRKNYIRIL GD+V VE+SPY
Sbjct: 1  MAKDDVIEVDGKVVDTMPNAMFTVELENGHQVLATISGKIRKNYIRILPGDKVQVELSPY 60
```

```
Query: 61 DLTRGRITYRFK 72
          DLTRGRITYRFK
Sbjct: 61 DLTRGRITYRFK 72
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6317> which encodes the amino acid sequence <SEQ ID 6318>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3253(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 = 67/67 (100%), Positives = 67/67 (100%)

Query: 6  VIEIEGKVVETMPNAMFTVELENGHQILATVSGKIRKNYIRILVGDRVTVEMSPYDLTRG 65
          VIEIEGKVVETMPNAMFTVELENGHQILATVSGKIRKNYIRILVGDRVTVEMSPYDLTRG
Sbjct: 1  VIEIEGKVVETMPNAMFTVELENGHQILATVSGKIRKNYIRILVGDRVTVEMSPYDLTRG 60

Query: 66 RITYRFK 72
          RITYRFK
Sbjct: 61 RITYRFK 67
```

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

Example 2046

A DNA sequence (GBSx2157) was identified in *S.agalactiae* <SEQ ID 6319> which encodes the amino acid sequence <SEQ ID 6320>. This protein is predicted to be adenylate kinase (adk). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

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

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

```
>GP:CAA41940 GB:X59250 adenylate kinase [Lactococcus lactis]
 Identities = 146/214 (68%), Positives = 170/214 (79%), Gaps = 6/214 (2%)

Query: 1   MNLLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVP 60
           MNLLIMGLPGAGKGTQA  IV+ +GV HISTGDMFRAAM N+TEMG+LAKS+IDKGELVP
Sbjct: 1   MNLLIMGLPGAGKGTQAEFIVKNYGVNHISTGDMFRAAMKNETEMGKLAKSFIDKGELVP 60

Query: 61  DEVTNGIVKERLAEDDIAEKGFLLDGYPRTIEQAHALDATLEELGLRLDGVINIKVDPSC 120
           DEVTNGIVKERLA+DDI   GFLLDGYPRTI+QAHALD  LEELG++LD V+NI V+P+
Sbjct: 61  DEVTNGIVKERLAQDDIKASGFLLDGYPRTIDQAHALDTMLEELGIKLDAVVNIVVNPNI 120

Query: 121 LIERLSGRIINRKTGETFHKVFNPPV------DYKEEDYYQREDDKPETVKRRLDVNIAQ 174
           L++RLSGR I R  G T+HK+FNP              D YQR DD PETVK RLDVNI +
Sbjct: 121 LVDRLSGRYICRNCGATYHKIFNPTKVEGTCDVCGSHDLYQRADDVPETVKNRLDVNIKE 180

Query: 175 GEPILEHYRKLGLVTDIEGNQEITEVFADVEKAL 208
             PI+EHY +LGLV +IEG QEI++V  D++K L
```

```
Sbjct: 181 SAPIIEHYTELGLVKNIEGEQEISQVTDDIKKVL 214
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6321> which encodes the amino acid sequence <SEQ ID 6322>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

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

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

```
 Identities = 208/212 (98%), Positives = 212/212 (99%)

Query: 1   MNLLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVP 60
           MNLLIMGLPGAGKGTQAAKIVEEFG+AHISTGDMFRAAMANQTEMGRLAKSYIDKGELVP
Sbjct: 1   MNLLIMGLPGAGKGTQAAKIVEEFGIAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVP 60

Query: 61  DEVTNGIVKERLAEDDIAEKGFLLDGYPRTIEQAHALDATLEELGLRLDGVINIKVDPSC 120
           DEVTNGIVKERLAEDDIAEKGFLLDGYPRTIEQAHALDATLEELGLRLDGVINIKVDPSC
Sbjct: 61  DEVTNGIVKERLAEDDIAEKGFLLDGYPRTIEQAHALDATLEELGLRLDGVINIKVDPSC 120

Query: 121 LIERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIAQGEPILE 180
           L+ERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVN+AQGEPILE
Sbjct: 121 LVERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNMAQGEPILE 180

Query: 181 HYRKLGLVTDIEGNQEITEVFADVEKALLELK 212
           HYRKLGLVTDIEGNQEIT+VFADVEKALLELK
Sbjct: 181 HYRKLGLVTDIEGNQEITDVFADVEKALLELK 212
```

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 8967> and protein <SEQ ID 8968> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
McG: Discrim Score:     -1.04
GvH: Signal Score (-7.5): -1.08
     Possible site: 17
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 0 value:   6.79 threshold:  0.0
   PERIPHERAL  Likelihood =  6.79     106
 modified ALOM score:  -1.86

*** 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:

```
over 213aa

                                                                Lactococcus lactis
  EGAD|8612| adenylate kinase Insert characterized
  SP|P27143|KAD_LACLA ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE). Edit
characterized
  GP|44074|emb|CAA41940.1||X59250 adenylate kinase Insert characterized
  PIR|S17987|S17987 adenylate kinase (EC 2.7.4.3) - subsp. lactis Insert characterized
  PIR|B44812|B44812 adenylate kinase (EC 2.7.4.3) - Insert characterized
```

```
ORF01658(301 - 924 of 1236)
EGAD|8612|8416(1 - 214 of 215) adenylate kinase {Lactococcus lactis}SP|P27143|KAD_LACLA
ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE).GP|44074|emb|CAA41940.1||X59250
adenylate kinase {Lactococcus lactis}PIR|S17987|S17987 adenylate kinase (EC 2.7.4.3) -
Lactococcus lactis subsp. lactisPIR|B44812|B44812 adenylate kinase (EC 2.7.4.3) -
Lactococcus lactis
%Match = 34.8
%Identity = 69.5  %Similarity = 81.0
Matches = 146  Mismatches = 38  Conservative Sub.s = 24

132       162       192       222       252       282       312       342
QAYSF*LQRVLKV*NNSRAIF*RDAMLDS*IQQNRI*VDSVNLLFCFLISPTCCVGFI*KQNKETIMNLLIMGLPGAGKG
                                                                 ||||||||||||||
                                                                 MNLLIMGLPGAGKG
                                                                           10

372       402       432       462       492       522       552       582
TQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVPDEVTNGIVKERLAEDDIAEKGFLLDGYPRTIEQA
|||  ||: :|| |||||||||||| |:||||:||||:|||||||||||||||||||||:|||   ||||||||||:||
TQAEFIVKNYGVNHISTGDMFRAAMKNETEMGKLAKSFIDKGELVPDEVTNGIVKERLAQDDIKASGFLLDGYPRTIDQA
               30        40        50        60        70        80        90

612       642       672       702       732                 774       804
HALDATLEELGLRLDGVINIKVDPSCLIERLSXRIINRKTGETFHKVFNPP-----VDY-KEEDYYQREDDKPETVKRRL
||||  |||||::|| |:|| |:|: |::||| | | |  | |:||:|||       |     | ||| || ||||| ||
HALDTMLEELGIKLDAVVNIVVNPNILVDRLSGRYICRNCGATYHKIFNPTKVEGTCDVCGSHDLYQRADDVPETVKNRL
              110       120       130       140       150       160       170

834       864       894       924       954       984      1014      1044
DVNIAQGEPILEHYRKLGLVTDIEGNQEITEVFADVEKALLELK*IMLIYLHK*ISNDILS*SDL*LLPLYRGHQIEI*G
||||  :  ||:||| :|||| :||| |||::|  |::| |
DVNIKESAPIIEHYTELGLVKNIEGEQEISQVTDDIKKVLG
              190       200       210
```

SEQ ID 8968 (GBS114) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 9; MW 26.9kDa).

The GBS114-His fusion product was purified (Figure 108A; see also Figure 200, lane 8) and used to immunise mice (lane 1+2+3 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 108B), FACS (Figure 108C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Example 2047

A DNA sequence (GBSx2158) was identified in *S.agalactiae* <SEQ ID 6323> which encodes the amino acid sequence <SEQ ID 6324>. This protein is predicted to be preprotein translocase secy subunit (secY). Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-14.01   Transmembrane  217 - 233 ( 209 - 240)
   INTEGRAL    Likelihood = -8.65   Transmembrane  314 - 330 ( 307 - 334)
   INTEGRAL    Likelihood = -6.16   Transmembrane  369 - 385 ( 363 - 392)
   INTEGRAL    Likelihood = -5.36   Transmembrane   19 -  35 (  17 -  40)
   INTEGRAL    Likelihood = -3.93   Transmembrane  180 - 196 ( 179 - 199)
   INTEGRAL    Likelihood = -3.03   Transmembrane  395 - 411 ( 392 - 412)
   INTEGRAL    Likelihood = -2.55   Transmembrane  151 - 167 ( 151 - 168)
   INTEGRAL    Likelihood = -2.02   Transmembrane  117 - 133 ( 117 - 133)
   INTEGRAL    Likelihood = -0.64   Transmembrane  270 - 286 ( 269 - 286)

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

```
bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9467> which encodes amino acid sequence <SEQ ID 9468> was also identified.

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

```
>GP:CAA41939 GB:X59250 SecY protein [Lactococcus lactis]
 Identities = 292/433 (67%), Positives = 361/433 (82%), Gaps = 2/433 (0%)

Query: 1   MFLKLLRDALKVKMVRNKILFTIFILLVFRIGTHITVPGINVKSLEQMGELPFLNMLNLV 60
           MF K L++A KVK VR +ILFTIFIL VFR+G HIT PG+NV++L+Q+ +LPFL+M+NLV
Sbjct: 1   MFFKTLKEAFKVKDVRARILFTIFILFVFRLGAHITAPGVNVQNLQQVADLPFLSMMNLV 60

Query: 61  SGNAMRNFSVFSMGVSPYITASIVVQLLQMDILPKFVEWGKQGEVGRRKLNQATRYISLF 120
           SGNAM+N+S+F+MGVSPYITASI+VQLLQMDILPKFVEW KQGE+GRRKLNQATRYI+L
Sbjct: 61  SGNAMQNYSLFAMGVSPYITASIIVQLLQMDILPKFVEWSKQGEIGRRKLNQATRYITLV 120

Query: 121 LAFVQSIGITAGFNTLSSVALVKTPNVQTYLLIGAILTTGSMVVTWLGEQITDKGFGNGV 180
           LA  QSIGITAGF  +SS+ +V+ PN Q+YL+IG +LTTGSMVVTW+GEQI +KGFG+GV
Sbjct: 121 LAMAQSIGITAGFQAMSSLNIVQNPNWQSYLMIGVLLTTGSMVVTWMGEQINEKGFGSGV 180

Query: 181 SMIIFAGIISSIPSAITTIYEDFFVNVRSSAITNSYIFVGILIVAVLAIVFFTTFIQQAE 240
           S+IIFAGI+S IPSAI ++Y++ F+NVR S I  S+IFV  LI++ + I++ TTF+QQAE
Sbjct: 181 SVIIFAGIVSGIPSAIKSVYDEKFLNVRPSEIPMSWIFVIGLILSAIVIIYVTTFVQQAE 240

Query: 241 YKIPIQYTKLVQGAPTSSYLPLKVNPAGVIPVIFASSITTIPSTIIPFFQ--NGKEIPWL 298
            K+PIQYTKL QGAPTSSYLPL+VNPAGVIPVIFA SITT P+TI+ F Q   G  + WL
Sbjct: 241 RKVPIQYTKLTQGAPTSSYLPLRVNPAGVIPVIFAGSITTAPATILQFLQRSQGSNVGWL 300

Query: 299 TKLQELLNYQTPVGMIIYAILIILFSFFYTFVQVNPEKTAENLQKNSSYIPSIRPGRETE 358
           + LQ  L+Y T  GM+ YA+LI+LF+FFY+FVQVNPEK AENLQK  SYIPS+RPG+ TE
Sbjct: 301 STLQNALSYTTWTGMLFYALLIVLFTFFYSFVQVNPEKMAENLQKQGSYIPSVRPGKGTE 360

Query: 359 EYMSSLLKKLATIGSVFLAFISLLPIIAQQALHLSSSIALGGTSLLILIATGIEGMKQLE 418
           +Y+S LL +LAT+GS+FL  IS++PI AQ    L   +ALGGTSLLILI   I+ +KQLE
Sbjct: 361 KYVSRLLMRLATVGSLFLGLISIIPIAAQNVWGLPKIVALGGTSLLILIQVAIQAVKQLE 420

Query: 419 GYLLKRRYVGFMN 431
           GYLLKR+Y GFM+
Sbjct: 421 GYLLKRKYAGFMD 433
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3987> which encodes the amino acid sequence <SEQ ID 3988>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-14.70   Transmembrane  233 - 249 ( 226 - 255)
   INTEGRAL    Likelihood = -8.12   Transmembrane  330 - 346 ( 323 - 350)
   INTEGRAL    Likelihood = -6.10   Transmembrane  384 - 400 ( 378 - 403)
   INTEGRAL    Likelihood = -5.20   Transmembrane   35 -  51 (  33 -  56)
   INTEGRAL    Likelihood = -4.09   Transmembrane  199 - 215 ( 195 - 215)
   INTEGRAL    Likelihood = -3.56   Transmembrane  167 - 183 ( 165 - 184)
   INTEGRAL    Likelihood = -1.65   Transmembrane  411 - 427 ( 411 - 428)
   INTEGRAL    Likelihood = -1.49   Transmembrane  133 - 149 ( 133 - 149)
   INTEGRAL    Likelihood = -0.64   Transmembrane  286 - 302 ( 285 - 302)

----- Final Results -----
              bacterial membrane --- Certainty=0.6880(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 = 377/434 (86%), Positives = 417/434 (95%)

Query: 1   MFLKLLRDALKVKMVRNKILFTIFILLVFRIGTHITVPGINVKSLEQMGELPFLNMLNLV 60
           MFLK+L+DALK+K VRNKI FTIFI+LVFRIGTHITVPG+N KSLEQ+ ELPFLNMLNLV
Sbjct: 17  MFLKILKDALKIKTVRNKIFFTIFIILVFRIGTHITVPGVNAKSLEQLSELPFLNMLNLV 76

Query: 61  SGNAMRNFSVFSMGVSPYITASIVVQLLQMDILPKFVEWGKQGEVGRRKLNQATRYISLF 120
           SGNAMRNFSVFSMGVSPYITASIVVQLLQMDILPKFVEWGKQGEVGRRKLNQATRYISL
Sbjct: 77  SGNAMRNFSVFSMGVSPYITASIVVQLLQMDILPKFVEWGKQGEVGRRKLNQATRYISLV 136

Query: 121 LAFVQSIGITAGFNTLSSVALVKTPNVQTYLLIGAILTTGSMVVTWLGEQITDKGFGNGV 180
           LAF QSIGITAGFNTLS+VALVKTP+++TYLLIGA+LTTGS++VTWLGEQITDKGFGNGV
Sbjct: 137 LAFAQSIGITAGFNTLSNVALVKTPDIKTYLLIGALLTTGSVIVTWLGEQITDKGFGNGV 196

Query: 181 SMIIFAGIISSIPSAITTIYEDFFVNVRSSAITNSYIFVGILIVAVLAIVFFTTFIQQAE 240
           SMIIFAGIISSIPSAI TI ED+FVNV++S + +SY+ VGILI+AVLAIVFFTT++QQAE
Sbjct: 197 SMIIFAGIISSIPSAIATIREDYFVNVKASDLHSSYLIVGILIIAVLAIVFFTTYVQQAE 256

Query: 241 YKIPIQYTKLVQGAPTSSYLPLKVNPAGVIPVIFASSITTIPSTIIPFFQNGKEIPWLTK 300
           YKIPIQYTKL+QGAPTSSYLPLKVNPAGVIPVIFASSITTIPSTIIPF QNG+++PWL +
Sbjct: 257 YKIPIQYTKLMQGAPTSSYLPLKVNPAGVIPVIFASSITTIPSTIIPFVQNGRDLPWLNR 316

Query: 301 LQELLNYQTPVGMIIYAILIILFSFFYTFVQVNPEKTAENLQKNSSYIPSIRPGRETEEY 360
           LQE+ NYQTPVGMI+YA+LIILFSFFYTFVQVNPEKTAENLQKNSSYIPS+RPGRETE++
Sbjct: 317 LQEIFNYQTPVGMIVYALLIILFSFFYTFVQVNPEKTAENLQKNSSYIPSVRPGRETEQF 376

Query: 361 MSSLLKKLATIGSVFLAFISLLPIIAQQALHLSSSIALGGTSLLILIATGIEGMKQLEGY 420
           MS+LLKKLAT+G++FLAFISL PI AQQAL+LSSSIALGGTSLLILI+TGIEGMKQLEGY
Sbjct: 377 MSALLKKLATVGAIFLAFISLAPIAAQQALNLSSSIALGGTSLLILISTGIEGMKQLEGY 436

Query: 421 LLKRRYVGFMNTTE 434
           LLKR+YVGFMNT E
Sbjct: 437 LLKRKYVGFMNTAE 450
```

A related GBS gene <SEQ ID 8969> and protein <SEQ ID 8970> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      6.16
GvH: Signal Score (-7.5): -4.32
     Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 9 value: -14.01 threshold:  0.0
   INTEGRAL    Likelihood =-14.01   Transmembrane  217 - 233 ( 209 - 240)
   INTEGRAL    Likelihood = -9.98   Transmembrane  311 - 327 ( 307 - 334)
   INTEGRAL    Likelihood = -6.16   Transmembrane  369 - 385 ( 363 - 392)
   INTEGRAL    Likelihood = -5.36   Transmembrane   19 -  35 (  17 -  40)
   INTEGRAL    Likelihood = -3.93   Transmembrane  180 - 196 ( 179 - 199)
   INTEGRAL    Likelihood = -3.03   Transmembrane  395 - 411 ( 392 - 412)
   INTEGRAL    Likelihood = -2.55   Transmembrane  151 - 167 ( 151 - 168)
   INTEGRAL    Likelihood = -2.02   Transmembrane  117 - 133 ( 117 - 133)
   INTEGRAL    Likelihood = -0.64   Transmembrane  270 - 286 ( 269 - 286)
   PERIPHERAL  Likelihood =  0.95       69
 modified ALOM score:   3.30

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
ORF01657(301 - 1596 of 1902)
EGAD|6545|6344(1 - 434 of 439) preprotein translocase secy subunit {Lactococcus lactis}
SP|P27148|SECY_LACLA PREPROTEIN TRANSLOCASE SECY SUBUNIT. GP|44073|emb|CAA41939.1||X59250
```

```
SecY protein {Lactococcus lactis} PIR|S17985|S17985 preprotein translocase secY -
Lactococcus lactis subsp. lactis
%Match = 46.6
%Identity = 67.0  %Similarity = 84.1
Matches = 290  Mismatches = 68  Conservative Sub.s = 74

72        102       132       162       192       222       252       282
HQCKRICSCEP*PIKCL*RWY*SNSSCS*RSWNRAC*KIRR*NSW*W*IN*EIVC*SS*IF*IC*SSYHC*RWFNRSHLI

312       342       372       402       432       462       492       522
NER*LIMFLKLLRDALKVKMVRNKILFTIFILLVFRIGTHITVPGINVKSLEQMGELPFLNMLNLVSGNAMRNFSVFSMG
      ||:| |::|:||| || :||||||||:|||:| ||| ||:||::|:|: :||||:|:||||||||:|:|:|:||
     MFFKTLKEAFKVKDVRARILFTIFILFVFRLGAHITAPGVNVQNLQQVADLPFLSMMNLVSGNAMQNYSLFAMG
               10        20        30        40        50        60        70

552       582       612       642       672       702       732       762
VSPYITASIVVQLLQMDILPKFVEWGKQGEVGRRKLNQATRYISLFLAFVQSIGITAGFNTLSSVALVKTPNVQTYLLIG
|||||||||:|||||||||||||||| ||||:||||||||||||:| ||  |||||||||  :||: :|: || |:||:||
VSPYITASIIVQLLQMDILPKFVEWSKQGEIGRRKLNQATRYITLVLAMAQSIGITAGFQAMSSLNIVQNPNWQSYLMIG
              90        100       110       120       130       140       150

792       822       852       882       912       942       972       1002
AILTTGSMVVTWLGEQITDKGFGNGVSMIIFAGIISSIPSAITTIYEDFFVNVRSSAITNSYIFVGILIVAVLAIVFFTT
 :|||||||||:|||| :||||:|||:||||||:| ||||| ::|:: |:||| | |  |:||| ||:: : |:: ||
VLLTTGSMVVTWMGEQINEKGFGSGVSVIIFAGIVSGIPSAIKSVYDEKFLNVRPSEIPMSWIFVIGLILSAIVIIYVTT
              170       180       190       200       210       220       230

1032      1062      1092      1122      1152      1176      1206      1236
FIQQAEYKIPIQYTKLVQGAPTSSYLPLKVNPAGVIPVIFASSITTIPSTIIPFFQ--NGKEIPWLTKLQELLNYQTPVG
|:|||| |:||||||| |||||||||||:|||||||||||| |||| |:||: |:|   |  : ||: ||  |:| |  |
FVQQAERKVPIQYTKLTQGAPTSSYLPLRVNPAGVIPVIFAGSITTAPATILQFLQRSQGSNVGWLSTLQNALSYTTWTG
              250       260       270       280       290       300       310

1266      1296      1326      1356      1386      1416      1446      1476
MIIYAILIILFSFFXTFXQVNPEKTAENLQKNSSYIPSIRPGRETEEYMSSLLKKLATIGSVFLAFISLLPIIAQQALHL
|: ||:||:||:|| :| |||||| |||||  |||||:|||: ||:|:| || :|||:||:|| :||::|| ||    |
MLFYALLIVLFTFFYSFVQVNPEKMAENLQKQGSYIPSVRPGKGTEKYVSRLLMRLATVGSLFLGLISIIPIAAQNVWGL
              330       340       350       360       370       380       390

1506      1536      1566      1596      1626      1656      1686      1716
SSSIALGGTSLLILIATGIEGMKQLEGYLLKRRYVGFMNTTE*NIG*LCQPSILFFNKSDMLCWIYLKTK*GDYNESFNY
   :||||||||||   |: :||||||||||:| |||:
PKIVALGGTSLLILIQVAIQAVKQLEGYLLKRKYAGFMDNPLETK
              410       420       430
```

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

Example 2048

A DNA sequence (GBSx2159) was identified in *S.agalactiae* <SEQ ID 6325> which encodes the amino acid sequence <SEQ ID 6326>. This protein is predicted to be 50S ribosomal protein L15 (rplO). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5259(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:AAB54021 GB:U96620 ribosomal protein L15 [Staphylococcus aureus]
```

```
 Identities = 116/146 (79%), Positives = 128/146 (87%)

Query: 1   MKLHELKPAEGSRKVRNRVGRGTSSGNGKTSGRGQKGQKARSGGGVRLGFEGGQTPLFRR 60
            MKLHELKPAEGSRK RNRVGRG ++GNGKTSGRG KGQKARSGGGVR GFEGGQ PLFRR
Sbjct: 1   MKLHELKPAEGSRKERNRVGRGVATGNGKTSGRGHKGQKARSGGGVRPGFEGGQLPLFRR 60

Query: 61  MPKRGFSNINAKEYALVNLDQLNVFEDGTEVTPVVLKEAGIVRAEKSGVKILGNGELTKK 120
            +PKRGF+NIN KEYA+VNLDQLN FEDGTEVTP +L E+G+V+ EKSG+KILGNG L KK
Sbjct: 61  LPKRGFTNINRKEYAIVNLDQLNKFEDGTEVTPALLVESGVVKNEKSGIKILGNGSLDKK 120

Query: 121 LSVKAAKFSKSAEAAITAKGGSIEVI 146
            L+VKA KFS SA  AI AKGG+ EVI
Sbjct: 121 LTVKAHKFSASAAEAIDAKGGAHEVI 146
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6327> which encodes the amino acid sequence <SEQ ID 6328>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5329(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 = 135/146 (92%), Positives = 142/146 (96%)

Query: 1   MKLHELKPAEGSRKVRNRVGRGTSSGNGKTSGRGQKGQKARSGGGVRLGFEGGQTPLFRR 60
            MKLHELK AEGSRKVRNRVGRGTSSGNGKTSGRGQKGQKARSGGGVRLGFEGGQTPLFRR
Sbjct: 1   MKLHELKAAEGSRKVRNRVGRGTSSGNGKTSGRGQKGQKARSGGGVRLGFEGGQTPLFRR 60

Query: 61  MPKRGFSNINAKEYALVNLDQLNVFEDGTEVTPVVLKEAGIVRAEKSGVKILGNGELTKK 120
            +PKRGF+NIN KEYALVNLDQLNVF+DGTEVTP +LK+AGIVRAEKSGVK+LGNGELTKK
Sbjct: 61  IPKRGFTNINTKEYALVNLDQLNVFDDGTEVTPAILKDAGIVRAEKSGVKVLGNGELTKK 120

Query: 121 LSVKAAKFSKSAEAAITAKGGSIEVI 146
            L+VKAAKFSKSAEAAI AKGGSIEVI
Sbjct: 121 LTVKAAKFSKSAEAAIIAKGGSIEVI 146
```

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

Example 2049

A DNA sequence (GBSx2160) was identified in *S.agalactiae* <SEQ ID 6329> which encodes the amino acid sequence <SEQ ID 6330>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1162(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:AAB54020 GB:U96620 ribosomal protein L30 [Staphylococcus
          aureus]
 Identities = 40/58 (68%), Positives = 46/58 (78%)

Query: 1  MAQIKITLTKSPIGRKPEQRKTVVALGLGKLNSSVVKEDNAAIRGMVNAISHLVTVEE 58
           MA+++ITLT+S IGR   QRKTV ALGL K NSSVV EDN AIRG +N + HLVTVEE
```

```
Sbjct: 1  MAKLQITLTRSVIGRPETQRKTVEALGLKKTNSSVVVEDNPAIRGQINKVKHLVTVEE 58
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6331> which encodes the amino acid sequence <SEQ ID 6332>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1088(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
Identities = 56/58 (96%), Positives = 57/58 (97%)

Query: 1  MAQIKITLTKSPIGRKPEQRKTVVALGLGKLNSSVVKEDNAAIRGMVNAISHLVTVEE 58
          MAQIKITLTKSPIGRKPEQRKTVVALGLGKLNSSVVKEDNAAIRGMV AISHLVTVE+
Sbjct: 1  MAQIKITLTKSPIGRKPEQRKTVVALGLGKLNSSVVKEDNAAIRGMVTAISHLVTVED 58
```

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

Example 2050

A DNA sequence (GBSx2161) was identified in *S.agalactiae* <SEQ ID 6333> which encodes the amino acid sequence <SEQ ID 6334>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3226(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 2051

A DNA sequence (GBSx2162) was identified in *S.agalactiae* <SEQ ID 6335> which encodes the amino acid sequence <SEQ ID 6336>. This protein is predicted to be 30S ribosomal protein S5 (rpsE). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3179(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:AAA22699 GB:M57621 ribosomal protein S5 [Bacillus
          stearothermophilus]
```

```
 Identities = 119/158 (75%), Positives = 139/158 (87%)

Query: 6   NAVELEERVVAINRVTKVVKGGRRLRFAALVVVGDRNGRVGFGTGKAQEVPEAIRKAVEA 65
           N +ELEERVVA+NRV KVVKGGRRLRF+ALVVVGD+NG VGFGTGKAQEVPEAIRKA+E
Sbjct: 7   NKLELEERVVAVNRVAKVVKGGRRLRFSALVVVGDKNGHVGFGTGKAQEVPEAIRKAIED 66

Query: 66  AKKNMVEVPMVGTTIPHEVRSEFGGAKVLLKPAVEGAGVAAGGAVRAVIELAGVADITSK 125
           AKKN++EVP+VGTTIPHEV   FG  +++LKPA EG GV AGG  RAV+ELAG++DI SK
Sbjct: 67  AKKNLIEVPIVGTTIPHEVIGHFGAGEIILKPASEGTGVIAGGPARAVLELAGISDILSK 126

Query: 126 SLGSNTPINIVRATVEGLKQLKRAEEVAALRGISVSDL 163
           S+GSNTPIN+VRAT +GLKQLKRAE+VA LRG +V +L
Sbjct: 127 SIGSNTPINMVRATFDGLKQLKRAEDVAKLRGKTVEEL 164
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6337> which encodes the amino acid sequence <SEQ ID 6338>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3179(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 158/164 (96%), Positives = 161/164 (97%)

Query: 1   MAFKDNAVELEERVVAINRVTKVVKGGRRLRFAALVVVGDRNGRVGFGTGKAQEVPEAIR 60
           MAFKDNAVELEERVVAINRVTKVVKGGRRLRFAALVVVGD NGRVGFGTGKAQEVPEAIR
Sbjct: 1   MAFKDNAVELEERVVAINRVTKVVKGGRRLRFAALVVVGDGNGRVGFGTGKAQEVPEAIR 60

Query: 61  KAVEAAKKNMVEVPMVGTTIPHEVRSEFGGAKVLLKPAVEGAGVAAGGAVRAVIELAGVA 120
           KAVEAAKKNM+EVPMVGTTIPHEV + FGGAKVLLKPAVEG+GVAAGGAVRAVIELAGVA
Sbjct: 61  KAVEAAKKNMIEVPMVGTTIPHEVYTNFGGAKVLLKPAVEGSGVAAGGAVRAVIELAGVA 120

Query: 121 DITSKSLGSNTPINIVRATVEGLKQLKRAEEVAALRGISVSDLA 164
           DITSKSLGSNTPINIVRATVEGLKQLKRAEEVAALRGISVSDLA
Sbjct: 121 DITSKSLGSNTPINIVRATVEGLKQLKRAEEVAALRGISVSDLA 164
```

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

Example 2052

A DNA sequence (GBSx2163) was identified in *S.agalactiae* <SEQ ID 6339> which encodes the amino acid sequence <SEQ ID 6340>. This protein is predicted to be 50S ribosomal protein L18 (rplR). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4488(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 9465> which encodes amino acid sequence <SEQ ID 9466> was also identified.

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

```
>GP:AAB06815 GB:L47971 ribosomal protein L18 [Bacillus subtilis]
 Identities = 86/120 (71%), Positives = 97/120 (80%), Gaps = 2/120 (1%)

Query: 4   VISKPDKNKIRQKRHRRVRGKLSGTADRPRLNIFRSNTGIYAQVIDDVAGVTLASASTLD 63
           +I+K  KN  R KRH RVR KLSGTA+RPRLN+FRSN  IYAQ+IDDV GVTLASASTLD
Sbjct: 1   MITKTSKNAARLKRHARVRAKLSGTAERPRLNVFRSNKHIYAQIIDDVNGVTLASASTLD 60

Query: 64  KE--VSNGTKTEQAVVVGKLVAERAVAKGISEVVFDRGGYLYHGRVKALADSARENGLKF 121
           K+  V +   T  A  VG+LVA+RA  KGIS+VVFDRGGYLYHGRVKALAD+ARE GLKF
Sbjct: 61  KDLNVESTGDTSAATKVGELVAKRAAEKGISDVVFDRGGYLYHGRVKALADAAREAGLKF 120
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6341> which encodes the amino acid sequence <SEQ ID 6342>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4488(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 = 116/121 (95%), Positives = 120/121 (98%)

Query: 1   MKIVISKPDKNKIRQKRHRRVRGKLSGTADRPRLNIFRSNTGIYAQVIDDVAGVTLASAS 60
           +KIVISKPDKNKIRQKRHRRVRGKLSGTADRPRLN+FRSNTGIYAQVIDDVAGVTLASAS
Sbjct: 1   VKIVISKPDKNKIRQKRHRRVRGKLSGTADRPRLNVFRSNTGIYAQVIDDVAGVTLASAS 60

Query: 61  TLDKEVSNGTKTEQAVVVGKLVAERAVAKGISEVVFDRGGYLYHGRVKALADSARENGLKF 121
           TLDK+VS GTKTEQAVVVGKLVAERAVAKGISEVVFDRGGYLYHGRVKALAD+ARENGLKF
Sbjct: 61  TLDKDVSKGTKTEQAVVVGKLVAERAVAKGISEVVFDRGGYLYHGRVKALADAARENGLKF 121
```

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

Example 2053

A DNA sequence (GBSx2164) was identified in *S.agalactiae* <SEQ ID 6343> which encodes the amino acid sequence <SEQ ID 6344>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1530(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:AAA22700 GB:M57622 ribosomal protein L6 [Bacillus
          stearothermophilus]
 Identities = 108/178 (60%), Positives = 133/178 (74%)

Query: 1   MSRIGNKVITLPAGVEIINKDNVVTVKGPKGQLTREFNKNIGITVEGTEVTVTRPNDSKE 60
           M R+G K I +PAGV +     N VTVKGPKG+LTR F+ ++ ITVEG  +TVTRP+D K
Sbjct: 1   MXRVGKKPIEIPAGVTVTVNGNTVTVKGPKGELTRTFHPDMTITVEGNVITVTRPSDEKH 60

Query: 61  MKTIHGTTRANLNNMVVGVSEGFKKALEMRGVGYRAQLQGSKLVLSVGKSHQDEVEAPEG 120
           + +HGTTR+ L NMV GVS+G++KALE+ GVGYRA  QG KLVLSVG SH  E+E  EG
Sbjct: 61  HRALHGTTRSLLANMVEGVSKGYEKALELVGVGYRASKQGKKLVLSVGYSHPVEIEPEEG 120

Query: 121 VTFEVPTPTTTINVIGINKESVGQTAAYVRSLRSPEPYKGKGIRYVGEFVRRKEGKTGK 178
```

```
            +  EVP+ T I V G +K+ VG+ AA +R++R PEPYKGKGIRY GE VR KEGKTGK
Sbjct: 121 LEIEVPSQTKIIVKGADKQRVGELAANIRAVRPPEPYKGKGIRYEGELVRLKEGKTGK 178
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6345> which encodes the amino acid sequence <SEQ ID 6346>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1704(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 = 153/178 (85%), Positives = 166/178 (92%)

Query: 1   MSRIGNKVITLPAGVEIINKDNVVTVKGPKGQLTREFNKNIGITVEGTEVTVTRPNDSKE 60
           MSRIGNKVIT+PAGVE+ N +NV+TVKGPKG+LTREFNKNI I VEGTE+TV RPNDSKE
Sbjct: 1   MSRIGNKVITMPAGVELTNNNNVITVKGPKGELTREFNKNIEIKVEGTEITVVRPNDSKE 60

Query: 61  MKTIHGTTRANLNNMVVGVSEGFKKALEMRGVGYRAQLQGSKLVLSVGKSHQDEVEAPEG 120
           MKTIHGTTRANLNNMVVGVSEGFKK LEM+GVGYRAQLQG+KLVLSVGKSHQDEVEAPEG
Sbjct: 61  MKTIHGTTRANLNNMVVGVSEGFKKDLEMKGVGYRAQLQGTKLVLSVGKSHQDEVEAPEG 120

Query: 121 VTFEVPTPTTTINVIGINKESVGQTAAYVRSLRSPEPYKGKGIRYVGEFVRRKEGKTGK 178
           +TF V  PT+I+V GINKE VGQTAAY+RSLRSPEPYKGKGIRYVGE+VR KEGKTGK
Sbjct: 121 ITFTVANPTSISVEGINKEVVGQTAAYIRSLRSPEPYKGKGIRYVGEYVRLKEGKTGK 178
```

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

Example 2054

A DNA sequence (GBSx2165) was identified in *S.agalactiae* <SEQ ID 6347> which encodes the amino acid sequence <SEQ ID 6348>. This protein is predicted to be 30S ribosomal protein S8 (rpsH). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4356(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:AAB06813 GB:L47971 ribosomal protein S8 [Bacillus subtilis]
 Identities = 100/132 (75%), Positives = 116/132 (87%)

Query: 1   MVMTDPIADFLTRIRNANQAKHEVLEVPASNIKKGIADILKREGFVKNVEVIEDDKQGII 60
           MVMTDPIAD LTRIRNAN  +HE LE+PAS +K+ IA+ILKREGF+++VE +ED KQGII
Sbjct: 1   MVMTDPIADMLTRIRNANMVRHEKLEIPASKLKREIAEILKREGFIRDVEFVEDSKQGII 60

Query: 61  RVFLKYGQNGERVITNLKRISKPGLRVYTKHEDMPKVLNGLGIAIVSTSEGLLTDKEARQ 120
           RVFLKYGQN ERVIT LKRISKPGLRVY K  ++P+VLNGLGIAI+STS+G+LTDKEAR
Sbjct: 61  RVFLKYGQNNERVITGLKRISKPGLRVYAKSNEVPRVLNGLGIAIISTSQGVLTDKEARA 120

Query: 121 KNIGGEVLAYIW 132
           K  GEVLAY+W
Sbjct: 121 KQAGGEVLAYVW 132
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6349> which encodes the amino acid sequence <SEQ ID 6350>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4327(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 = 122/132 (92%), Positives = 129/132 (97%)

Query: 1   MVMTDPIADFLTRIRNANQAKHEVLEVPASNIKKGIADILKREGFVKNVEVIEDDKQGII 60
           MVMTDPIADFLTRIRNANQ KHEVLEVPASNIKKGIA+ILKREGFVKNVEVIEDDKQGII
Sbjct: 1   MVMTDPIADFLTRIRNANQVKHEVLEVPASNIKKGIAEILKREGFVKNVEVIEDDKQGII 60

Query: 61  RVFLKYGQNGERVITNLKRISKPGLRVYTKHEDMPKVLNGLGIAIVSTSEGLLTDKEARQ 120
           RVFLKYG+NGERVITNLKRISKPGLRVY K +DMPKVLNGLGIAI+STSEGLLTDKEARQ
Sbjct: 61  RVFLKYGKNGERVITNLKRISKPGLRVYAKRDDMPKVLNGLGIAIISTSEGLLTDKEARQ 120

Query: 121 KNIGGEVLAYIW 132
           KN+GGEV+AY+W
Sbjct: 121 KNVGGEVIAYVW 132
```

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

Example 2055

A DNA sequence (GBSx2166) was identified in *S.agalactiae* <SEQ ID 6351> which encodes the amino acid sequence <SEQ ID 6352>. This protein is predicted to be ribosomal protein S14 (rpsN). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3833(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:CAB11905 GB:Z99104 ribosomal protein S14 [Bacillus subtilis]
 Identities = 47/61 (77%), Positives = 53/61 (86%)

Query: 1  MAKKSMIAKNKRPAKFSTQAYTRCEKCGRPHSVYRKFQLCRVCFRDLAYKGQVPGVTKAS 60
          MAKKSMIAK +R  KF  Q YTRCE+CGRPHSV RKF+LCR+CFR+LAYKGQ+PGV KAS
Sbjct: 1  MAKKSMIAKQQRTPKFKVQEYTRCERCGRPHSVIRKFKLCRICFRELAYKGQIPGVKKAS 60

Query: 61 W 61
          W
Sbjct: 61 W 61
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6353> which encodes the amino acid sequence <SEQ ID 6354>. Analysis of this protein sequence reveals the following:

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

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

```
                bacterial cytoplasm --- Certainty=0.4747(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 = 55/61 (90%), Positives = 59/61 (96%)

Query: 1  MAKKSMIAKNKRPAKFSTQAYTRCEKCGRPHSVYRKFQLCRVCFRDLAYKGQVPGVTKAS 60
          +AKKSMIAKNKRPAK STQAYTRCEKCGRPHSVYRKF+LCRVCFR+LAYKGQ+PGV KAS
Sbjct: 1  LAKKSMIAKNKRPAKHSTQAYTRCEKCGRPHSVYRKFKLCRVCFRELAYKGQIPGVVKAS 60

Query: 61 W 61
          W
Sbjct: 61 W 61
```

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

Example 2056

A DNA sequence (GBSx2167) was identified in *S.agalactiae* <SEQ ID 6355> which encodes the amino acid sequence <SEQ ID 6356>. This protein is predicted to be 50S ribosomal protein L5 (rplE). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1845(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:BAB03865 GB:AP001507 ribosomal protein L5 (BL6) [Bacillus halodurans]
 Identities = 143/178 (80%), Positives = 162/178 (90%)

Query: 3   NRLKEKYTNEVVPALTEKFNYSSVMAVPKVEKIVLNMGVGDAVSNAKNLEKAAAELALIS 62
           NRLKEKY  E+VP+LTEKFNYSSVMAVPK+EKIV+NMGVGDAV NAK L+KA  EL  I+
Sbjct: 2   NRLKEKYQKEIVPSLTEKFNYSSVMAVPKLEKIVVNMGVGDAVQNAKALDKAVEELTEIT 61

Query: 63  GQKPLITKAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSVSLPRVRDFHGVPTKS 122
           GQKP+ITKAKKSIAGF+LREG+ IGAKVTLRGERMYEFLDKL+SVSLPRVRDF G+  K+
Sbjct: 62  GQKPIITKAKKSIAGFKLREGMPIGAKVTLRGERMYEFLDKLISVSLPRVRDFRGISKKA 121

Query: 123 FDGRGNYTLGVKEQLIFPEINFDDVDKVRGLDIVIVTTANTDEESRELLKGLGMPFAK 180
           FDGRGNYTLGVKEQLIFPEI++D VDKVRG+D+VIVTTA+TDEE+RELL  +GMPF K
Sbjct: 122 FDGRGNYTLGVKEQLIFPEIDYDKVDKVRGMDVVIVTTASTDEEARELLSQMGMPFQK 179
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6357> which encodes the amino acid sequence <SEQ ID 6358>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1793(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 177/180 (98%), Positives = 180/180 (99%)
```

```
Query: 1   MANRLKEKYTNEVVPALTEKFNYSSVMAVPKVEKIVLNMGVGDAVSNAKNLEKAAAELAL 60
           MANRLKEKYTNEV+PALTEKFNY+SVMAVPKVEKIVLNMGVGDAVSNAKNLEKAAAELAL
Sbjct: 1   MANRLKEKYTNEVIPALTEKFNYTSVMAVPKVEKIVLNMGVGDAVSNAKNLEKAAAELAL 60

Query: 61  ISGQKPLITKAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSVSLPRVRDFHGVPT 120
           ISGQKPLITKAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSVSLPRVRDFHGVPT
Sbjct: 61  ISGQKPLITKAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSVSLPRVRDFHGVPT 120

Query: 121 KSFDGRGNYTLGVKEQLIFPEINFDDVDKVRGLDIVIVTTANTDEESRELLKGLGMPFAK 180
           KSFDGRGNYTLGVKEQLIFPEI+FDDVDKVRGLDIVIVTTANTDEESRELLKGLGMPFAK
Sbjct: 121 KSFDGRGNYTLGVKEQLIFPEISFDDVDKVRGLDIVIVTTANTDEESRELLKGLGMPFAK 180
```

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

Example 2057

A DNA sequence (GBSx2169) was identified in *S.agalactiae* <SEQ ID 6359> which encodes the amino acid sequence <SEQ ID 6360>. This protein is predicted to be 50S ribosomal protein L24 (rplX). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1850(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:AAD33285 GB:AF126061 RpL24 [Streptococcus pneumoniae]
 Identities = 89/101 (88%), Positives = 94/101 (92%)

Query: 1   MFVKKGDKVRVIAGKDKGTEAVVLKALPKVNKVVVEGVALIKKHQKPNNENPQGAIVEKE 60
           MFVKKGDKVRVIAGKDKGTEAVVL ALPKVNKV+VEGV ++KKHQ+P NE PQG I+EKE
Sbjct: 1   MFVKKGDKVRVIAGKDKGTEAVVLTALPKVNKVIVEGVNIVKKHQRPTNELPQGGIIEKE 60

Query: 61  APIHVSNVQVLDKNGVAGRVGYKVVDGKKVRYNKKSGEVLD 101
           A IHVSNVQVLDKNGVAGRVGYK VDGKKVRYNKKSGEVLD
Sbjct: 61  AAIHVSNVQVLDKNGVAGRVGYKFVDGKKVRYNKKSGEVLD 101
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6361> which encodes the amino acid sequence <SEQ ID 6362>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1850(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 95/101 (94%), Positives = 99/101 (97%)

Query: 1   MFVKKGDKVRVIAGKDKGTEAVVLKALPKVNKVVVEGVALIKKHQKPNNENPQGAIVEKE 60
           MFVKKGDKVRVIAGKDKGTEAVVLKALPKVNKV+VEGV +IKKHQKPN ENPQGAIVEKE
Sbjct: 1   MFVKKGDKVRVIAGKDKGTEAVVLKALPKVNKVIVEGVGMIKKHQKPNTENPQGAIVEKE 60

Query: 61  APIHVSNVQVLDKNGVAGRVGYKVVDGKKVRYNKKSGEVLD 101
           APIHVSNVQVLDKNGVAGR+GYKVVDGKKVRY+KKSGEVLD
```

```
Sbjct: 61  APIHVSNVQVLDKNGVAGRIGYKVVDGKKVRYSKKSGEVLD 101
```

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

Example 2058

A DNA sequence (GBSx2170) was identified in *S.agalactiae* <SEQ ID 6363> which encodes the amino acid sequence <SEQ ID 6364>. This protein is predicted to be 50S ribosomal protein L14 (rplN). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1004(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:AAD33284 GB:AF126061 RpL14 [Streptococcus pneumoniae]
 Identities = 116/122 (95%), Positives = 120/122 (98%)

Query: 1   MIQQETRLKVADNSGAREILTIKVLGGSGRKFANIGDVIVASVKQATPGGAVKKGDVVKA 60
           MIQ ETRLKVADNSGAREILTIKVLGGSGRKFANIGDVIVASVKQATPGGAVKKGDVVKA
Sbjct: 1   MIQTETRLKVADNSGAREILTIKVLGGSGRKFANIGDVIVASVKQATPGGAVKKGDVVKA 60

Query: 61  VIVRTKTGARRPDGSYIKFDDNAAVIIRDDKTPRGTRIFGPVARELREGGYMKIVSLAPE 120
           VIVRTK+GARR DGSYIKFD+NAAVIIR+DKTPRGTRIFGPVARELREGG+MKIVSLAPE
Sbjct: 61  VIVRTKSGARRADGSYIKFDENAAVIIREDKTPRGTRIFGPVARELREGGFMKIVSLAPE 120

Query: 121 VL 122
           VL
Sbjct: 121 VL 122
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6365> which encodes the amino acid sequence <SEQ ID 6366>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1004(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 = 122/122 (100%), Positives = 122/122 (100%)

Query: 1   MIQQETRLKVADNSGAREILTIKVLGGSGRKFANIGDVIVASVKQATPGGAVKKGDVVKA 60
           MIQQETRLKVADNSGAREILTIKVLGGSGRKFANIGDVIVASVKQATPGGAVKKGDVVKA
Sbjct: 1   MIQQETRLKVADNSGAREILTIKVLGGSGRKFANIGDVIVASVKQATPGGAVKKGDVVKA 60

Query: 61  VIVRTKTGARRPDGSYIKFDDNAAVIIRDDKTPRGTRIFGPVARELREGGYMKIVSLAPE 120
           VIVRTKTGARRPDGSYIKFDDNAAVIIRDDKTPRGTRIFGPVARELREGGYMKIVSLAPE
Sbjct: 61  VIVRTKTGARRPDGSYIKFDDNAAVIIRDDKTPRGTRIFGPVARELREGGYMKIVSLAPE 120

Query: 121 VL 122
           VL
Sbjct: 121 VL 122
```

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

Example 2059

A DNA sequence (GBSx2171) was identified in *S.agalactiae* <SEQ ID 6367> which encodes the amino acid sequence <SEQ ID 6368>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3415(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:AAD33283 GB:AF126061 RpS17 [Streptococcus pneumoniae]
 Identities = 82/86 (95%), Positives = 83/86 (96%)

Query: 1  MERNQRKTLYGRVVSDKMDKTITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV 60
          MERN RK L GRVVSDKMDKTITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV
Sbjct: 1  MERNNRKVLVGRVVSDKMDKTITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV 60

Query: 61 RIMETRPLSATKRFRLVEVVEKAVII 86
          RIMETRPLSATKRFRLVEVVE+AVII
Sbjct: 61 RIMETRPLSATKRFRLVEVVEEAVII 86
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6369> which encodes the amino acid sequence <SEQ ID 6370>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3415(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 = 86/86 (100%), Positives = 86/86 (100%)

Query: 1  MERNQRKTLYGRVVSDKMDKTITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV 60
          MERNQRKTLYGRVVSDKMDKTITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV
Sbjct: 1  MERNQRKTLYGRVVSDKMDKTITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV 60

Query: 61 RIMETRPLSATKRFRLVEVVEKAVII 86
          RIMETRPLSATKRFRLVEVVEKAVII
Sbjct: 61 RIMETRPLSATKRFRLVEVVEKAVII 86
```

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

Example 2060

A DNA sequence (GBSx2172) was identified in *S.agalactiae* <SEQ ID 6371> which encodes the amino acid sequence <SEQ ID 6372>. Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
        bacterial cytoplasm --- Certainty=0.4329(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:AAD33282 GB:AF126061 RpL29 [Streptococcus pneumoniae]
 Identities = 58/68 (85%), Positives = 64/68 (93%)

Query: 1  MKLQEIKDFVKELRGLSQEELAKKENELKKELFDLRFQAAAGQLEKTARLDEVKKQIARV 60
          MKL E+K+FVKELRGLSQEELAK+ENELKKELF+LRFQAA GQLE+TARL EVKKQIAR+
Sbjct: 1  MKLNEVKEFVKELRGLSQEELAKRENELKKELFELRFQAATGQLEQTARLKEVKKQIARI 60

Query: 61 KTVQSEMK 68
          KTVQSE K
Sbjct: 61 KTVQSEAK 68
```

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

Example 2061

A DNA sequence (GBSx2174) was identified in *S.agalactiae* <SEQ ID 6373> which encodes the amino acid sequence <SEQ ID 6374>. This protein is predicted to be RpL16 (rplP). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
        bacterial cytoplasm --- Certainty=0.4574(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:AAD33263 GB:AF126059 RpL16 [Streptococcus pneumoniae]
 Identities = 135/137 (98%), Positives = 137/137 (99%)

Query: 1   MLVPKRVKHRREFRGKMRGEAKGGKEVSFGEYGLQATTSHWITNRQIEAARIAMTRYMKR 60
           MLVPKRVKHRREFRGKMRGEAKGGKEV+FGEYGLQATTSHWITNRQIEAARIAMTRYMKR
Sbjct: 1   MLVPKRVKHRREFRGKMRGEAKGGKEVAFGEYGLQATTSHWITNRQIEAARIAMTRYMKR 60

Query: 61  GGKVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVKRGKVMFEIAGVSEEVAREALRL 120
           GGKVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVKRGKVMFEIAGVSEE+AREALRL
Sbjct: 61  GGKVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVKRGKVMFEIAGVSEEIAREALRL 120

Query: 121 ASHKLPVKCKFVKREAE 137
           ASHKLPVKCKFVKREAE
Sbjct: 121 ASHKLPVKCKFVKREAE 137
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6375> which encodes the amino acid sequence <SEQ ID 6376>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
        bacterial cytoplasm --- Certainty=0.4574(Affirmative) < succ>
         bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 136/137 (99%), Positives = 137/137 (99%)

Query: 1   MLVPKRVKHRREFRGKMRGEAKGGKEVSFGEYGLQATTSHWITNRQIEAARIAMTRYMKR 60
           MLVPKRVKHRREFRGKMRGEAKGGKEVSFGEYGLQATTSHWITNRQIEAARIAMTRYMKR
Sbjct: 1   MLVPKRVKHRREFRGKMRGEAKGGKEVSFGEYGLQATTSHWITNRQIEAARIAMTRYMKR 60

Query: 61  GGKVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVKRGKVMFEIAGVSEEVAREALRL 120
           GGKVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVKRGKVMFEIAGVSEE+AREALRL
Sbjct: 61  GGKVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVKRGKVMFEIAGVSEEIAREALRL 120

Query: 121 ASHKLPVKCKFVKREAE 137
           ASHKLPVKCKFVKREAE
Sbjct: 121 ASHKLPVKCKFVKREAE 137
```

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

Example 2062

A DNA sequence (GBSx2175) was identified in *S.agalactiae* <SEQ ID 6377> which encodes the amino acid sequence <SEQ ID 6378>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3758(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:AAD33280 GB:AF126061 RpS3 [Streptococcus pneumoniae]
 Identities = 200/208 (96%), Positives = 203/208 (97%)

Query: 10  MRVGIIRDWDAKWYAEKEYADYLHEDLAIRKFINKELADASVSTIEIERAVNKVIVSLHT 69
           MRVGIIRDWDAKWYAEKEYADYLHEDLAIRKF+ KELADA+VSTIEIERAVNKV VSLHT
Sbjct: 1   MRVGIIRDWDAKWYAEKEYADYLHEDLAIRKFVQKELADAAVSTIEIERAVNKVNVSLHT 60

Query: 70  AKPGMVIGKGGANVDALRGQLNKLTGKQVHINIIEIKQPDLDAHLVGENIARQLEQRVAF 129
           AKPGMVIGKGGANVDALR +LNKLTGKQVHINIIEIKQPDLDAHLVGE IARQLEQRVAF
Sbjct: 61  AKPGMVIGKGGANVDALRAKLNKLTGKQVHINIIEIKQPDLDAHLVGEGIARQLEQRVAF 120

Query: 130 RRAQKQAIQRTMRAGAKGIKTQVSGRLNGADIARAEGYSEGTVPLHTLRADIDYAWEEAD 189
           RRAQKQAIQR MRAGAKGIKTQVSGRLNGADIARAEGYSEGTVPLHTLRADIDYAWEEAD
Sbjct: 121 RRAQKQAIQRAMRAGAKGIKTQVSGRLNGADIARAEGYSEGTVPLHTLRADIDYAWEEAD 180

Query: 190 TTYGKLGVKVWIYRGEVLPARKNTKGGK 217
           TTYGKLGVKVWIYRGEVLPARKNTKGGK
Sbjct: 181 TTYGKLGVKVWIYRGEVLPARKNTKGGK 208
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6379> which encodes the amino acid sequence <SEQ ID 6380>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3758(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 2063

A DNA sequence (GBSx2176) was identified in *S.agalactiae* <SEQ ID 6381> which encodes the amino acid sequence <SEQ ID 6382>. This protein is predicted to be 50S ribosomal protein L22 (rplV). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2704(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:AAD33279 GB:AF126061 RpL22 [Streptococcus pneumoniae]
 Identities = 99/114 (86%), Positives = 106/114 (92%)

Query: 1   MAEITSAKAMARTVRVSPRKTRLVLDLIRGKNVADAIAILKFTPNKAARVIEKTLNSAIA 60
           MAEITSAKAMARTVRVSPRK+RLVLD IRGK+VADAIAIL FTPNKAA +I K LNSA+A
Sbjct: 1   MAEITSAKAMARTVRVSPRKSRLVLDNIRGKSVADAIAILTFTPNKAAEIILKVLNSAVA 60

Query: 61  NAENNFGLEKANLVVSETFANEGPTMKRFRPRAKGSASPINKRTTHVTVVVSEK 114
           NAENNFGL+KANLVVSE FANEGPTMKRFRPRAKGSASPINKRT H+TV V+EK
Sbjct: 61  NAENNFGLDKANLVVSEAFANEGPTMKRFRPRAKGSASPINKRTAHITVAVAEK 114
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6383> which encodes the amino acid sequence <SEQ ID 6384>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2794(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 113/114 (99%), Positives = 113/114 (99%)

Query: 1   MAEITSAKAMARTVRVSPRKTRLVLDLIRGKNVADAIAILKFTPNKAARVIEKTLNSAIA 60
           MAEITSAKAMARTVRVSPRKTRLVLDLIRGK VADAIAILKFTPNKAARVIEKTLNSAIA
Sbjct: 1   MAEITSAKAMARTVRVSPRKTRLVLDLIRGKKVADAIAILKFTPNKAARVIEKTLNSAIA 60

Query: 61  NAENNFGLEKANLVVSETFANEGPTMKRFRPRAKGSASPINKRTTHVTVVVSEK 114
           NAENNFGLEKANLVVSETFANEGPTMKRFRPRAKGSASPINKRTTHVTVVVSEK
Sbjct: 61  NAENNFGLEKANLVVSETFANEGPTMKRFRPRAKGSASPINKRTTHVTVVVSEK 114
```

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

Example 2064

A DNA sequence (GBSx2177) was identified in *S.agalactiae* <SEQ ID 6385> which encodes the amino acid sequence <SEQ ID 6386>. This protein is predicted to be 30S ribosomal protein S19 (rpsS). Analysis of this protein sequence reveals the following:

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

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

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

The protein is similar to ribosomal protein S19 from *S.pneumoniae*.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6387> which encodes the amino acid sequence <SEQ ID 6388>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3319(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
Identities = 92/92 (100%), Positives = 92/92 (100%)

Query: 1   MGRSLKKGPFVDEHLMKKVEAQANDEKKKVIKTWSRRSTIFPSFIGYTIAVYDGRKHVPV 60
            MGRSLKKGPFVDEHLMKKVEAQANDEKKKVIKTWSRRSTIFPSFIGYTIAVYDGRKHVPV
Sbjct: 19  MGRSLKKGPFVDEHLMKKVEAQANDEKKKVIKTWSRRSTIFPSFIGYTIAVYDGRKHVPV 78

Query: 61  YIQEDMVGHKLGEFAPTRTYKGHAADDKKTRR 92
            YIQEDMVGHKLGEFAPTRTYKGHAADDKKTRR
Sbjct: 79  YIQEDMVGHKLGEFAPTRTYKGHAADDKKTRR 110
```

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

Example 2065

A DNA sequence (GBSx2178) was identified in *S.agalactiae* <SEQ ID 6389> which encodes the amino acid sequence <SEQ ID 6390>. This protein is predicted to be L2 (rplB). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3182(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:AAC45959 GB:U43929 L2 [Bacillus subtilis]
 Identities = 208/277 (75%), Positives = 239/277 (86%)

Query: 1   MGIKVYKPTTNGRRNMTSLDFAEITTNTPEKSLLVSLKNKAGRNNNGRITVRHQGGGHKR 60
            M IK YKP++NGRR MT+ DFAEITT+ PEKSLL  L  K GRNN G++TVRHQGGGHKR
Sbjct: 1   MAIKKYKPSSNGRRGMTTSDFAEITTDKPEKSLLAPLHKKGGRNNQGKLTVRHQGGGHKR 60

Query: 61  HYRLIDFKRNKDGVEAVVKTIEYDPNRTANIALVHYTDGVKAYILAPKGLEVGQRIISGP 120
             YR+IDFKR+KDG+   V T+EYDPNR+ANIAL++Y DG K YILAPKG++VG  ++SGP
Sbjct: 61  QYRVIDFKRDKDGIPGRVATVEYDPNRSANIALINYADGEKRYILAPKGIQVGTEVMSGP 120

Query: 121 EADIKVGNALPLANIPVGTVIHNIELQPGKGAELIRAAGASAQVLGQEGKYVLVRLQSGE 180
            EADIKVGNALPL NIPVGTV+HNIEL+PGKG +L+R+AG SAQVLG+EGKYVLVRL SGE
Sbjct: 121 EADIKVGNALPLINIPVGTVVHNIELKPGKGGQLVRSAGTSAQVLGKEGKYVLVRLNSGE 180

Query: 181 VRMILGTCRATIGTVGNEQQSLVNIGKAGRNRWKGVRPTVRGSVMNPNDHPHGGGEGKAP 240
```

```
            VRMIL  CRA+IG VGNEQ  L+NIGKAGR+RWKG+RPTVRGSVMNPNDHPHGGGEG+AP
Sbjct: 181 VRMILSACRASIGQVGNEQHELINIGKAGRSRWKGIRPTVRGSVMNPNDHPHGGGEGRAP 240

Query: 241 VGRKAPSTPWGKPALGLKTRNKKAKSDKLIVRRRNQK 277
           +GRK+P +PWGKP LG KTR KK KSDK IVRRR  K
Sbjct: 241 IGRKSPMSPWGKPTLGFKTRKKKNKSDKFIVRRRKNK 277
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6391> which encodes the amino acid sequence <SEQ ID 6392>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2560(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 = 264/277 (95%), Positives = 276/277 (99%)

Query: 1   MGIKVYKPTTNGRRNMTSLDFAEITTNTPEKSLLVSLKNKAGRNNNGRITVRHQGGGHKR 60
           +GIKVYKPTTNGRRNMTSLDFAEITT+TPEKSLLVSLK+KAGRNNNGRITVRHQGGGHKR
Sbjct: 1   VGIKVYKPTTNGRRNMTSLDFAEITTSTPEKSLLVSLKSKAGRNNNGRITVRHQGGGHKR 60

Query: 61  HYRLIDFKRNKDGVEAVVKTIEYDPNRTANIALVHYTDGVKAYILAPKGLEVGQRIISGP 120
           HYRLIDFKRNKDGVEAVVKTIEYDPNRTANIALVHYTDGVKAYI+APKGLEVGQRI+SGP
Sbjct: 61  HYRLIDFKRNKDGVEAVVKTIEYDPNRTANIALVHYTDGVKAYIIAPKGLEVGQRIVSGP 120

Query: 121 EADIKVGNALPLANIPVGTVIHNIELQPGKGAELIRAAGASAQVLGQEGKYVLVRLQSGE 180
           +ADIKVGNALPLANIPVGTV+HNIEL+PGKG EL+RAAGASAQVLGQEGKYVLVRLQSGE
Sbjct: 121 DADIKVGNALPLANIPVGTVVHNIELKPGKGGELVRAAGASAQVLGQEGKYVLVRLQSGE 180

Query: 181 VRMILGTCRATIGTVGNEQQSLVNIGKAGRNRWKGVRPTVRGSVMNPNDHPHGGGEGKAP 240
           VRMILGTCRATIGTVGNEQQSLVNIGKAGR+RWKG+RPTVRGSVMNPNDHPHGGGEGKAP
Sbjct: 181 VRMILGTCRATIGTVGNEQQSLVNIGKAGRSRWKGIRPTVRGSVMNPNDHPHGGGEGKAP 240

Query: 241 VGRKAPSTPWGKPALGLKTRNKKAKSDKLIVRRRNQK 277
           VGRKAPSTPWGKPALGLKTRNKKAKSDKLIVRRRN+K
Sbjct: 241 VGRKAPSTPWGKPALGLKTRNKKAKSDKLIVRRRNEK 277
```

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

Example 2066

A DNA sequence (GBSx2180) was identified in *S.agalactiae* <SEQ ID 6393> which encodes the amino acid sequence <SEQ ID 6394>. This protein is predicted to be 50S ribosomal protein L23 (rplW). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1669(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:BAB03855 GB:AP001507 ribosomal protein L23 [Bacillus halodurans]
 Identities = 56/92 (60%), Positives = 67/92 (71%), Gaps = 1/92 (1%)
```

```
Query: 2  NLYDVIKKPVITEKSMVALEAGKYTFEVDTRAHKLLIKQAVEAAFDGVKVASVNTVTVKP 61
          N  DVIK+PVITE+S   +   KYTFEVD RA+K  IK A+E  FD VKVA VNT+  K
Sbjct: 3  NARDVIKRPVITERSTEVMGDKKYTFEVDVRANKTQIKDAIEEIFD-VKVAKVNTMNYKG 61

Query: 62 KAKRVGRYTGFTSKTKKAIITLTADSKAIELF 93
          K KR GRYTGFT++ KKAI+TLT DSK ++ F
Sbjct: 62 KPKRFGRYTGFTARRKKAIVTLTPDSKELDFF 93
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6395> which encodes the amino acid sequence <SEQ ID 6396>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1617(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 = 96/98 (97%), Positives = 97/98 (98%)

Query: 1  MNLYDVIKKPVITEKSMVALEAGKYTFEVDTRAHKLLIKQAVEAAFDGVKVASVNTVTVK 60
          MNLYDVIKKPVITEKSM+ALEAGKYTFEVDTRAHKLLIKQAVEAAFDGVKVASVNTV VK
Sbjct: 1  MNLYDVIKKPVITEKSMIALEAGKYTFEVDTRAHKLLIKQAVEAAFDGVKVASVNTVNVK 60

Query: 61 PKAKRVGRYTGFTSKTKKAIITLTADSKAIELFAAEAE 98
          PKAKRVGRYTGFTSKTKKAIITLTADSKAIELFAAEAE
Sbjct: 61 PKAKRVGRYTGFTSKTKKAIITLTADSKAIELFAAEAE 98
```

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

Example 2067

A DNA sequence (GBSx2181) was identified in *S.agalactiae* <SEQ ID 6397> which encodes the amino acid sequence <SEQ ID 6398>. This protein is predicted to be 50S ribosomal protein L4 (rplD). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.54   Transmembrane  140 - 156 ( 139 - 156)

----- Final Results -----
               bacterial membrane --- Certainty=0.1617(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAC45957 GB:U43929 L4 [Bacillus subtilis]
 Identities = 130/207 (62%), Positives = 160/207 (76%)

Query: 1   MANVKLFDQTGKEVSSVELNEAIFGIEPNESVVFDVVISQRASLRQGTHAVKNRSAVSGG 60
           M  V L++Q G      +ELN ++FGIEPNESVVFD ++ QRASLRQGTH VKNRS V GG
Sbjct: 1   MPKVALYNQNGSTAGDIELNASVFGIEPNESVVFDAILMQRASLRQGTHKVKNRSEVRGG 60

Query: 61  GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLPQKVRRLALKSVYSAKVAE 120
           GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSY YKLP+KVRRLA+KSV S+KV +
Sbjct: 61  GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYSYKLPKKVRRLAIKSVLSSKVID 120

Query: 121 DKFVAVENLSFAAPKTAEFASVLSALSIDSKVLVILEEGNEFAALSARNLPNVTVATATT 180
           +  + +E+L+     KT E A++L  LS++ K L++  + NE  ALSARN+P VTV  A
```

```
Sbjct: 121 NNIIVLEDLTLDTAKTKEMAAILKGLSVEKKALIVTADANEAVALSARNIPGVTVVEANG 180

Query: 181 ASVLDIVNADKLLVTKEAISTIEGVLA 207
           +VLD+VN +KLL+TK A+  +E VLA
Sbjct: 181 INVLDVVNHEKLLITKAAVEKVEEVLA 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6399> which encodes the amino acid sequence <SEQ ID 6400>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2544(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 = 199/207 (96%), Positives = 203/207 (97%)

Query: 1   MANVKLFDQTGKEVSSVELNEAIFGIEPNESVVFDVVISQRASLRQGTHAVKNRSAVSGG 60
           MANVKLFDQTGKEVSSVELN+AIFGIEPNESVVFDVVISQRASLRQGTHAVKNRSAVSGG
Sbjct: 1   MANVKLFDQTGKEVSSVELNDAIFGIEPNESVVFDVVISQRASLRQGTHAVKNRSAVSGG 60

Query: 61  GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLPQKVRRLALKSVYSAKVAE 120
           GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLPQKVRRLALKSVYSAKVAE
Sbjct: 61  GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLPQKVRRLALKSVYSAKVAE 120

Query: 121 DKFVAVENLSFAAPKTAEFASVLSALSIDSKVLVILEEGNEFAALSARNLPNVTVATATT 180
           DKFVAVE LSFAAPKTAEFA VLSALSID+KVLV++EEGNEFAALSARNLPNVTVATA T
Sbjct: 121 DKFVAVEGLSFAAPKTAEFAKVLSALSIDTKVLVLVEEGNEFAALSARNLPNVTVATAAT 180

Query: 181 ASVLDIVNADKLLVTKEAISTIEGVLA 207
           ASVLDIVNADKLLVTKEAISTIE VLA
Sbjct: 181 ASVLDIVNADKLLVTKEAISTIEEVLA 207
```

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

Example 2068

A DNA sequence (GBSx2183) was identified in *S.agalactiae* <SEQ ID 6401> which encodes the amino acid sequence <SEQ ID 6402>. This protein is predicted to be 50S ribosomal protein L3 (rplC). Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2090(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:AAC45956 GB:U43929 L3 [Bacillus subtilis]
 Identities = 157/208 (75%), Positives = 180/208 (86%), Gaps = 2/208 (0%)

Query: 1   MTKGILGKKVGMTQIFTESGEFIPVTVIEATPNVVLQVKTVETDGYEAVQVGFDDKREVL 60
           MTKGILG+K+GMTQ+F E+G+ IPVTVIEA PNVVLQ KT E DGYEA+Q+GFDDKRE L
Sbjct: 1   MTKGILGRKIGMTQVFAENGDLIPVTVIEAAPNVVLQKKTAENDGYEAIQLGFDDKREKL 60

Query: 61  SNKPAKGHVAKANTAPKRFIREFKNIE--GLEVGAELSVEQFEAGDVVDVTGTSKGKGFQ 118
```

```
            SNKP KGHVAKA TAPKRF++E + +E     EVG E+ VE F AG++VDVTG SKGKGFQ
Sbjct: 61  SNKPEKGHVAKAETAPKRFVKELRGVEMDAYEVGQEVKVEIFSAGEIVDVTGVSKGKGFQ 120

Query: 119 GVIKRHGQSRGPMAHGSRYHRRPGSMGPVAPNRVFKNKRLAGRMGGNRVTVQNLEIVQVI 178
            G IKRHGQSRGPM+HGSRYHRRPGSMGPV PNRVFK K L GRMGG ++TVQNLEIV+V
Sbjct: 121 GAIKRHGQSRGPMSHGSRYHRRPGSMGPVDPNRVFKGKLLPGRMGGEQITVQNLEIVKVD 180

Query: 179 PEKNVVLIKGNVPGAKKSLITIKSAVKA 206
             E+N++LIKGNVPGAKKSLIT+KSAVK+
Sbjct: 181 AERNLLLIKGNVPGAKKSLITVKSAVKS 208
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6403> which encodes the amino acid sequence <SEQ ID 6404>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2090(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 = 205/208 (98%), Positives = 207/208 (98%)

Query: 1   MTKGILGKKVGMTQIFTESGEFIPVTVIEATPNVVLQVKTVETDGYEAVQVGFDDKREVL 60
            MTKGILGKKVGMTQIFTESGEFIPVTVIEATPNVVLQVKTVETDGYEAVQVGFDDKREVL
Sbjct: 1   MTKGILGKKVGMTQIFTESGEFIPVTVIEATPNVVLQVKTVETDGYEAVQVGFDDKREVL 60

Query: 61  SNKPAKGHVAKANTAPKRFIREFKNIEGLEVGAELSVEQFEAGDVVDVTGTSKGKGFQGV 120
            SNKPAKGHVAKANTAPKRFIREFKNIEGLEVGAELSVEQFEAGDVVDVTG SKGKGFQGV
Sbjct: 61  SNKPAKGHVAKANTAPKRFIREFKNIEGLEVGAELSVEQFEAGDVVDVTGISKGKGFQGV 120

Query: 121 IKRHGQSRGPMAHGSRYHRRPGSMGPVAPNRVFKNKRLAGRMGGNRVTVQNLEIVQVIPE 180
            IKRHGQSRGPMAHGSRYHRRPGSMGPVAPNRVFKNKRLAGRMGGNRVTVQNLEIVQVIPE
Sbjct: 121 IKRHGQSRGPMAHGSRYHRRPGSMGPVAPNRVFKNKRLAGRMGGNRVTVQNLEIVQVIPE 180

Query: 181 KNVVLIKGNVPGAKKSLITIKSAVKAAK 208
            KNV+L+KGNVPGAKKSLITIKSAVKAAK
Sbjct: 181 KNVILVKGNVPGAKKSLITIKSAVKAAK 208
```

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

Example 2069

A DNA sequence (GBSx2184) was identified in *S.agalactiae* <SEQ ID 6405> which encodes the amino acid sequence <SEQ ID 6406>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.43   Transmembrane    5 -  21 (   5 -  21)

----- Final Results -----
             bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has 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 2070

A DNA sequence (GBSx2185) was identified in *S.agalactiae* <SEQ ID 6407> which encodes the amino acid sequence <SEQ ID 6408>. This protein is predicted to be 30S ribosomal protein S10 (rpsJ). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3160(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:AAB46363 GB:L29637 S10 ribosomal protein [Streptococcus mutans]
 Identities = 98/102 (96%), Positives = 102/102 (99%)

Query: 1   MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGATVAGPVPLPTERSLYTIIRATHKYKD 60
           MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGA+VAGPVPLPTERSLYT+IRATHKYKD
Sbjct: 1   MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGASVAGPVPLPTERSLYTVIRATHKYKD 60

Query: 61  SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPSGVNVEIKL 102
           SREQFEMRTHKRL+DI+NPTQKTVDALMKLDLPSGVNVEIKL
Sbjct: 61  SREQFEMRTHKRLIDIVNPTQKTVDALMKLDLPSGVNVEIKL 102
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6409> which encodes the amino acid sequence <SEQ ID 6410>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3160(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 = 102/102 (100%), Positives = 102/102 (100%)

Query: 1   MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGATVAGPVPLPTERSLYTIIRATHKYKD 60
           MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGATVAGPVPLPTERSLYTIIRATHKYKD
Sbjct: 1   MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGATVAGPVPLPTERSLYTIIRATHKYKD 60

Query: 61  SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPSGVNVEIKL 102
           SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPSGVNVEIKL
Sbjct: 61  SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPSGVNVEIKL 102
```

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

Example 2071

A DNA sequence (GBSx2186) was identified in *S.agalactiae* <SEQ ID 6411> which encodes the amino acid sequence <SEQ ID 6412>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
```

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2538(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 2072

A DNA sequence (GBSx2187) was identified in *S.agalactiae* <SEQ ID 6413> which encodes the amino acid sequence <SEQ ID 6414>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.41   Transmembrane   88 - 104 (  79 - 110)
   INTEGRAL    Likelihood = -8.39   Transmembrane  304 - 320 ( 300 - 324)
   INTEGRAL    Likelihood = -6.58   Transmembrane  185 - 201 ( 180 - 206)
   INTEGRAL    Likelihood = -5.63   Transmembrane  338 - 354 ( 331 - 357)
   INTEGRAL    Likelihood = -5.52   Transmembrane  240 - 256 ( 237 - 259)
   INTEGRAL    Likelihood = -4.99   Transmembrane  383 - 399 ( 375 - 407)
   INTEGRAL    Likelihood = -3.82   Transmembrane   49 -  65 (  48 -  73)
   INTEGRAL    Likelihood = -2.87   Transmembrane  127 - 143 ( 121 - 144)
   INTEGRAL    Likelihood = -2.81   Transmembrane  159 - 175 ( 159 - 177)
   INTEGRAL    Likelihood = -2.18   Transmembrane   30 -  46 (  30 -  47)

----- Final Results -----
              bacterial membrane --- Certainty=0.5564(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:BAB06655 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 132/423 (31%), Positives = 210/423 (49%), Gaps = 16/423 (3%)

Query: 7   IIQLAIPAMIENILQMLMGVVDNYLVAQLGVVAVSGVSVANNIITIYQAIF--IALGASI 64
           +  L  P  IE +L MLMG  D  +++Q    AV+ V V+N I+ +   +F  +A G SI
Sbjct: 11  LFALTWPIFIEILLHMLMGNADTLMLSQYSDDAVAAVGVSNQILAVIIVMFGFVATGTSI 70

Query: 65  ASLLAKSLAGSKKDDAISVCSQAIFLTLLIGAVLGIISIVFGQTFFKLLGTTKSVAQVGG 124
             L+A+ L    ++++A  V   +I   L+ G VLG++ I FG    K +    S+ Q
Sbjct: 71  --LVAQHLGAKERENAGKVAVVSIGANLIFGIVLGLLLIAFGPPILKAMQLDDSLLQEAT 128

Query: 125 LYLAIVGGGVVTLGMLTTLGSFLRVQGQPRLPMYVSIFVNFLNAVLSGFAIFEWR----Y 180
           LYL IVGG  V   ++ T G+ LR     +  MYV+I +N LN + +   IF
Sbjct: 129 LYLQIVGGFSVVQSLIMTAGAILRSHSFTKDVMYVTIGMNILNVIGNYLFIFGPFGIPVL 188

Query: 181 GLVGVAVSTLIARLIGICILAKYL--------PIKKIIKRMTWKISAQIWNLALPSAGER 232
           G+ GVA+ST+++R IG+ ++A  L         P   ++KR         +  + +PSAGE+
Sbjct: 189 GVTGVALSTVVSRTIGLFVIAILLYKRIRGELPFAYLLKRFPRVELRNLLKIGIPSAGEQ 248

Query: 233 LMMRAGDVVIVAIVVQLGTNVVAGNAIGETLTQFNYMPGLGIATATIILTAKYVGQKNRE 292
           L   A  +VI   +  +GT  +      + L  F ++  + I   T IL    VG K  +
Sbjct: 249 LSYNASQLVITYFIAMMGTEALTTKVYTQNLMMFVFLFAVAIGQGTQILIGHQVGAKQIQ 308

Query: 293 SIEETIQSSYYIGLVLMILISSFMLLAGKPLTQLFTNNPSAIKGSLIVILLSFVGVPATI 352
           +       S +I + + + ++         PL  +FT+NP  +     ++LL+ +  P
Sbjct: 309 AAYVRCFRSLWIAMTVSVSMAVVFFAPSTPLLGIFTDNPDILSLGTTLLLLTIILEPGRA 368
```

```
Query: 353 GTLVYTAAWQGLGNAKLPFYTTTIGMWLIRVVLGYLLGIVFELGLLGVWMATIADNIFRW 412
           LV  ++ +  G+ K P Y   + MW I V + YLLG+   LGL+GVW+A IAD  FR
Sbjct: 369 CNLVVISSLRAAGDVKFPVYLAIVSMWGIAVPIAYLLGLPLGLGLIGVWIAFIADEWFRG 428

Query: 413 LFL 415
           L +
Sbjct: 429 LLM 431
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6415> which encodes the amino acid sequence <SEQ ID 6416>. Analysis of this protein sequence reveals the following:

```
     Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.26   Transmembrane   89 - 105 (  85 - 108)
   INTEGRAL    Likelihood = -4.35   Transmembrane  305 - 321 ( 302 - 322)
   INTEGRAL    Likelihood = -3.82   Transmembrane  161 - 177 ( 161 - 180)
   INTEGRAL    Likelihood = -3.82   Transmembrane  192 - 208 ( 189 - 208)
   INTEGRAL    Likelihood = -3.77   Transmembrane  129 - 145 ( 128 - 151)
   INTEGRAL    Likelihood = -3.24   Transmembrane  242 - 258 ( 240 - 258)
   INTEGRAL    Likelihood = -2.81   Transmembrane  378 - 394 ( 377 - 394)
   INTEGRAL    Likelihood = -2.66   Transmembrane  339 - 355 ( 338 - 358)
   INTEGRAL    Likelihood = -2.60   Transmembrane   58 -  74 (  58 -  75)
   INTEGRAL    Likelihood = -2.50   Transmembrane   32 -  48 (  32 -  49)

----- 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>
```

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

```
>GP:BAB06655 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 119/435 (27%), Positives = 214/435 (48%), Gaps = 14/435 (3%)

Query: 9   IFSLALPSMIENILQMLMGMVDNYLVAQIGLVAVSGVSIANNIISIYQSLFIALGAAVSS 68
           +F+L  P  IE +L MLMG  D  +++Q    AV+ V ++N I+++   +F  +    S
Sbjct: 11  LFALTWPIFIEILLHMLMGNADTLMLSQYSDDAVAAVGVSNQILAVIIVMFGFVATGTSI 70

Query: 69  LIARSIGENNQNKQLNYMAGVLQVTLLLSVGLGLLSVAGHHQVLEWLGAEASVTLVGGQY 128
           L+A+ +G   +          +   L+  + LGLL +A    +L+ +  + S+      Y
Sbjct: 71  LVAQHLGAKERENAGKVAVVSIGANLIFGIVLGLLLIAFGPPILKAMQLDDSLLQEATLY 130

Query: 129 LSIVGGMIVSLGLLTSLGAIVRAQGYPKIPMQVSLLINVLNAIFSALSIY----VWGFGL 184
           L IVGG  V   L+ + GAI+R+  + K  M V++ +N+LN I + L I+    +  G+
Sbjct: 131 LQIVGGFSVVQSLIMTAGAILRSHSFTKDVMYVTIGMNILNVIGNYLFIFGPFGIPVLGV 190

Query: 185 LGVAWATVLSRLVGVFLLCQF--------IPIKQVAKRLMRPLDKIIFDLSLPAAGERLM 236
            GVA +TV+SR +G+F++            +P   + KR  R   +  + +P+AGE+L
Sbjct: 191 TGVALSTVVSRTIGLFVIAILLYKRIRGELPFAYLLKRFPRVELRNLLKIGIPSAGEQLS 250

Query: 237 MRAGDVLIIGIVVRFGTTALAGNAIGETLTQFNYMPGLAMATATIILVARQLGGGKVTEI 296
            A  ++I   +  GT AL      + L  F ++  +A+   T IL+  Q+G  ++
Sbjct: 251 YNASQLVITYFIAMMGTEALTTKVYTQNLMMFVFLFAVAIGQGTQILIGHQVGAKQIQAA 310

Query: 297 RYIIREAFILSTLMMLVMGALTYLLGPSLLPLFTQNTDAQRSAMIVLLFSLLGAPATAGT 356
            +  ++ + M  + +     LL +FT N D       +LL +++  P  A
Sbjct: 311 YVRCFRSLWIAMTVSVSMAVVFFAFSTPLLGIFTDNPDILSLGTTLLLLTIILEPGRACN 370

Query: 357 LVYTAVWQGLGKAKLPFYATTIGMWVIRIGLGYVIGVVWQYGLIGVWMATVLDNTSRWFI 416
           LV  +  G  K P Y   + MW I + + Y++G+    GLIGVW+A + D   R  +
Sbjct: 371 LVVISSLRAAGDVKFPVYLAIVSMWGIAVPIAYLLGLPLGLGLIGVWIAFIADEWFRGLL 430

Query: 417 LSKHFK--KYQEITF 429
           +   ++  K+QE++F
Sbjct: 431 MIWRWRKGKWQEMSF 445
```

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

```
Identities = 219/418 (52%), Positives = 316/418 (75%)

Query: 5   KEIIQLAIPAMIENILQMLMGVVDNYLVAQLGVVAVSGVSVANNIITIYQAIFIALGASI 64
           ++I  LA+P+MIENILQMLMG+VDNYLVAQ+G+VAVSGVS+ANNII+IYQ++FIALGA++
Sbjct: 7   RKIFSLALPSMIENILQMLMGMVDNYLVAQIGLVAVSGVSIANNIISIYQSLFIALGAAV 66

Query: 65  ASLLAKSLAGSKKDDAISVCSQAIFLTLLIGAVLGIISIVFGQTFFKLLGTTKSVAQVGG 124
           +SL+A+S+  + ++  ++  +  + +TLL+   LG++S+       + LG   SV VGG
Sbjct: 67  SSLIARSIGENNQNKQLNYMAGVLQVTLLLSVGLGLLSVAGHHQVLEWLGAEASVTLVGG 126

Query: 125 LYLAIVGGGVVTLGMLTTLGSFLRVQGQPRLPMYVSIFVNFLNAVLSGFAIFEWRYGLVG 184
            YL+IVGG +V+LG+LT+LG+ +R QG P++PM VS+ +N LNA+ S  +I+ W +GL+G
Sbjct: 127 QYLSIVGGMIVSLGLLTSLGAIVRAQGYPKIPMQVSLLINVLNAIFSALSIYVWGFGLLG 186

Query: 185 VAVSTLIARLIGICILAKYLPIKKIIKRMTWKISAQIWNLALPSAGERLMMRAGDVVIVA 244
           VA +T+++RL+G+ +L +++PIK++ KR+   +   I++L+LP+AGERLMMRAGDV+I+
Sbjct: 187 VAWATVLSRLVGVFLLCQFIPIKQVAKRLMRPLDKIIFDLSLPAAGERLMMRAGDVLIIG 246

Query: 245 IVVQLGTNVVAGNAIGETLTQFNYMPGLGIATATIILTAKYVGQKNRESIEETIQSSYYI 304
           IVV+ GT  +AGNAIGETLTQFNYMPGL +ATATIIL A+ +G      I   I+ ++ +
Sbjct: 247 IVVRFGTTALAGNAIGETLTQFNYMPGLAMATATIILVARQLGGGKVTEIRYIIREAFIL 306

Query: 305 GLVLMILISSFMLLAGKPLTQLFTNNPSAIKGSLIVILLSFVGVPATIGTLVYTAAWQGL 364
             ++M+++ +   L G  L  LFT N  A + ++IV+L S +G PAT GTLVYTA WQGL
Sbjct: 307 STLMMLVMGALTYLLGPSLLPLFTQNTDAQRSAMIVLLFSLLGAPATAGTLVYTAVWQGL 366

Query: 365 GNAKLPFYTTTIGMWLIRVVLGYLLGIVFELGLLGVWMATIADNIFRWLFLKVHYHRY 422
           G AKLPFY TTIGMW+IR+ LGY++G+V++ GL+GVWMAT+ DN  RW  L  H+ +Y
Sbjct: 367 GKAKLPFYATTIGMWVIRIGLGYVIGVVWQYGLIGVWMATVLDNTSRWFILSKHFKKY 424
 Identities = 48/211 (22%), Positives = 89/211 (41%), Gaps = 29/211 (13%)

Query: 213 MTWKISAQIWNLALPSAGERLMMRAGDVVIVAIVVQLGTNVVAGNAIGETLTQFNYMPGL 272
           M +    +I++LALPS  E ++     +V   +V Q+G   V+G +I   +       +
Sbjct: 1   MIYNNRRKIFSLALPSMIENILQMLMGMVDNYLVAQIGLVAVSGVSIANNIISIYQSLFI 60

Query: 273 GIATATIILTAKYVGQKNRESIEETIQSSYYIGLVLMILISSFML--------------L 318
            + A   L A+ +G+ N+       Q +Y  G++ + L+ S  L              L
Sbjct: 61  ALGAAVSSLIARSIGENNQNK-----QLNYMAGVLQVTLLLSVGLGLLSVAGHHQVLEWL 115

Query: 319 AGKPLTQLFTNNPSAIKGSLIVILLSFVGVPATIGTLVYTAAWQGLGNAKLPFYTTTIGM 378
             +    L      +I G +IV L    G+  ++G  ++G +V     +  G  K+P   + + +
Sbjct: 116 GAEASVTLVGGQYLSIVGGMIVSL----GLLTSLGAIV-----RAQGYPKIPMQVSLL-I 165

Query: 379 WLIRVVLGYLLGIVFELGLLGVWMATIADNI 409
           ++  +  L   V+  GLLGV  AT+   +
Sbjct: 166 NVLNAIFSALSIYVWGFGLLGVAWATVLSRL 196
```

A related GBS gene <SEQ ID 8071> and protein <SEQ ID 8072> were also identified. Analysis of this

```
   INTEGRAL    Likelihood = -2.18   Transmembrane   30 -  46 (  30 -  47)
   PERIPHERAL  Likelihood =  0.32       11
 modified ALOM score:   2.78

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5564(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
ORF01629(313 - 1533 of 1878)
EGAD|165726|TM0815(20 - 436 of 464) conserved hypothetical protein {Thermotoga maritima}
OMNI|TM0815 conserved hypothetical protein GP|4981345|gb|AAD35897.1|AE001748_13|AE001748
conserved hypothetical protein {Thermotoga maritima} PIR|H72331|H72331 conserved
hypothetical protein - Thermotoga maritima (strain MSB8)
%Match = 13.9
%Identity = 29.4  %Similarity = 53.7
Matches = 120  Mismatches = 183  Conservative Sub.s = 99

48        78        108       138       168       198       228       258
YK*RRDTGFRCYFNLKRFVRCFFT*GGYRSTKGRSNP*NGSTYLKYARNG*RVSRFETIIKIRLF*NI*SEKETF*KFSH

                                                                          M

288       318       348       378       408       438       468       498
HSLFNDPG**KGDTVRYSKEIIQLAIPAMIENILQMLMGVVDNYLVAQLGVVAVSGVSVANNIITIYQAIFIALGASIAS
                  ||:|:||:||| ||:|||| |: |  :: :    |:||| ::| :  : | ::||
RYSLFKNYLPKEEVPEIRKELIKLALPAMGENVLQMLFGMADTAFLGHYSWKAMSGVGLSNQVFWVVQVVLIAASMGATV
                 20        30        40        50        60        70        80

528       558       588       609       639       669       699       729
LLAKSLAGSKKDDAISVCSQAIFLTLLIGAVL---GIISIVFGQTFFKLLGTTKSVAQVGGLYLAIVGGGVVTLGMLTTL
 :| ::    :    |:   ::|| :: | :|     :| |:   |  | |  :| |     || ::  | : : ::  :
TIANAIGAGNRKAVRSLAWNSVFLAIFTGVILTALTPLSDVLINIFPNLEGEIESSA---KEYLKVILSGSMGFSIMAVF
                100       110       120       130       140       150

759       789       819       837       867       897       909       939
GSFLRVQGQPRLPMYVSIFVNFLNAVLSGFAIF----EWRYGLVGVAVSTLIARLIGICILA------KYLPIKKIIKRM
 : ||  |  | || |: : ||||  |    ||       |: | ||:|:::|::|  ||       : : ::| :
SAMLRGAGDTRTPMIVTGLTNFLNIFLDYAMIFGKFGFPEMGVRGAAVATILSRFVGAGILTYVIFKREEFQLRKGLVPP
         170       180       190       200       210       220       230

969       999       1029      1059      1089      1119      1149      1179
TWKISAQIWNLALPSAGERLMMRAGDVVIVAIVVQLGTNVVAGNAIGETLTQFNYMPGLGIATATIILTAKYVGQKNRES
 |    :|  : :|:| | ::   | ::   |::  |    ||: ||  :  :::|| :||: |   |  :| |  |:|
KWSSQKEILRVGFPTAIENFVFSTGVLMFANILLIAGAEAYAGHRIGINVESLSFMPAFGISVAITTLVGRYNGMGNKEH
         250       260       270       280       290       300       310

1209      1239      1269      1299      1329      1359      1383      1413
IEETIQSSYYIGLVLMILISSFMLLAGKPLTQLFTNNPSAIKGSLIVILLSFVGVPATIGTLVYT--AAWQGLGNAKLPF
:   |:  : : |:: : :   ::|  :|| ::||::|  |: | : :    :|:      :  |   | :| ||   |
VLGVIRQGWILSLLFQVTVGIIIFLFPEPLIRIFTSDPQIIEISKLPV--KIIGLFQFFLAIDSTMNGALRGTGNTLPPM
         330       340       350       360       370       380       390

1443      1473      1503      1533      1563      1593      1623      1653
YTTTIGMWLIRVVLGYLLGIVFELGLLGVWMATIADNIFRWLFLKVHYHRYIQKM*PEMVAFFSKIIK*CLRVLFFFDII
  | | :|  |: : :::   |:|||||  |:  ||| |||
IITFISIWTARLPVAFVMVKYFQLGLLGAWIGMIADIIFRSTLKLLFFLSGKWEKRAVLTRERVKELG
          410       420       430       440       450       460
```

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

Example 2073

A DNA sequence (GBSx2188) was identified in *S.agalactiae* <SEQ ID 6417> which encodes the amino acid sequence <SEQ ID 6418>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2200(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:AAD05671 GB:AE001448 THREONINE SYNTHASE [Helicobacter pylori
          J99]
 Identities = 161/479 (33%), Positives = 259/479 (53%), Gaps = 17/479 (3%)

Query: 14  KVTASQAILKGLADDGGLFTPITFPKVDLDFTKLKDASYQEVAKLVLSAFFDDFTEQELD 73
           K+   +A+L   A  GGL+T   F    L++      SY E+ + V      +  +  L
Sbjct: 13  KIDFIEAVLNPNAPKGGLYTLEHFET--LEWQDCLGMSYSELVEHVFELLNLEIPKNLLA 70

Query: 74  YCISQAYDTKFDTTEIAPIVKIGDRYHL-ELFHGPTIAFKDMALSILPYLLTTAAKKQGV 132
             + + Y+   +    API  + +R  + EL+HGP++AFKDMAL  L  L +  A   G
Sbjct: 71  SALKR-YENFDNPKNPAPIFALNERLFVQELYHGPSLAFKDMALQPLASLFSNLAV--GK 127

Query: 133 DNKIVILTATSGDTGKAAMAGFADVPGTEIIVFYPKNGVSYIQELQMITQAGQNTHVVAI 192
           + K ++L +TSGDTG A + G A +P   ++  YPK+G S +Q+LQM+TQ   N  V  +
Sbjct: 128 NEKYLVLVSTSGDTGPATLEGLAGMPNVFVVCLYPKDGTSLVQKLQMVTQNASNLKVFGV 187

Query: 193 EGNFDDAQTSVKEMFNNSLLRLKLSQQHMQLSSANSMNIGRLVPQIVYYIYAYAQLVKSK 252
            G+FDDAQ ++K +  +      L  + ++LS ANS+N GR+  QIVY+I+ + +L K
Sbjct: 188 SGDFDDAQNALKNLLKDDDFNEALKARQLKLSVANSVNFGRIAFQIVYHIWGFLELYKKG 247

Query: 253 EISIGQPINFSVPTGNFGNILAAYYASQIGLPVTKLICASNDNNVLTDFFKTQTYD-KNR 311
           I+  + I  ++P+GNFGN L A+YA ++GL + K+    +N N+VL +F +T  YD   R
Sbjct: 248 AINSKEKITLAIPSGNFGNALGAFYAKKMGLNIAKIKVVTNSNDVLREFIETGRYDLTKR 307

Query: 312 EFKVTSSPSMDILVSSNLERLIFHLLGDDAETTKKLMEDLVTTGEYALEARQANIL-ESF 370
             K T SP+MDIL SSN+ER +F L G   E T +LM+ L      YAL+ ++  +L E F
Sbjct: 308 SLKQTFSPAMDILKSSNVERALFSLFG--FERTLELMQALEEEKFYALKPKELALLQEHF 365

Query: 371 VAGFATEQFVELDIKHLFDQYQYIEDPHTAVASAVYQAYQTETKDQTPAVIVSTASPYKF 430
                +++     I+ ++ ++QY+ DPHTA       A      K      ++ +TAS  KF
Sbjct: 366 SCASCSDEDCLKTIQEVYAEHQYLIDPHTAT------ALNASLKTHEKTLVSATASYEKF 419

Query: 431 PCVVTKAIT-NKEEIQDFAAISILNDLSGVSLPKAVTDLQKAEVIHRTVVPTSNMRETV 488
           P     A+   K+   D AA+  L +       + + DL +  + H+ V+  + ++ ++
Sbjct: 420 PKTTLLALNEQKKNDDDKAALETLKNSYNTPDSQRLDDLFERGIKHQEVLKLNEIKSSI 478
```

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 2074

A DNA sequence (GBSx2189) was identified in *S.agalactiae* <SEQ ID 6419> which encodes the amino acid sequence <SEQ ID 6420>. Analysis of this protein sequence reveals the following:

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

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

```
                bacterial cytoplasm --- Certainty=0.3153(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 9279> which encodes amino acid sequence <SEQ ID 9280> was also identified.

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

```
>GP:AAF40975 GB:AE002410 alcohol dehydrogenase, propanol-preferring
           [Neisseria meningitidis MC58]
 Identities = 202/282 (71%), Positives = 228/282 (80%), Gaps = 1/282 (0%)

Query: 1   MGHEGIGIVEEIGEGVTSLRVGDRVSIAWFFEGCGHCEYCTTGRETLCRSVKNAGYSVDG 60
           +GHEGIG+V+E+ +GV +L+VGDRVSIAW F+ CG CEYC TGRETLCRSV NAGY+ DG
Sbjct: 60  LGHEGIGLVKEVADGVKNLKVGDRVSIAWLFQSCGSCEYCNTGRETLCRSVLNAGYTADG 119

Query: 61  GMSEYAIVTADYAVKVPEGLDPAQASSITCAGVTTYKAIKEAGAAPGQWIAVYGAGGLGN 120
           GM+ + IV+ADYAVKVPEGLDPAQASSITCAGVTTYKAIK +G  PGQWIA+YGAGGLGN
Sbjct: 120 GMATHCIVSADYAVKVPEGLDPAQASSITCAGVTTYKAIKVSGVRPGQWIAIYGAGGLGN 179

Query: 121 LAVQYAKKVFNAHVVAVDINADKLQLAKEVGADLTVNGKEIKDVAAYIQEKTGGCHGVVV 180
           L VQYAKKVF AHVVA+DIN DKL  AKE GADL VN  + +D A  IQEKTGG H  VV
Sbjct: 180 LGVQYAKKVFGAHVVAIDINDDKLAFAKETGADLVVNAAK-EDAAKVIQEKTGGAHAAVV 238

Query: 181 TAVSKVAFNQAIDSVRAGGTVVAVGLPSEYMELSIVKTVLDGIRVVGSLVGTRKDLEEAF 240
           TAVS  AFN A++ VRAGG VVA+GLP E M+LSI + VLDGI VVGSLVGTRKDLEEAF
Sbjct: 239 TAVSAAAFNSAVNCVRAGGRVVAIGLPPESMDLSIPRLVLDGIEVVGSLVGTRKDLEEAF 298

Query: 241 AFGAEGLVVPVVEKVPVDTAPQVFDEMERGLIQGRKVLDFTK 282
            FGAEGLVVP V+    +D AP +F EM  G I GR V+D  K
Sbjct: 299 QFGAEGLVVPKVQLRALDEAPAIFQEMREGKITGRMVIDMKK 340
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6421> which encodes the amino acid sequence <SEQ ID 6422>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2356(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 = 263/280 (93%), Positives = 273/280 (96%)

Query: 1   MGHEGIGIVEEIGEGVTSLRVGDRVSIAWFFEGCGHCEYCTTGRETLCRSVKNAGYSVDG 60
           +GHEGIGIVEEIGEGVTSL+VGDRVSIAWFFEGCGHCEYCTTGRETLCRSVKNAGYSVDG
Sbjct: 76  LGHEGIGIVEEIGEGVTSLKVGDRVSIAWFFEGCGHCEYCTTGRETLCRSVKNAGYSVDG 135

Query: 61  GMSEYAIVTADYAVKVPEGLDPAQASSITCAGVTTYKAIKEAGAAPGQWIAVYGAGGLGN 120
           GMSEYA+VTADYAVKVPEGLDPAQASSITCAGVTTYKAIKEAGAAPGQWI ++GAGGLGN
Sbjct: 136 GMSEYAVVTADYAVKVPEGLDPAQASSITCAGVTTYKAIKEAGAAPGQWIVIFGAGGLGN 195

Query: 121 LAVQYAKKVFNAHVVAVDINADKLQLAKEVGADLTVNGKEIKDVAAYIQEKTGGCHGVVV 180
           LAVQYAKKVFNAHVVAVDIN DKL+LAKEVGAD+ VNGKEI+DV  YIQEKTGG HGVVV
Sbjct: 196 LAVQYAKKVFNAHVVAVDINNDKLELAKEVGADILVNGKEIEDVPGYIQEKTGGAHGVVV 255

Query: 181 TAVSKVAFNQAIDSVRAGGTVVAVGLPSEYMELSIVKTVLDGIRVVGSLVGTRKDLEEAF 240
           TAVSKVAFNQAIDSVRAGGTVVAVGLPSEYMELSIVKTVLDGI+VVGSLVGTRKDLEEAF
Sbjct: 256 TAVSKVAFNQAIDSVRAGGTVVAVGLPSEYMELSIVKTVLDGIKVVGSLVGTRKDLEEAF 315

Query: 241 AFGAEGLVVPVVEKVPVDTAPQVFDEMERGLIQGRKVLDF 280
           AFGAEGLV PVVEKVPVDTAP+VFDEMERGLIQGRKVLDF
```

```
Sbjct: 316 AFGAEGLVAPVVEKVPVDTAPEVFDEMERGLIQGRKVLDF 355
```

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

Example 2075

A DNA sequence (GBSx2190) was identified in *S.agalactiae* <SEQ ID 6423> which encodes the amino acid sequence <SEQ ID 6424>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.82   Transmembrane   83 -  99 (  76 - 108)
   INTEGRAL    Likelihood = -7.27   Transmembrane   46 -  62 (  43 -  65)
   INTEGRAL    Likelihood = -7.22   Transmembrane  187 - 203 ( 182 - 209)
   INTEGRAL    Likelihood = -6.00   Transmembrane  243 - 259 ( 229 - 262)
   INTEGRAL    Likelihood = -4.25   Transmembrane  404 - 420 ( 402 - 422)
   INTEGRAL    Likelihood = -3.98   Transmembrane  120 - 136 ( 119 - 136)
   INTEGRAL    Likelihood = -3.88   Transmembrane  308 - 324 ( 307 - 324)
   INTEGRAL    Likelihood = -2.13   Transmembrane  378 - 394 ( 376 - 394)
   INTEGRAL    Likelihood = -1.38   Transmembrane  152 - 168 ( 152 - 168)
   INTEGRAL    Likelihood = -1.17   Transmembrane  271 - 287 ( 271 - 287)

----- Final Results -----
               bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9371> which encodes amino acid sequence <SEQ ID 9372> was also identified.

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

```
>GP:AAC17857 GB:AF026147 YojI [Bacillus subtilis]
 Identities = 183/432 (42%), Positives = 266/432 (61%), Gaps = 1/432 (0%)

Query: 1   MKLFIPVLIYQFANFSATFIDSVMTGQYSQLHLAGVSTASNLWTPFFALLVGMISALVPV 60
           + + IP+ I Q      TF+D+VM+G+ S   LAGV+  S+LWTP +  L G++ A+ P+
Sbjct: 15  LHILIPIFITQAGLSLITFLDTVMSGKVSPADLAGVAIGSSLWTPVYTGLAGILMAVTPI 74

Query: 61  VGQHLGRGNKEQIRTEFHQFLYLGLILSLILFLIMQFIAQPVLGSLGLEDEVLAVGRGYL 120
           V Q LG   K++I     Q +Y+  +LS+ + +I       +LG L L+  V  + + +L
Sbjct: 75  VAQLLGAEKKQKIPFTVLQAVYVAALLSIAVLVIGYAAVDLILGRLNLDIHVHQIAKHFL 134

Query: 121 NYMLIGIMPLVLFSICRSFFDALGLTRLSMYLMLLILPFNSFFNYMLIYGKFGMPRLGGA 180
            ++ +GI PL ++++ RSF D+LG TR++M + L  LP N   NY+ I+GKFGMP LGG
Sbjct: 135 GFLSLGIFPLFVYTVLRSFIDSLGKTRVTMMITLSSLPINFVLNYVFIFGKFGMPALGGV 194

Query: 181 GAGLGTSLTYWAIFIVIIIVMSLHPQIKTYHIW-TLERIKAPLIIEDIRLGLPIGLQIFA 239
           GAGL ++LTYW I I+   ++  +     Y I+ T+ +         +++GLPIG  +F
Sbjct: 195 GAGLASALTYWCICIISFFIIHKNAPFSEYGIFLTMYKFSWKACKNLLKIGLPIGFAVFF 254

Query: 240 EVAIFAVVGLFMAKFSSIIIAAHQAAMNFSSLMYAFPLSISTALAITISFEVGAERFQDA 299
           E +IFA V L M+ F ++ IA+HQAAMNF+SL+Y  PLS+S AL I + FE GA RF+DA
Sbjct: 255 ETSIFAAVTLLMSHFHTVTIASHQAAMNFASLLYMLPLSVSMALTIVVGFEAGAARFKDA 314

Query: 300 NTYSRIGRLTAVGITSGTLLFLFLFRENVAAMYNSDPHFVAITAQFLTYSLFFQFADAYA 359
            +YS IG + A+G +  T   + LFRE +A MY SDP  + +T  FL Y+LFFQ +DA A
Sbjct: 315 RSYSLIGIMMAIGFSLFTAACILLFREQIAGMYTSDPDVLRLTQHFLIYALFFQLSDAVA 374

Query: 360 APVQGILRGYKDTTKPFMIGAGSYWLCALPLAVILEKNSQLGPFAYWIGLITGIFVCGLF 419
           AP+QG LRGYKD          SYW+  LP+  ++   + LG F YWIGLI G+    +
Sbjct: 375 APIQGALRGYKDVNYTLAAAFVSYWVIGLPVGYMVGTFTSLGAFGYWIGLIAGLAAGAVG 434

Query: 420 LNQRLQKIKKLY 431
```

```
          L  RL K++K Y
Sbjct: 435 LFFRLAKLQKRY 446
```

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

Example 2076

A DNA sequence (GBSx2191) was identified in *S.agalactiae* <SEQ ID 6425> which encodes the amino acid sequence <SEQ ID 6426>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60   Transmembrane   23 -  39 (  23 -  39)

----- 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 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 2077

A DNA sequence (GBSx2192) was identified in *S.agalactiae* <SEQ ID 6427> which encodes the amino acid sequence <SEQ ID 6428>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3829(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:AAC06891 GB:AE000703 hypothetical protein [Aquifex aeolicus]
 Identities = 72/213 (33%), Positives = 115/213 (53%), Gaps = 11/213 (5%)

Query: 36  RPKILMHVCCAPCSTYTLEYLSQ---WADVTIYFANSNIHPKDEYYRREYVTQKFVHDFN 92
           + KIL+H+CCAP + Y L+ L +    +++  YF + NIHP +EY  R   T++   +
Sbjct: 3   KSKILVHICCAPDAIYFLKKLREDYPESEIIGYFYDPNIHPYEEYRLRYLETERICKELG 62

Query: 93  KNTGYSVQFLSAPYEPNEFFKIVHGLEEEPEGGDRCKVCYDFRLDKTAEKAVELGFDYFG 152
            N       +   Y+   + + V G E+EPE G RC++C+D+RL+K+AE A ELG D
Sbjct: 63  IN------LIEGEYDLENWLERVKGYEDEPERGKRCQICFDYRLEKSAEVAKELGCDALT 116

Query: 153 SALTISPHKNSQTINTIGIDVQKIYDTQYLPSDLKKNKGYQRSVEMCKDYDIYRQCYCGC 212
           + L +SP K+    +   G +  K    ++L  D +K  G Q   ++ K +IY+Q YCGC
Sbjct: 117 TTLLMSPKKSIPQLKKAGEEATKRTGIEFLAPDYRKGGGTQEMFKLSKEREIYQQDYCGC 176

Query: 213 IFGAKDQGINLLQIKKDAKAFVSDKDGKEEFPN 245
           I+G   Q     +I  D   F+  + G +E  N
Sbjct: 177 IYGLFKQKNG--KIFWDLVGFLGRRPGSKEERN 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6429> which encodes the amino acid sequence <SEQ ID 6430>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3498(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 254-256
```

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

```
>GP:AAC06891 GB:AE000703 hypothetical protein [Aquifex aeolicus]
 Identities = 65/182 (35%), Positives = 106/182 (57%), Gaps = 9/182 (4%)

Query: 39  RPSILMHVCCAPCSTYTLEYLTQF---ADITVYFANSNIHPKDEYHRRAYVTQQFVSEFN 95
           +  IL+H+CCAP + Y L+ L +     ++I  YF + NIHP +EY  R   T++   E
Sbjct: 3   KSKILVHICCAPDAIYFLKKLREDYPESEIIGYFYDPNIHPYEEYRLRYLETERICKELG 62

Query: 96  AKTGNTVQFLEADYVPNEYVRQVRGLEEEPEGGDRCRVCFDYRLDKTAQKAVELGFDYFA 155
                 +  +E +Y    ++ +V+G E+EPE G RC++CFDYRL+K+A+ A ELG D
Sbjct: 63  ------INLIEGEYDLENWLERVKGYEDEPERGKRCQICFDYRLEKSAEVAKELGCDALT 116

Query: 156 SALTISPHKNSQTINDVGIDVQKVYTTKYLPSDFKKNNGYRRSVEMCEEYDIYRQCYCGC 215
           + L +SP K+  +   G +  K    ++L  D++K  G +  ++ +E +IY+Q YCGC
Sbjct: 117 TTLLMSPKKSIPQLKKAGEEATKRTGIEFLAPDYRKGGGTQEMFKLSKEREIYQQDYCGC 176

Query: 216 VY 217
           +Y
Sbjct: 177 IY 178
```

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

```
 Identities = 184/255 (72%), Positives = 219/255 (85%)

Query: 1   MIDVENILEKMKPNQKINYDWVMQQMVKQWQASDIRPKILMHVCCAPCSTYTLEYLSQWA 60
           MID++ IL  M PNQKINYD VMQQM K W+   +RP ILMHVCCAPCSTYTLEYL+Q+A
Sbjct: 4   MIDLQEILANMNPNQKINYDRVMQQMAKVWEKESVRPSILMHVCCAPCSTYTLEYLTQFA 63

Query: 61  DVTIYFANSNIHPKDEYYRREYVTQKFVHDFNKNTGYSVQFLSAPYEPNEFFKIVHGLEE 120
           D+T+YFANSNIHPKDEY+RR YVTQ+FV +FN  TG +VQFL A Y PNE+ + V GLEE
Sbjct: 64  DITVYFANSNIHPKDEYHRRAYVTQQFVSEFNAKTGNTVQFLEADYVPNEYVRQVRGLEE 123

Query: 121 EPEGGDRCKVCYDFRLDKTAEKAVELGFDYFGSALTISPHKNSQTINTIGIDVQKIYDTQ 180
           EPEGGDRC+VC+D+RLDKTA+KAVELGFDYF SALTISPHKNSQTIN +GIDVQK+Y T+
Sbjct: 124 EPEGGDRCRVCFDYRLDKTAQKAVELGFDYFASALTISPHKNSQTINDVGIDVQKVYTTK 183

Query: 181 YLPSDLKKNKGYQRSVEMCKDYDIYRQCYCGCIFGAKDQGINLLQIKKDAKAFVSDKDGK 240
           YLPSD KKN GY+RSVEMC++YDIYRQCYCGC++ AK QGI+L+Q+KKDAKAF++DKD
Sbjct: 184 YLPSDFKKNNGYRRSVEMCEEYDIYRQCYCGCVYAAKMQGIDLVQVKKDAKAFMADKDLD 243

Query: 241 EEFPNIRFTFNGKSM 255
            +F +IRF++ G  M
Sbjct: 244 NDFTHIRFSYRGDEM 258
```

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

Example 2078

A DNA sequence (GBSx2193) was identified in *S.agalactiae* <SEQ ID 6431> which encodes the amino acid sequence <SEQ ID 6432>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4216(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:CAB14809 GB:Z99118 excinuclease ABC (subunit C) [Bacillus subtilis]
 Identities = 189/333 (56%), Positives = 244/333 (72%)

Query: 1   MNELIKHKLELLPDSPGCYLHKDKNGTIIYVGKAKNLKNRVKSYFHGSHNTKTELLVSEI 60
           MN+ +K KL LLPD PGCYL KD+  T+IYVGKAK LKNRV+SYF GSH+ KT+ LV+EI
Sbjct: 1   MNKQLKEKLALLPDQPGCYLMKDRQQTVIYVGKAKVLKNRVRSYFTGSHDAKTQRLVTEI 60

Query: 61  EDFEYIVTTSNTEALLLEINLIQENMPKYNIRLKDDKSYPYIKITNERYPRLMITRQVKK 120
           EDFEYIVT+SN EAL+LE+NLI+++ PKYN+ LKDDK+YP+IK+T+ER+PRL++TR VKK
Sbjct: 61  EDFEYIVTSSNLEALILEMNLIKKHDPKYNVMLKDDKTYPFIKLTHERHPRLIVTRNVKK 120

Query: 121 SDGTYFGPYPDSGAATEIKRLLDRLFPFKKCTNPANKVCFYYHLGQCNAHTVCQTNKAYW 180
             G YFGPYP+  AA E K+LLDRL+P +KC+   ++VC YYHLGQC A  V   ++
Sbjct: 121 DKGRYFGPYPNVQAARETKKLLDRLYPLRKCSKLPDRVCLYYHLGQCLAPCVKDISEETN 180

Query: 181 DSLREDVKQFLNGKDNKIVNGLTEKMKSAAMTMEFERAAEYRDLIEAISLLRTKQRVIHQ 240
             L E + +FL G  N++   L EKM  AA  +EFERA E RD I  I     KQ++
Sbjct: 181 RELVESITRFLRGGYNEVKKELEEKMHEAAENLEFERAKELRDQIAHIESTMEKQKMTMN 240

Query: 241 DMKDRDVFGYFVDKGWMCVQVFFVRNGKLIQRDVNMFPYYNEPEEDFLTYIGQFYQDTKH 300
           D+ DRDVF Y  DKGWMCVQVFF+R GKLI+RDV+MFP Y E +E+FLT+IGQFY    H
Sbjct: 241 DLVDRDVFAYAYDKGWMCVQVFFIRQGKLIERDVSMFPLYQEADEEFLTFIGQFYSKNNH 300

Query: 301 FLPKEVFIPQDIDAKSVETIVGCKIVKPQRGKR 333
           FLPKE+ +P  ID   +E ++   + +P++G +
Sbjct: 301 FLPKEILVPDSIDQSMIEQLLETNVHQPKKGPK 333
```

There is also homology to SEQ ID 2568.

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

Example 2079

A DNA sequence (GBSx2194) was identified in *S.agalactiae* <SEQ ID 6433> which encodes the amino acid sequence <SEQ ID 6434>. This protein is predicted to be maltose operon transcriptional repressor (rbsR). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3761(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 9393> which encodes amino acid sequence <SEQ ID 9394> was also identified.

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

```
>GP:AAD02112 GB:AF039082 putative maltose operon transcriptional
            repressor [Lactococcus lactis]
 Identities = 64/166 (38%), Positives = 105/166 (62%), Gaps = 13/166 (7%)
```

```
Query: 1   MGKSAIDYLYKKGHKSIQFVTDDLNSEVSEERYLGYFKGARKLGLNQKPALLFDRGNPQV 60
           +G+ A+  L +  H++I FVTD   +EV EERY G+   A +LGL+    LLF   N  +
Sbjct: 169 LGREAVRLLAQLNHQNISFVTDTKETEVFEERYQGFKDEAERLGLSHD--LLFMDSNFSL 226

Query: 61  LEEFINRVKEEETTALIVIGDTVSVRVMQFLSFYKLKVPDDISIMTFNNSLFSHLIHPYL 120
             E          TAL+V+ D +S++V++ L    L VP+D+S++T+NNS+F  +IHPYL
Sbjct: 227 RNE----------TALVVMDDVLSLKVVERLRSQGLNVPEDVSLITYNNSIFGAMIHPYL 276

Query: 121 STFDINVNNLGRTSVRRLIDIIKSPDKVFSETIIVPFTLEERESVR 166
           +TFDI++  LG +++++++D+  + + +  +TII PF L  RES +
Sbjct: 277 TTFDIHIEQLGASAIKKILDLRDNKENLPEKTII-PFELIVRESTK 321
```

There is also homology to SEQ ID 5082.

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

Example 2080

A DNA sequence (GBSx2195) was identified in *S.agalactiae* <SEQ ID 6435> which encodes the amino acid sequence <SEQ ID 6436>. This protein is predicted to be 4-alpha-glucanotransferase (malQ). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2003(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:AAA26923 GB:J01796 amylomaltase [Streptococcus pneumoniae]
 Identities = 250/500 (50%), Positives = 329/500 (65%), Gaps = 4/500 (0%)

Query: 1   MKKRASGVLMHITSLPGDLGIGTFGREAYAFVDFLVETDQKFWQILPLTTTSFGDSPYQS 60
           MKKR SGVLMHI+SLPG  GIG+FG+ AY FVDFLV T Q++WQILPL  TS+GDSPYQS
Sbjct: 1   MKKRQSGVLMHISSLPGAYGIGSFGQSAYDFVDFLVRTKQRYWQILPLGATSYGDSPYQS 60

Query: 61  FSAVAGNTHLIDFDLLTLEGFISKDDYQNISFGQDPEVVDYAGLFEKRRPVLEKAVKNFL 120
           FSA AGNTH ID D+L  +G +   D + + FG D   VDYA ++  RRP+LEKAVK F
Sbjct: 61  FSAFAGNTHFIDLDILVEQGLLEASDLEGVDFGSDASEVDYAKIYYARRPLLEKAVKRFF 120

Query: 121 QEERATRMLSDFLQE-EKWVTDFAEFMAIKEHFGNKALQEWDDKAIIRREEEALAGYRQK 179
            E    +    F Q+ + W+  FAE+MAIKE+F N A  EW D     R+  AL  YR++
Sbjct: 121 -EVGDVKDFEKFAQDNQSWLELFAEYMAIKEYFDNLAWTEWPDADARARKASALESYREQ 179

Query: 180 LSEVIKYHEVTQYFFYKQWFELKEYANDKGIQIIGDMPIYVSADSVEVWTMPELFKLDRD 239
           L++ + YH VTQYFF++QW +LK YAND  I+I+GDMPIYV+ DS ++W  P LFK D +
Sbjct: 180 LADKLVYHRVTQYFFFQQWLKLKAYANDNHIEIVGDMPIYVAEDSSDMWANPHLFKTDVN 239

Query: 240 KQPLAIAGVPADDFSDDGQLWGNPIYNWDYHKESDFDWWIYRIQSGVKMYDYLRIDHFKG 299
            +   IAG P D+FS  GQLWGNPIY+W+   +  + WWI R++   K+YD +RIDHF+G
Sbjct: 240 GKATCIAGCPPDEFSVTGQLWGNPIYDWEAMDKDGYKWWIERLRESFKIYDIVRIDHFRG 299

Query: 300 FSDYWEIRGDYQTANDGSWQPAPGPELFATIKEKLGDLPIIAENLGYIDERAERLLAGTG 359
           F  YWEI     TA  G W   PG +LFA +KE+LG+L IIAE+LG++ +    L   TG
Sbjct: 300 FESYWEIPAGSDTAAPGEWVKGPGYKLFAAVKEELGELNIIAEDLGFMTDEVIELRERTG 359

Query: 360 FPGMKIMEFGFYDTTGNSIDIPHNYTENTIAYAGTHDNEVINGWFEN-LTVEQKAYAENY 418
           FPGMKI++F F +     SID PH     N++ Y GTHDN  + GW+ N +     + Y
Sbjct: 360 FPGMKILQFAF-NPEDESIDSPHLAPANSVMYTGTHDNNTVLGWYRNEIDDATREYMARY 418
```

```
Query: 419 MRRLPNEPITETVLRTLYATVSQTTITCMQDLLDKPADSRMNMPNTVGGNWQWRMRKEDL 478
           R   E +   +LRT++++VS   I  MQDLL+    +RMN P+T+GGNW WRM ++ L
Sbjct: 419 TNRKEYETVVHAMLRTVFSSVSFMAIATMQDLLELDEAARMNFPSTLGGNWSWRMTEDQL 478

Query: 479 TENRKAFLKEITTIYNRGNK 498
           T   +  L ++TTIY R N+
Sbjct: 479 TPAVEEGLLDLTTIYRRINE 498
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6437> which encodes the amino acid sequence <SEQ ID 6438>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.85   Transmembrane  435 - 451 ( 435 - 451)

----- 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>
```

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

```
 Identities = 313/495 (63%), Positives = 387/495 (77%)

Query: 1   MKKRASGVLMHITSLPGDLGIGTFGREAYAFVDFLVETDQKFWQILPLTTTSFGDSPYQS 60
           M KRASG+LMHI+SLPG  GIGTFG+ A+ FVDFL ET Q +WQILPLTTTSFGDSPYQS
Sbjct: 1   MNKRASGILMHISSLPGKFGIGTFGKSAFEFVDFLAETKQTYWQILPLTTTSFGDSPYQS 60

Query: 61  FSAVAGNTHLIDFDLLTLEGFISKDDYQNISFGQDPEVVDYAGLFEKRRPVLEKAVKNFL 120
           FSA+AGNTH IDF+LL  +  +   D  +I+FG +PE VDYA LF+ RRP+LEKAV+ F+
Sbjct: 61  FSAIAGNTHFIDFELLVDDELLEAADLCDITFGTNPEAVDYAQLFQVRRPLLEKAVRAFV 120

Query: 121 QEERATRMLSDFLQEEKWVTDFAEFMAIKEHFGNKALQEWDDKAIIRREEEALAGYRQKL 180
            E+     L  F     W+TDFAEFMA+KE+F NKALQ+WDD+ +I+R+E++L  YR+ L
Sbjct: 121 AEQENVCKLEAFETASSWLTDFAEFMALKEYFNNKALQDWDDETVIKRQEDSLNNYRELL 180

Query: 181 SEVIKYHEVTQYFFYKQWFELKEYANDKGIQIIGDMPIYVSADSVEVWTMPELFKLDRDK 240
           ++ I YH+V QYFFY+QW  LK YAN KGI+IIGDMPIYVSADSVEVWTMPELFK+D DK
Sbjct: 181 AKKITYHKVCQYFFYQQWSALKTYANHKGIEIIGDMPIYVSADSVEVWTMPELFKVDSDK 240

Query: 241 QPLAIAGVPADDFSDDGQLWGNPIYNWDYHKESDFDWWIYRIQSGVKMYDYLRIDHFKGF 300
           +PL IAGVPAD FS+DGQLWGNP YNW  H++S+F WWIYRIQ   K+YD LRIDHFKGF
Sbjct: 241 KPLFIAGVPADGFSEDGQLWGNPTYNWSAHEKSNFAWWIYRIQESFKLYDQLRIDHFKGF 300

Query: 301 SDYWEIRGDYQTANDGSWQPAPGPELFATIKEKLGDLPIIAENLGYIDERAERLLAGTGF 360
           SD+WEI     +TA +G W  APG  LF+ ++E LG+LPIIAENLGYIDE+AE+LLA TGF
Sbjct: 301 SDFWEIPAGDKTARNGHWASAPGIALFSAVREALGELPIIAENLGYIDEKAEQLLASTGF 360

Query: 361 PGMKIMEFGFYDTTGNSIDIPHNYTENTIAYAGTHDNEVINGWFENLTVEQKAYAENYMR 420
           PGMKI+EFG +D T  SID+PH Y  N +AY GTHDNEV+NGW++NL+ EQ  +  NY+
Sbjct: 361 PGMKILEFGLFDITSQSIDLPHYYDRNCVAYTGTHDNEVVNGWYDNLSEEQVHFVNNYLH 420

Query: 421 RLPNEPITETVLRTLYATVSQTTITCMQDLLDKPADSRMNMPNTVGGNWQWRMRKEDLTE 480
           +  +E IT+ +LRT++A+V  T I C+QDLLDK    SRMNMPNT+GGNWQWRM   +L +
Sbjct: 421 KHADESITKAMLRTIFASVCDTAILCIQDLLDKDGKSRMNMPNTIGGNWQWRMLDGELNQ 480

Query: 481 NRKAFLKEITTIYNR 495
           + K +L  +T +Y R
Sbjct: 481 DHKDYLIYLTDLYGR 495
```

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

Example 2081

A DNA sequence (GBSx2196) was identified in *S.agalactiae* <SEQ ID 6439> which encodes the amino acid sequence <SEQ ID 6440>. This protein is predicted to be glycogen phosphorylase (malP). Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2678(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:AAC00218 GB:AF008220 glycogen phosphorylase [Bacillus subtilis]
 Identities = 297/776 (38%), Positives = 452/776 (57%), Gaps = 41/776 (5%)

Query: 13  GKVLSELTNEEIYVELLNFVKEEAAA--------KSKNSSQRKVYYISAEFLIGKLLSNNL 65
           GK   +    + Y  L N V+E  +A        KS+++S ++ YY+S EFL+G+LL  NL
Sbjct: 21  GKSFKDSAKLDQYKTLGNMVREYISADWIETNEKSRSNSGKQTYYLSIEFLLGQLLEQNL 80

Query: 66  INLGIYKDVKKELELVGKSIAEIEDVEPEPSLGNGGLGRLASCFIDSISSLGINGEGVGL 125
           +NLG+   V+  L+ +G ++ EI  +E +  LGNGGLGRLA+CF+DS++SL + G G+G+
Sbjct: 81  MNLGVRDVVEAGLKEIGINLEEILQIENDAGLGNGGLGRLAACFLDSLASLNLPGHGMGI 140

Query: 126 NYHCGLFKQVFRNNQQEAEANYWIEN-NSWLVPT-DISYDVPF--------RDFTLKSRL 175
            Y  GLF+Q   + Q      W++N N W V   D + DVPF        +   L  R
Sbjct: 141 RYKHGLFEQKIVDGHQVELPEQWLKNGNVWEVRNADQAVDVPFWGEVHMTEKSGRLHFRH 200

Query: 176 DR----------IDVLGYKKDTKNYLNLFDIDGLDYNLIEKGITFDKTEIKKNLTLFLYP 225
           ++          I ++GY+  T N L L++ +   Y     G          + ++ FLYP
Sbjct: 201 EQATIVTAVPYDIPIIGYETGTVNTLRLWNAE--PYAHYHGGNILSYKRETEAVSEFLYP 258

Query: 226 DDSDKNGELLRIYQQYFMVSNAAQLLIDEAIERGSNLHDLAEYAYVQINDTHPSMVIPEL 285
           DD+   G++LR+ QQYF+V  + + +++   +   +L  L +   + INDTHP++ +PEL
Sbjct: 259 DDTHDEGKILRLKQQYFLVCASLKSIVNNYRKTHKSLSGLHKKVSIHINDTHPALAVPEL 318

Query: 286 IRLLTEKHGFEFDEAVSVVRNMVGYTNHTILAEALEKWPLEYLNEVVPHLVTIIKKLDQM 345
           +R+L ++    ++EA  +  + + YTNHT L+EALEKWP+     ++P +  II+++++
Sbjct: 319 MRILLDEENMSWEEAWHITVHTISYTNHTTLSEALEKWPIHLFKPLLPRMYMIIEEINER 378

Query: 346 IRE--------EQTNPEVQIIDEAGRVHMAHMDIHFSTSVNGVAALHTEILKNSELKVFY 397
                    +   E   I   G V MAH+ I  S SVNGVA +H++ILK  E++ F+
Sbjct: 379 FCRAVWEKYPGDWKRIENMAITAHGVVKMAHLAIVGSYSVNGVAKIHSDILKEREMRDFH 438

Query: 398 DIYPDKFNNKTNGITFRRWLEFANQDLADYLKELIGDSYLTDATQLEKLLTYADSNEVHD 457
            ++P++FNNKTNGI  RRWL  AN  L+  + E IGD ++     L +L  YA      +
Sbjct: 439 LLFPNRFNNKTNGIAHRRWLLKANPGLSAIITEAIGDEWVKQPESLIRLEPYATDPAFIE 498

Query: 458 KLAAIKFKNKLALKRYLKENKGIELDEYSIIDTQIKRFHEYKRQQMNALYVIHKYLEIKR 517
           +   K K K  L   +   G+ ++  SI D Q+KR H YKRQ +N L++++ Y  +K
Sbjct: 499 QFQNNKSKKKQELADLIFCTAGVVVNPESIFDVQVKRLHAYKRQLLNVLHIMYLYNRLKE 558

Query: 518 GH-FPSRKLTVIFGGKAAPAYTIAQDIIHLILCLSELINNDPEVNKYLNVHLVENYNVTV 576
              F     T IFG KA+P+Y  A+ II LI  ++E +N DP V + + V  +ENY V++
Sbjct: 559 DSGFSIYPQTFIFGAKASPSYYYAKKIIKLIHSVAEKVNYDPAVKQLIKVVFLENYRVSM 618

Query: 577 AEKLIPATDISEQISLASKEASGTGNMKFMLNGALTLGTMDGANVEIAELAGKENIYTFG 636
           AE++ PA+D+SEQIS ASKEASGTGNMKFM+NGALT+GT DGAN+EI E  G + IYTFG
Sbjct: 619 AERIFPASDVSEQISTASKEASGTGNMKFMMNGALTIGTHDGANIEILERVGPDCIYTFG 678

Query: 637 KDSDTIINLYETSGYRSKDYYDKDKVIREAVDFIISDDIVSLGNAERLKRLHDELV-GKD 695
            +D +++   E  GYRS++YY  D+ IR+  D +I+      G A+  + + D L+  D
Sbjct: 679 LKADEVLSYQENGGYRSREYYQHDRRIRQVADQLINGFFE--GEADEFESIFDSLLPHND 736
```

```
Query: 696 WFMTLIDLKEYIAVKEQVLADYEDYESWNKKVIHNIAKAGFFSSDRTIEQYNQDIW 751
           +  L D   Y   +E++ ADY +   W++  I NIA +G+FSSDRTI +Y +DIW
Sbjct: 737 EYFVLKDFSSYADAQERIQADYRERRKWSEHSIVNIAHSGYFSSDRTIREYAKDIW 792
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6441> which encodes the amino acid sequence <SEQ ID 6442>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.71   Transmembrane  538 - 554 ( 538 - 554)

----- Final Results -----
               bacterial membrane --- Certainty=0.2084(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 = 629/754 (83%), Positives = 696/754 (91%), Gaps = 2/754 (0%)

Query: 1   MTRNFTTYVGQQ-GKVLSELTNEEIYVELLNFVKEEAAAKSKNSSQRKVYYISAEFLIGK 59
           MTR FT YV  + GK L++ +NEEIY+ LLNFVKEEA+ K+KNS++RKVYYISAEFLIGK
Sbjct: 1   MTR-FTEYVETKLGKSLTQASNEEIYLSLLNFVKEEASHKAKNSAKRKVYYISAEFLIGK 59

Query: 60  LLSNNLINLGIYKDVKKELELVGKSIAEIEDVEPEPSLGNGGLGRLASCFIDSISSLGIN 119
           LLSNNLINLGIYKD+K+EL   GKSIAE+EDVE EPSLGNGGLGRLASCFIDSI+SLGIN
Sbjct: 60  LLSNNLINLGIYKDIKEELAAAGKSIAEVEDVELEPSLGNGGLGRLASCFIDSIASLGIN 119

Query: 120 GEGVGLNYHCGLFKQVFRNNQQEAEANYWIENNSWLVPTDISYDVPFRDFTLKSRLDRID 179
           GEGVGLNYHCGLFKQVF++N+QEAE N+WIE++SWLVPTDISYDVPF++FTLKSRLDRID
Sbjct: 120 GEGVGLNYHCGLFKQVFKHNEQEAEPNFWIEDDSWLVPTDISYDVPFKNFTLKSRLDRID 179

Query: 180 VLGYKKDTKNYLNLFDIDGLDYNLIEKGITFDKTEIKKNLTLFLYPDDSDKNGELLRIYQ 239
           VLGYK+DTKNYLNLFDI+G+DY LI+ GI+FDKT+I KNLTLFLYPDDSDKNGELLRIYQ
Sbjct: 180 VLGYKRDTKNYLNLFDIEGVDYGLIKDGISFDKTQIAKNLTLFLYPDDSDKNGELLRIYQ 239

Query: 240 QYFMVSNAAQLLIDEAIERGSNLHDLAEYAYVQINDTHPSMVIPELIRLLTEKHGFEFDE 299
           QYFMVSNAAQL+IDEAIERGSNLHDLA+YAYVQINDTHPSMVIPELIRLLTEKHGF+FDE
Sbjct: 240 QYFMVSNAAQLIIDEAIERGSNLHDLADYAYVQINDTHPSMVIPELIRLLTEKHGFDFDE 299

Query: 300 AVSVVRNMVGYTNHTILAEALEKWPLEYLNEVVPHLVTIIKKLDQMIREEQTNPEVQIID 359
           AV+VV+NMVGYTNHTILAEALEKWP  YLNEVVPHLVTII+KLD ++R E ++P VQIID
Sbjct: 300 AVAVVKNMVGYTNHTILAEALEKWPTAYLNEVVPHLVTIIEKLDALVRSEVSDPAVQIID 359

Query: 360 EAGRVHMAHMDIHFSTSVNGVAALHTEILKNSELKVFYDIYPDKFNNKTNGITFRRWLEF 419
           E+GRVHMAHMDIHF+TSVNGVAALHTEILKNSELK FYD+YP+KFNNKTNGITFRRWLEF
Sbjct: 360 ESGRVHMAHMDIHFATSVNGVAALHTEILKNSELKAFYDLYPEKFNNKTNGITFRRWLEF 419

Query: 420 ANQDLADYLKELIGDSYLTDATQLEKLLTYADSNEVHDKLAAIKFKNKLALKRYLKENKG 479
           ANQDLADY+KELIGD YLTDAT+LEKL+ +AD   VH KLA IKF NKLALKRYLK+NK
Sbjct: 420 ANQDLADYIKELIGDEYLTDATKLEKLMAFADDKAVHAKLAEIKFNNKLALKRYLKDNKD 479

Query: 480 IELDEYSIIDTQIKRFHEYKRQQMNALYVIHKYLEIKRGHFPSRKLTVIFGGKAAPAYTI 539
           IELDE+SIIDTQIKRFHEYKRQQMNALYVIHKYLEIK+G+ P RK+TVIFGGKAAPAY I
Sbjct: 480 IELDEHSIIDTQIKRFHEYKRQQMNALYVIHKYLEIKKGNLPKRKITVIFGGKAAPAYII 539

Query: 540 AQDIIHLILCLSELINNDPEVNKYLNVHLVENYNVTVAEKLIPATDISEQISLASKEASG 599
           AQDIIHLILCLSELINNDPEV+ YLNVHLVENYNVTVAE LIPATDISEQISLASKEASG
Sbjct: 540 AQDIIHLILCLSELINNDPEVSPYLNVHLVENYNVTVAEHLIPATDISEQISLASKEASG 599

Query: 600 TGNMKFMLNGALTLGTMDGANVEIAELAGKENIYTFGKDSDTIINLYETSGYRSKDYYDK 659
           TGNMKFMLNGALTLGTMDGANVEIAELAG ENIYTFGKDSDTIINLY T+ Y +KDYYD
Sbjct: 600 TGNMKFMLNGALTLGTMDGANVEIAELAGMENIYTFGKDSDTIINLYATASYVAKDYYDN 659

Query: 660 DKVIREAVDFIISDDIVSLGNAERLKRLHDELVGKDWFMTLIDLKEYIAVKEQVLADYED 719
              I+ AV+FIIS ++++ GN ERL RL+ EL+ KDWFMTLIDL+EYI VKE++LADYED
Sbjct: 660 HPAIKAAVNFIISPELLAFGNEERLDRLYKELISKDWFMTLIDLEEYIEVKEKMLADYED 719
```

```
Query: 720 YESWNKKVIHNIAKAGFFSSDRTIEQYNQDIWHS 753
           + W  KV+HNIAKAGFFSSDRTIEQYN+DIWHS
Sbjct: 720 QDLWMTKVVHNIAKAGFFSSDRTIEQYNEDIWHS 753
```

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

Example 2082

A DNA sequence (GBSx2197) was identified in *S.agalactiae* <SEQ ID 6443> which encodes the amino acid sequence <SEQ ID 6444>. This protein is predicted to be glycerol-3-phosphatase transporter (glpT). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.88   Transmembrane  339 - 355 ( 333 - 359)
   INTEGRAL    Likelihood = -8.01   Transmembrane  432 - 448 ( 426 - 450)
   INTEGRAL    Likelihood = -6.74   Transmembrane   92 - 108 (  91 - 127)
   INTEGRAL    Likelihood = -6.69   Transmembrane  194 - 210 ( 190 - 214)
   INTEGRAL    Likelihood = -3.77   Transmembrane  367 - 383 ( 364 - 385)
   INTEGRAL    Likelihood = -2.81   Transmembrane  111 - 127 ( 109 - 127)
   INTEGRAL    Likelihood = -2.28   Transmembrane  407 - 423 ( 406 - 424)
   INTEGRAL    Likelihood = -2.02   Transmembrane  165 - 181 ( 165 - 182)
   INTEGRAL    Likelihood = -0.64   Transmembrane   29 -  45 (  29 -  45)

----- Final Results -----
               bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAC44575 GB:U28354 IS629 ORFB fused with sequences similar to E.
           coli GlpT and UhpT proteins, Swiss-Prot Accession Number
           P08194 and P09836; Method: conceptual translation
           supplied by author [Shig
 Identities = 174/321 (54%), Positives = 241/321 (74%), Gaps = 4/321 (1%)

Query: 109 GVIPSVITSIWLFTIMYLINGWLQGMGYPPGARTLVYWYDNKERIKYATIWNLSHNFGGA 168
           GV P V + + +    YL+NGW+QGMGYPPGA+TLV+WY+++ERI +AT+WNLSHN GGA
Sbjct: 12  GVGP-VCSELHIAPSTYLLNGWIQGMGYPPGAKTLVFWYEHRERISWATLWNLSHNVGGA 70

Query: 169 IAPILTGVGLALAGNDSLNQARAAYWFPGVVACLLAVLVYFLQEDTPESIGLPPIEEYHK 228
           +AP+L G      G+ +L+ ARAA+ FPGV+   ++VL+YF+Q D P S+GLPPIEE+
Sbjct: 71  LAPVLIGFSFGFFGDSALDHARAAFIFPGVLCMAMSVLIYFIQVDRPVSVGLPPIEEWKG 130

Query: 229 EQYTNVVDSSDILEEPEVLGMGEIIKKYILPNTKLMWASLYSIFVYILRYGIVSWTPKFL 288
              ++        E+   L + +II+K+I+ N KL++  +Y  FVYILRYGIVSW PKFL
Sbjct: 131 NVVSHPAKGR---EQGPRLSIPDIIRKHIIRNNKLIYCCIYGSFVYILRYGIVSWAPKFL 187

Query: 289 ATSVQDGGKGITATAGMGGFSLFEIGGIIGMLTAGYLSAKVFKNSKPLTNVAFLVVAILL 348
           + S+  GGK +   A MGG S+FEIGG+ GML AGYLS ++F+NSKPLTN  FL + I+L
Sbjct: 188 SDSLDVGGKDMGKLASMGGGSVFEIGGVAGMLLAGYLSVRLFRNSKPLTNTLFLALTIIL 247

Query: 349 LAAYWFIPAGPQYMALDFIILLGLGASIYGPVMMVGLYAMELVPKAAAGAASGLTGTFSY 408
           L AYW++P+G +Y+ L++ IL+ LG ++YGPVM +GLY+MELVPK AAGAASGL+GTFSY
Sbjct: 248 LIAYWYVPSGNEYLWLNYTILILLGLAVYGPVMFIGLYSMELVPKEAAGAASGLSGTFSY 307

Query: 409 VGGATIATLAIGIIIDHFGWG 429
           + G+ +ATL +G+++D+ GWG
Sbjct: 308 IFGSIVATLGMGLVVDYLGWG 328
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6445> which encodes the amino acid sequence <SEQ ID 6446>. Analysis of this protein sequence reveals the following:

```
      Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.37   Transmembrane  185 - 201 ( 175 - 208)
   INTEGRAL    Likelihood = -9.13   Transmembrane  114 - 130 (  90 - 134)
   INTEGRAL    Likelihood = -7.75   Transmembrane  322 - 338 ( 320 - 345)
   INTEGRAL    Likelihood = -6.79   Transmembrane  421 - 437 ( 419 - 439)
   INTEGRAL    Likelihood = -6.37   Transmembrane   91 - 107 (  90 - 113)
   INTEGRAL    Likelihood = -5.36   Transmembrane  163 - 179 ( 161 - 181)
   INTEGRAL    Likelihood = -5.20   Transmembrane  350 - 366 ( 347 - 371)
   INTEGRAL    Likelihood = -4.41   Transmembrane   23 -  39 (  22 -  41)
   INTEGRAL    Likelihood = -3.77   Transmembrane  257 - 273 ( 249 - 273)
   INTEGRAL    Likelihood = -1.33   Transmembrane   61 -  77 (  61 -  77)
   INTEGRAL    Likelihood = -1.28   Transmembrane  383 - 399 ( 383 - 399)
   INTEGRAL    Likelihood = -0.90   Transmembrane  299 - 315 ( 299 - 315)

----- Final Results -----
               bacterial membrane --- Certainty=0.5946(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAF96050 GB:AE004355 glycerol-3-phosphate transporter [Vibrio cholerae]
 Identities = 128/438 (29%), Positives = 215/438 (48%), Gaps = 17/438 (3%)

Query: 1   LFMEEDYNKREP-EKFTQFLRRQKVVFFVAFF-GYVCAYLVRNNFKLMSNTIMVQNGWDK 58
           LF    + +R P +K      R +   F+  F GY   YL R NF L +   +++ G+ +
Sbjct: 21  LFKPAAHTQRLPSDKVDSVYSRLRWQLFIGIFVGYAGYYLGRKNFSL-AMPYLIEQGFSR 79

Query: 59  AQIAILLSCLTVSYGLAKFYMGALGDRVSLRKLFSISLGASALICILIGFF---NSSMVV 115
             + + L  ++++YGL+KF MG + DR + R   S  L  SAL+    GF      S+
Sbjct: 80  GDLGVALGAVSIAYGLSKFLMGNVSDRSNPRYFLSAGLLLSALVMFCFGFMPWATGSITA 139

Query: 116 LGILLVLCGVVQGALAPASQAMIANYFPNKTRGGAIAGWNISQNMGSALLPLTIALLTSM 175
           + ILL L G  QG    PA    + +++   K RG  ++ WN++ N+G  L    I  +  +
Sbjct: 140 MFILLFLNGWFQGMGWPACGRTMVHWWSRKERGEIVSVWNVAHNVGGGL----IGPIFLL 195

Query: 176 GLVVPANGNILLAFLIPGVLVFLFALCCWKLGGDNPESEGLDSLRTMYGDAGESAVASEE 235
           GL +  N +   AF +P     L A+  W +  D P+S GL  +     D  +    S E
Sbjct: 196 GLWM-FNDDWRTAFYVPAFFAVLVAVFTWLVMRDTPQSCGLPPIEEYKNDYPDDYDKSHE 254

Query: 236 EKHNLSYWQLIWKYVFCNPSLLLVAAVNVALYFVRFGIEDWMPIYLSQVANMSEAHIHFA 295
            +  ++  ++ +KYVF N  L  +A  N  +Y +R+G+ DW P+YL +  + +      +A
Sbjct: 255 NE--MTAKEIFFKYVFNNKLLWSIAIANAFVYLIRYGVLDWAPVYLKEAKHFTVDKSSWA 312

Query: 296 ISMLEWVAIPGSLVFAWLAVR-YPNKMAKVGAIGLFVLAAIVFVYERLTATGAPNYFLLL 354
             + EW  IPG+L+  W++ + +  + A  G + + ++    V VY      G P   +
Sbjct: 313 YFLYEWAGIPGTLLCGWISDKVFKGRRAPAGILFMVLVTLAVLVY-WFNPAGNPAVDMAA 371

Query: 355 VIAGILGSLIYGPQLIVNILTINFVPLNVAGTAIGFVGVTAYLIGNMGANWLMPILADGF 414
           V  ++A  +G LIYGP +++ +  +   P   AGTA G  G+  YL G + AN ++    D F
Sbjct: 372 LVA--IGFLIYGPVMLIGLYALELAPKKAAGTAAGLTGLFGYLGGAVAANAILGYTVDHF 429

Query: 415 GWFWSYIVVAALSAFSAV 432
           GW   ++V+ A    S +
Sbjct: 430 GWDGGFMVLVASCVLSVL 447
```

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

```
 Identities = 117/439 (26%), Positives = 203/439 (45%), Gaps = 27/439 (6%)

Query: 23  KYPRYRVQVLISIFVGYMGYYFVRNTTSILSGILNMS----ATEIGIITCASYIAYGLSK 78
           ++ R +  V    F GY+  Y VRN   ++S  + +        +I I+     ++YGL+K
Sbjct: 17  QFLRRQKVVFFVAFFGYVCAYLVRNNFKLMSNTIMVQNGWDKAQIAILLSCLTVSYGLAK 76
```

```
Query: 79  FISGLISDESNSKIFLPVGLFLTGLVNVLIGVIPSVITSIWLFTIMYLINGWLQGMGYPP 138
           F  G + D  + +    + L  + L+ +LIG   S   S+ +  I+ ++ G +QG   P
Sbjct: 77  FYMGALGDRVSLRKLFSISLGASALICILIGFFNS---SMVVLGILLVLCGVVQGALAPA 133

Query: 139 GARTLVYWYDNKERIKYATIWNLSHNFGGAIAPI----LTGVGLALAGNDSLNQARAAYW 194
               +  ++ NK R      WN+S N G A+ P+    LT +GL +  N ++     A+
Sbjct: 134 SQAMIANYFPNKTRGGAIAGWNISQNMGSALLPLTIALLTSMGLVVPANGNI---LLAFL 190

Query: 195 FPGVVACLLAVLVYFLQEDTPESIGLPPIEEYHKEQYTNVVDSSDILEEPEVLGMGEIIK 254
            PGV+  L A+  + L  D PES GL  +   + +   + V S    EE   L   ++I
Sbjct: 191 IPGVLVFLFALCCWKLGGDNPESEGLDSLRTMYGDAGESAVASE---EEKHNLSYWQLIW 247

Query: 255 KYILPNTKLMWASLYSIFVYILRYGIVSWTPKFLATSVQDGGKGITATAGMGGFSLFEIG 314
           KY+  N  L+  +  ++ +Y +R+GI  W P +L+          I          S+ E
Sbjct: 248 KYVFCNPSLLLVAAVNVALYFVRFGIEDWMPIYLSQVANMSEAHIHFA-----ISMLEWV 302

Query: 315 GIIGMLTAGYLSAKVFKNSKPLTNVAFLVVAILLLAAYWFIPAG-PQYMALDFIILLG-L 372
            I G L   +L+ +        +  +   V+A ++          G P Y  L  +++ G L
Sbjct: 303 AIPGSLVFAWLAVRYPNKMAKVGAIGLFVLAAIVFVYERLTATGAPNYFLL--LVIAGIL 360

Query: 373 GASIYGPVMMVGLYAMELVPKAAAGAASGLTGTFSYVGGATIATLAIGIIIDHFGWGVAF 432
           G+ IYGP ++V +  +  VP   AG A G  G  +Y+ G   A   + I+ D FGW  ++
Sbjct: 361 GSLIYGPQLIVNILTINFVPLNVAGTAIGFVGVTAYLIGNMGANWLMPILADGFGWFWSY 420

Query: 433 IIF-GISGFAAIVCTLLSR 450
           I+   +S F+A+   +L++
Sbjct: 421 IVVAALSAFSAVGYLILAK 439
```

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

Example 2083

A DNA sequence (GBSx2198) was identified in *S.agalactiae* <SEQ ID 6447> which encodes the amino acid sequence <SEQ ID 6448>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3202(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 6449> which encodes the amino acid sequence <SEQ ID 6450>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4473(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 54/100 (54%), Positives = 67/100 (67%)

Query: 1   MTYELCLEYGTYPLRPVDAWADEINTAPAFITEDKKLLELLEEVNTLFHELFLTIECSFH 60
           MTYELCLEYGTYPL  VDA+  E    P FI ED+ L   LE +N LFH+LF+TIE  FH
Sbjct: 1   MTYELCLEYGTYPLSRVDAYWGEDQNPPTFIQEDRLLCHKLETMNHLFHDLFVTIESQFH 60
```

```
Query: 61  YIGHDPPEKRAKITQIYHVIIEHLSIHYPEYDIKIESLLM 100
           Y+G + PEKRA+I  +Y  +   L   Y +Y IKIE+ L+
Sbjct: 61  YVGFNMPEKRAQIRILYQEVATILKSKYKDYPIKIETFLL 100
```

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

Example 2084

A DNA sequence (GBSx2199) was identified in *S.agalactiae* <SEQ ID 6451> which encodes the amino acid sequence <SEQ ID 6452>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2369(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:AAB81912 GB:U92974 unknown [Lactococcus lactis]
 Identities = 213/322 (66%), Positives = 260/322 (80%), Gaps = 5/322 (1%)

Query: 1   MSEKIRVLLYYKYVSIENAEEYAAKHLEFCKSIGLKGRILIADEGINGTVSGDYETTQKY 60
           M++  RVLLYY+YV IE+ E +A KHL  CK +GLKGRIL+ADEGINGTVSG  E T  Y
Sbjct: 1   MTQDYRVLLYYQYVPIEDGETFAQKHLADCKELGLKGRILVADEGINGTVSGTIEQTNAY 60

Query: 61  MDWVHSDERFADLWFKIDEENQQAFRKMFVRYKKEIVHLGLEDNNFDSDINPLETTGEYL 120
           M+ + +D RF+   FKIDE  Q AF+KM VRY+ E+V+L LED     D+NPLE TG YL
Sbjct: 61  MELMKNDPRFSSTIFKIDEAEQNAFKKMHVRYRPELVNLSLED-----DVNPLELTGAYL 115

Query: 121 NPKQFKEALLDEDTVVLDTRNDYEYDLGHFRGAIRPDIRNFRELPQWVRDNKDKFMEKRV 180
           +PK+F+EA+LDE+TVV+D RNDYE+DLGHFRGAIRP+IR+FRELPQW+RDNK++FMEKRV
Sbjct: 116 DPKEFREAMLDENTVVIDARNDYEFDLGHFRGAIRPEIRSFRELPQWIRDNKEQFMEKRV 175

Query: 181 VVYCTGGVRCEKFSGWMVREGFKDVGQLHGGIATYGKDPEVQGELWDGAMYVFDDRISVP 240
           + YCTGG+RCEKFSGW+VREGFKDVGQL GGIATYGKDPEVQG+LWDG MYVFD RI+VP
Sbjct: 176 LTYCTGGIRCEKFSGWLVREGFKDVGQLLGGIATYGKDPEVQGDLWDGQMYVFDSRIAVP 235

Query: 241 INHVNPTVISKDYFDGTPCERYVNCANPFCNKQIFASEENEAKYVRGCSPECRAHERNRY 300
           IN     ++ +D+FDG+PCERY+NC NP CN+Q+ ASEENEAKY+  CS ECR H  NRY
Sbjct: 236 INQKEHVIVGRDWFDGSPCERYINCGNPECNRQMLASEENEAKYLGACSHECRVHPNNRY 295

Query: 301 VQENGLSRQEWAERLEAIGESL 322
           ++ + LS QE  ERL  + + L
Sbjct: 296 IKAHQLSNQEVQERLALLEKDL 317
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6453> which encodes the amino acid sequence <SEQ ID 6454>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2443(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 = 321/324 (99%), Positives = 323/324 (99%)

Query: 1   MSEKIRVLLYYKYVSIENAEEYAAKHLEFCKSIGLKGRILIADEGINGTVSGDYETTQKY 60
```

```
          MSEKIRVLLYYKYVSIENA+EYAAKHLEFCKSIGLKGRILIADEGINGTVSGDYETTQKY
Sbjct: 1  MSEKIRVLLYYKYVSIENAQEYAAKHLEFCKSIGLKGRILIADEGINGTVSGDYETTQKY 60

Query: 61  MDWVHSDERFADLWFKIDEENQQAFRKMFVRYKKEIVHLGLEDNNFDSDINPLETTGEYL 120
          MDWVHSDERFADLWFKIDEENQQAFRKMFVRYKKEIVHLGLEDNNFDSDINPLETTGEYL
Sbjct: 61  MDWVHSDERFADLWFKIDEENQQAFRKMFVRYKKEIVHLGLEDNNFDSDINPLETTGEYL 120

Query: 121 NPKQFKEALLDEDTVVLDTRNDYEYDLGHFRGAIRPDIRNFRELPQWVRDNKDKFMEKRV 180
           NPKQFKEALLDEDTVVLDTRNDYEYDLGHFRGAIRPDIRNFRELPQWVRDNKDKFMEKRV
Sbjct: 121 NPKQFKEALLDEDTVVLDTRNDYEYDLGHFRGAIRPDIRNFRELPQWVRDNKDKFMEKRV 180

Query: 181 VVYCTGGVRCEKFSGWMVREGFKDVGQLHGGIATYGKDPEVQGELWDGAMYVFDDRISVP 240
           VVYCTGGVRCEKFSGWMVREGFKDVGQLHGGIATYGKDPEVQGELWDGAMYVFDDRISVP
Sbjct: 181 VVYCTGGVRCEKFSGWMVREGFKDVGQLHGGIATYGKDPEVQGELWDGAMYVFDDRISVP 240

Query: 241 INHVNPTVISKDYFDGTPCERYVNCANPFCNKQIFASEENEAKYVRGCSPECRAHERNRY 300
           INHVNPTVISKDYFDGTPCERYVNCANPFCNKQIFASEENE KYVRGCSPECRAHERNRY
Sbjct: 241 INHVNPTVISKDYFDGTPCERYVNCANPFCNKQIFASEENETKYVRGCSPECRAHERNRY 300

Query: 301 VQENGLSRQEWAERLEAIGESLPQ 324
           VQENGLSRQEWAERLEAIGESLP+
Sbjct: 301 VQENGLSRQEWAERLEAIGESLPE 324
```

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

Example 2085

A DNA sequence (GBSx2200) was identified in *S.agalactiae* <SEQ ID 6455> which encodes the amino acid sequence <SEQ ID 6456>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> 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:AAC83954 GB:L47648 putative [Bacillus subtilis]
 Identities = 54/192 (28%), Positives = 89/192 (46%), Gaps = 14/192 (7%)

Query: 5   QTIIIGAGAAGIGFGSAMQRLGLTNFLIIEKGHIGESFLRWPRTTQFITPSFTTNGFGFP 64
           + IIIG G  G+     ++++G+ + L+IEKG++  S   +P    F + S
Sbjct: 5   KAIIIGGGPCGLSAAIHLKQIGI-DALVIEKGNVVNSIYNYPTHQTFFSSSEKLE----- 58

Query: 65  DLNAVIPDTSPAFSFEKEHLSGVEYARYLQLVAAHYNLPIQNETSVLSIDK-RDSLFVIK 123
                I D   AF  E      ++   Y + V    N+ +     V  + K +++ FVI+
Sbjct: 59  -----IGDV--AFITENRKPVRIQALSYYREVVKRKNIRVNAFEMVRKVTKTQNNTFVIE 111

Query: 124 TSKGDFSADYLIMATGEFQNPNTIDIKGADLGMHYGQVDNFHIKSDNPFIIIGGNESACD 183
           TSK  ++  Y I+ATG + +PN + + G DL   +      H   D   ++IGG  S+ D
Sbjct: 112 TSKETYTTPYCIIATGYYDHPNYMGVPGEDLPKVFHYFKEGHPYFDKDVVVIGGKNSSVD 171

Query: 184 ALTHLVYLGNQV 195
           A   LV  G +V
Sbjct: 172 AALELVKSGARV 183
```

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

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

A related GBS gene <SEQ ID 8973> and protein <SEQ ID 8974> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 2
McG: Discrim Score:       5.05
GvH: Signal Score (-7.5): -3.14
     Possible site: 57
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 0 value:   0.26 threshold:  0.0
   PERIPHERAL  Likelihood =  0.26       6
 modified ALOM score:  -0.55

*** 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:

```
33.2/56.1% over 281aa
                                                                  Bacillus subtilis
  EGAD|109228| hypothetical protein Insert characterized
  GP|2635109|emb|CAB14605.1||Z99117 alternate gene name: yrdP Insert characterized
  GP|1934657|gb|AAB80908.1||U93876 hypothetical protein YrdP Insert characterized
  PIR|E69725|E69725 potassium uptake trkA - Insert characterized

ORF01799(310 - 1128 of 1725)
EGAD|109228|  S2656(2  -  283  of  345)  hypothetical  protein  {  acillus  subtilis}
GP|2635109|emb|CA  14605.1||Z99117  alternate  gene  name:  yrdP  {  acillus  subtilis}
GP|1934657|gb|AA  80908.1||U93876  hypothetical  protein  YrdP  {  acillus  subtilis}
PIR|E69725|E69725 potassium uptake trkA -  acillus subtilis
%Match = 6.1
%Identity = 33.2  %Similarity = 56.0
Matches = 77  Mismatches = 88  Conservative Sub.s = 53

270       300       330       360       390       417       444       474
YYC*LVKYFILHIYFCQGEDMKHYQTIIIGAGAAGIGFGSAMQRLGLTNFLIIEKGH-IGESFL-RWPRTTQFITPSFTT
                     MYDTIVIGAGQAGISIGYYLKQ-SDQKFIILDKSHEVGESWKDRYDSLVLFTSRMYSS
                               10        20        30        40        50

480       510       540       570       600       630       660       690
--------NGFGFPDLNAVIPDTSPAFSFEKEHLSGVEYARYLQLVAAHYNLPIQNETSVLSIDKRDSLFVIKTSKGDFS
LPGMHLEGEKHGFPSKNEIV-------------------AYLKKYVKKFEIPIQLRTEVISVLKIKNYFLIKTNREEYQ
          70                                80        90       100       110

720       750                               822       852       882       912
ADYLIMATGEFQNPNTIDIKGADLG-----MHYGQVDNF-HIKSDNPFIIIGGNESACDALTHLVYLGNQVELYTDTFGR
TKNLVIATGPFHTPNIPSIS-KDLSDNINQLHSSQYKNSKQLAYGNVLVVGGGNSGA----------------------
          130       140       150       160       170

942       969       996      1026
KESNPDPSISLS-PLTKERLKHIQ-DHKKEYYSISEGKKAI--EIKQIG------------------------------
---------QIAVELSKERVTYLACSNKLVYFPLMIGKRSIFWWFDKLGVLHASHTSIVGKFIQKKGDPVFGHELKHAIK
            180       190       200       210       220       230       240

1068      1098      1128      1158      1188      1218      1248
-------------KQYQVTFDDGSTAESFHKPILSTGFLNTCHLIDGIALFEYDKNQLPIVTEDDESTIVNNCFLIGPSL
QKEIILKKRVIAAKQNEIIFKDSSTLE-VNNIIWATGFRNPLCWINIKGVLDQEGRIIHHRGVSPVEGLYFIGLPWQHKR
            260       270       280       290       300       310       320
```

SEQ ID 8974 (GBS284) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 10; MW 42.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 9; MW 67.6kDa).

GBS284-GST was purified as shown in Figure 225, lane 7.

Example 2086

A DNA sequence (GBSx2201) was identified in *S.agalactiae* <SEQ ID 6457> which encodes the amino acid sequence <SEQ ID 6458>. This protein is predicted to be NrgA-like protein. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.73   Transmembrane    7 -  23 (   1 -  31)
   INTEGRAL    Likelihood = -6.42   Transmembrane   86 - 102 (  82 - 108)
   INTEGRAL    Likelihood = -6.42   Transmembrane  324 - 340 ( 318 - 342)
   INTEGRAL    Likelihood = -5.26   Transmembrane  210 - 226 ( 207 - 229)
   INTEGRAL    Likelihood = -5.10   Transmembrane  113 - 129 ( 112 - 133)
   INTEGRAL    Likelihood = -1.49   Transmembrane  246 - 262 ( 246 - 263)
   INTEGRAL    Likelihood = -1.17   Transmembrane  183 - 199 ( 183 - 199)
   INTEGRAL    Likelihood = -0.43   Transmembrane   41 -  57 (  41 -  57)
   INTEGRAL    Likelihood = -0.00   Transmembrane  265 - 281 ( 265 - 282)

----- Final Results -----
               bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9997> which encodes amino acid sequence <SEQ ID 9998> was also identified.

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

```
>GP:CAB15668 GB:Z99122 ammonium transporter [Bacillus subtilis]
 Identities = 105/378 (27%), Positives = 181/378 (47%), Gaps = 41/378 (10%)

Query: 3   VKKGLFVFLLLCILSMWLMIFGVAFVYFGSLH-QSLTSRIIYQFVLTVLLTTTAWFMGAY 61
           ++ G  VF+  C L +WLM  G+A +Y G +  +++ S  ++ F  ++ + +  W +  Y
Sbjct: 1   MQMGDTVFMFFCALLVWLMTPGLALFYGGMVKSKNVLSTAMHSFS-SIAIVSIVWVLFGY 59

Query: 62  FLAFEGHFKTVFQFQEADGKQI--------------VNCLFQLCFALYAVVMLIGSIIDR 107
            LAF      +   + A  K +               +  +FQ+ FA+     ++ G+  +R
Sbjct: 60  TLAFAPGNSIIGGLEWAGLKGVGFDPGDYSDTIPHSLFMMFQMTFAVLTTAIISGAFAER 119

Query: 108 VQTKRLLLAVVSWLFLVYTPLAYLIWNSEGVFAKMGVLDFSGGMIVHLSAGLSSYILAHV 167
           ++    LL  V W  LVYTP+A+ +W   G   ++G LDF+GG +VH+S+G++  +LA V
Sbjct: 120 MRFGAFLLFSVLWASLVYTPVAHWVWGG-GWIGQLGALDFAGGNVVHISSGVAGLVLAIV 178

Query: 168 IGK-----SEHQHNKVKNDSLFLGMILITFGWFGFNMGPVGEWNSQAIMILLNTIFAIIG 222
           +GK     +   HN +     FLG  LI FGWFGFN+G     +  A+    +NT  A
Sbjct: 179 LGKRKDGTASSPHNLIYT---FLGGALIWFGWFGFNVGSALTLDGVAMYAFINTNTAAAA 235

Query: 223 GGLAWTLAAKWNGEEEKTGSLLNGIIVGLVTSTAGVGYLLTWQLLAVTFFASLFTYFVTD 282
           G   W L      ++      ++G I GLV  T   G++  +  + +         ++
Sbjct: 236 GIAGWILVEWIINKKPTMLGAVSGAIAGLVAITPAAGFVTPFASIIIGIIGGAVCFWGVF 295

Query: 283 YVAKAFAIDDVVSSFGMNGIGGLLGSLGVGLFKLSHMP----------------VQLLAL 326
            + K F  DD + +FG++GIGG  G +  GLF  + +                  Q++A+
Sbjct: 296 SLKKKFGYDDALDAFGLHGIGGTWGGIATGLFATTSVNSAGADGLFYGDASLIWKQIVAI 355

Query: 327 ATTILLSIIMTYIISKAI 344
           A T +   I+T++I K +
Sbjct: 356 AATYVFVFIVTFVIIKIV 373
```

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

A related GBS gene <SEQ ID 8975> and protein <SEQ ID 8976> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:     17.19
GvH: Signal Score (-7.5): -4.07
     Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 9 value: -11.73 threshold:  0.0
   INTEGRAL    Likelihood =-11.73   Transmembrane    7 -  23 (   1 -  31)
   INTEGRAL    Likelihood = -6.42   Transmembrane   86 - 102 (  82 - 108)
   INTEGRAL    Likelihood = -6.42   Transmembrane  324 - 340 ( 318 - 342)
   INTEGRAL    Likelihood = -5.26   Transmembrane  210 - 226 ( 207 - 229)
   INTEGRAL    Likelihood = -5.10   Transmembrane  113 - 129 ( 112 - 133)
   INTEGRAL    Likelihood = -1.49   Transmembrane  246 - 262 ( 246 - 263)
   INTEGRAL    Likelihood = -1.17   Transmembrane  183 - 199 ( 183 - 199)
   INTEGRAL    Likelihood = -0.43   Transmembrane   41 -  57 (  41 -  57)
   INTEGRAL    Likelihood = -0.00   Transmembrane  265 - 281 ( 265 - 282)
   PERIPHERAL  Likelihood =  0.26     152
modified ALOM score:   2.85

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
ORF01800(307 - 1332 of 1641)
EGAD|19589|BS3646(1 - 373 of 404) probable ammonium transporter {Bacillus subtilis}
OMNI|NT01BS4254 ammonium transporter SP|Q07429|NRGA_BACSU PROBABLE AMMONIUM TRANSPORTER
(MEMBRANE PROTEIN NRGA). GP|143264|gb|AAA17399.1||L03216 membrane-associated protein
{Bacillus subtilis} GP|1684645|emb|CAB05374.1||Z82987 unknown {Bacillus subtilis}
GP|2636176|emb|CAB15668.1||Z99122 ammonium transporter {Bacillus subtilis}
PIR|A36865|A36865 ammonium transporter nrgA - Bacillus subtilis
%Match = 13.5
%Identity = 30.0  %Similarity = 54.8
Matches = 104  Mismatches = 149  Conservative Sub.s = 86

144       174       204       234       264       294       324       354
PFSMIRKFVSPNRCMAEPKPIPAAPAPIIMV**CFMSSP*QK*MCKIKYLTS*Q*YSLTNKRVFVKKGLFVFLLLCILSM
                                                          :: |  ||:::| | :
                                                         MQMGDTVFMFFCALLV
                                                               10

384       411       441       471       501       531
WLMIFGVAFYYFGSLH-QSLTSRIIYQFVLTVLLTTTAWFMGAYFLAFEGHFKTVFQFQEADGKQI--------------
|||  |:|::| | :  ::: |  :: |  :: : :  | :  | |||      :  :: |  | :
WLMTPGLALFYGGMVKSKNVLSTAMHSF-SSIAIVSIVWVLFGYTLAFAPGNSIIGGLEWAGLKGVGFDPGDYSDTIPHS
            30        40        50        60        70        80        90

579       609       639       669       699       729       759       789
VNCLFQLCFALYAVVMLIGSIIDRVQTKRLLLAVVSWLFLVYTPLAYLIWNSEGVFAKMGVLDFSGGMIVHLSAGLSSYI
:  :||: ||:    :: |:  :|::   :||  | |  |||||:|: :|   |   ::| |||:|| :||:|:|::  :
LFMMFQMTFAVLTTAIISGAFAERMRFGAFLLFSVLWASLVYTPVAHWVWGG-GWIGQLGALDFAGGNVVHISSGVAGLV
            110       120       130       140       150       160       170

819       849       873       903       933       963       993       1023
LAHVIGKSEHQHNKVKNDSLF--LGMILITFGWFGFNMGPVGEWNSQAIMILLNTIFAIIGGGLAWTLAAKWNGEEEKTG
|| |:|| :       :: ::  ||  || |||||||:|     :  |:  ::||  |   |   | |  :|   :: |
LAIVLGKRKDGTASSPHNLIYTFLGGALIWFGWFGFNVGSALTLDGVAMYAFINTNTAAAAGIAGWIL-VEWIINKKPTM
            190       200       210       220       230       240       250
```

```
1050       1080       1110       1140       1170       1200       1230       1260
-SLLNGIIVGLVTSTAGVGYLLTWQLLAVTFFASLFTYFVTDYXAKAFAIDDVVSSFGMNGIGGLLGSLGVGLFKLSHMP
   ::| | |||  |   |::  :  : :        ::       | |  || : :||::|||| | :  |||  : :
LGAVSGAIAGLVAITPAAGFVTPFASIIIGIIGGAVCFWGVFSLKKKFGYDDALDAFGLHGIGGTWGGIATGLFATTSVN
               270        280        290        300        310        320        330

1272       1302       1332       1362       1392       1422       1452
V----------------QLLALATTILLSIIMTYIISKAIFRK**IRLRCTSQPYLLF*QGE*LNRIINHFHY*TLSXX*
                  |::|:| | ::  |:|::| | :
SAGADGLFYGDASLIWKQIVAIAATYVFVFIVTFVIIKIVSLFLPLRATEEEESLGLDLTMHGEKAYQDSM
               350        360        370        380        390        400
```

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

Example 2087

A DNA sequence (GBSx2202) was identified in *S.agalactiae* <SEQ ID 6459> which encodes the amino acid sequence <SEQ ID 6460>. This protein is predicted to be dUTPase (dut). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2731(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 9471> which encodes amino acid sequence <SEQ ID 9472> was also identified.

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

```
>GP:CAA72644 GB:Y11901 dUTPase [Lactococcus lactis]
 Identities = 67/144 (46%), Positives = 90/144 (61%), Gaps = 8/144 (5%)

Query: 40  RGFELVSQFSNKELLPKRETAHAAGYDLKVAKKTVIEPGEITLVPTGIKAYMQPGEVLYL 99
           RGF+   +       +P+R T H+AGYD+   ++   I+P EI +V TG+   +   EVL L
Sbjct: 3   RGFK---KLDGNATIPERATKHSAGYDISASETVTIQPDEIKMVSTGLAVQLGDDEVLKL 59

Query: 100 YDRSSNPRKKGIVLINSVGVIDGDYYNNQVNEGHIFAQMQNITDQAVILEEGERIVQAVF 159
           YDRSSNP K+GI LINSVG+ID DYY  +           NI+ + V + +G+RI+Q VF
Sbjct: 60  YDRSSNPVKRGIALINSVGIIDSDYYPQEFK-----GLFMNISKEPVTISKGQRIMQGVF 114

Query: 160 APFLLADDDQATGMRTGGFGSTGK 183
             +L  DDD A G RTGGFGSTG+
Sbjct: 115 VKYLTIDDDNANGKRTGGFGSTGE 138
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6461> which encodes the amino acid sequence <SEQ ID 6462>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2519(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 = 115/148 (77%), Positives = 125/148 (83%)

Query: 36  MSKVRGFELVSQFSNKELLPKRETAHAAGYDLKVAKKTVIEPGEITLVPTGIKAYMQPGE 95
           M+K+RGFELVS F+N +LLPKRET HAAGYDL VA+    I PGEI LVPTG+KAYMQ GE
Sbjct: 1   MTKIRGFELVSSFTNPDLLPKRETTHAAGYDLSVAEAVTIAPGEIKLVPTGVKAYMQDGE 60

Query: 96  VLYLYDRSSNPRKKGIVLINSVGVIDGDYYNNQVNEGHIFAQMQNITDQAVILEEGERIV 155
           VLYLYDRSSNPRKKGI+LINSVGVID DYY N+ NEGHIFAQMQNITD  V L  GERIV
Sbjct: 61  VLYLYDRSSNPRKKGIILINSVGVIDADYYGNEANEGHIFAQMQNITDHPVTLAVGERIV 120

Query: 156 QAVFAPFLLADDDQATGMRTGGFGSTGK 183
           Q VF PFL+AD DQA G RTGGFGSTG+
Sbjct: 121 QGVFMPFLIADGDQARGERTGGFGSTGQ 148
```

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

Example 2088

A DNA sequence (GBSx2203) was identified in *S.agalactiae* <SEQ ID 6463> which encodes the amino acid sequence <SEQ ID 6464>. This protein is predicted to be RadA homolog (radA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2628(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:CAB11863 GB:Z99104 DNA repair protein homolog [Bacillus subtilis]
 Identities = 285/453 (62%), Positives = 358/453 (78%), Gaps = 4/453 (0%)

Query: 1   MAKKKSVFTCQECGYQSPKYLGRCPNCSAWSSFVEEVEVQEVKNARVSLNGEKSRPTKLK 60
           MAK KS F CQ CGY+SPK++G+CP C AW++ VEE+  +   N R + +        K
Sbjct: 1   MAKTKSKFICQSCGYESPKWMGKCPGCGAWNTMVEEMIKKAPANRRAAFSHSVQTVQKPS 60

Query: 61  DVSSINYS---RTKTDMDEFNRVLGGGVVPGSLVLIGGDPGIGKSTLLLQVSTQLA-NKG 116
            ++SI  S    R KT + EFNRVLGGGVV GSLVLIGGDPGIGKSTLLLQVS QL+ +
Sbjct: 61  PITSIETSEEPRVKTQLGEFNRVLGGGVVKGSLVLIGGDPGIGKSTLLLQVSAQLSGSSN 120

Query: 117 TVLYVSGEESAEQIKLRSERLGDIDNEFYLYAETNMQSIRSEIEKIKPDFLIIDSIQTIM 176
           +VLY+SGEES +Q KLR++RLG  +   ++ +ET+M+ I S I+++ P F+++DSIQT+
Sbjct: 121 SVLYISGEESVKQTKLRADRLGINNPSLHVLSETDMEYISSAIQEMNPSFVVVDSIQTVY 180

Query: 177 SPEVSSVQGSVSQVREVTAELMQLAKTNNIATFIVGHVTKEGTLAGPRMLEHMVDTVLYF 236
             +++S  GSVSQVRE TAELM++AKT  I  FIVGHVTKEG++AGPR+LEHMVDTVLYF
Sbjct: 181 QSDITSAPGSVSQVRECTAELMKIAKTKGIPIFIVGHVTKEGSIAGPRLLEHMVDTVLYF 240

Query: 237 EGERHHTFRILRAVKNRFGSTNEIGIFEMQSGGLVEVLNPSQVFLEERLDGATGSAIVVT 296
           EGERHHTFRILRAVKNRFGSTNE+GIFEM+  GL EVLNPS++FLEER  G+ GS+I  +
Sbjct: 241 EGERHHTFRILRAVKNRFGSTNEMGIFEMREEGLTEVLNPSEIFLEERSAGSAGSSITAS 300

Query: 297 MEGTRPILAEVQALVTPTVFGNAKRTTTGLDFNRVSLIMAVLEKRCGLLLQNQDAYLKSA 356
           MEGTRPIL E+QAL++PT FGN +R  TG+D NRVSL+MAVLEKR GLLLQNQDAYLK A
Sbjct: 301 MEGTRPILVEIQALISPTSFGNPRRMATGIDHNRVSLLMAVLEKRVGLLLQNQDAYLKVA 360

Query: 357 GGVKLDEPAIDLAVAVAIASSYKEKPTNPQESFIGEIGLTGEIRRVTRIEQRINEASKLG 416
           GGVKLDEPAIDLA+ ++IASS+++ P NP + FIGE+GLTGE+RRV+RIEQR+ EA+KLG
Sbjct: 361 GGVKLDEPAIDLAIVISIASSFRDTPPNPADCFIGEVGLTGEVRRVSRIEQRVKEAAKLG 420

Query: 417 FTKIYAPKNSLAGIEIPKGIDVIGVTTVSQVLK 449
```

```
            F ++  P  +L G   PKGI+VIGV  V++ L+
Sbjct: 421 FKRMIIPAANLDGWTKPKGIEVIGVANVAEALR 453
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6465> which encodes the amino acid sequence <SEQ ID 6466>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2191(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 = 416/453 (91%), Positives = 441/453 (96%)

Query: 1   MAKKKSVFTCQECGYQSPKYLGRCPNCSAWSSFVEEVEVQEVKNARVSLNGEKSRPTKLK 60
           MAKKK+ F CQECGYQSPKYLGRCPNCSAWSSFVEEVEV+EVKNARVSL GEKSRP KLK
Sbjct: 1   MAKKKATFICQECGYQSPKYLGRCPNCSAWSSFVEEVEVKEVKNARVSLAGEKSRPVKLK 60

Query: 61  DVSSINYSRTKTDMDEFNRVLGGGVVPGSLVLIGGDPGIGKSTLLLQVSTQLANKGTVLY 120
           DV +I+Y RT+TDM EFNRVLGGGVVPGSL+LIGGDPGIGKSTLLLQVSTQLANKGTVLY
Sbjct: 61  DVDNISYHRTQTDMSEFNRVLGGGVVPGSLILIGGDPGIGKSTLLLQVSTQLANKGTVLY 120

Query: 121 VSGEESAEQIKLRSERLGDIDNEFYLYAETNMQSIRSEIEKIKPDFLIIDSIQTIMSPEV 180
           VSGEESAEQIKLRSERLGDIDNEFYLYAETNMQ+IR+EIE IKPDFLIIDSIQTIMSP++
Sbjct: 121 VSGEESAEQIKLRSERLGDIDNEFYLYAETNMQAIRTEIENIKPDFLIIDSIQTIMSPDI 180

Query: 181 SSVQGSVSQVREVTAELMQLAKTNNIATFIVGHVTKEGTLAGPRMLEHMVDTVLYFEGER 240
           + VQGSVSQVREVTAELMQLAKTNNIATFIVGHVTKEGTLAGPRMLEHMVDTVLYFEGER
Sbjct: 181 TGVQGSVSQVREVTAELMQLAKTNNIATFIVGHVTKEGTLAGPRMLEHMVDTVLYFEGER 240

Query: 241 HHTFRILRAVKNRFGSTNEIGIFEMQSGGLVEVLNPSQVFLEERLDGATGSAIVVTMEGT 300
           HHTFRILRAVKNRFGSTNEIGIFEMQSGGLVEVLNPSQVFLEERLDGATGSA+VVTMEG+
Sbjct: 241 HHTFRILRAVKNRFGSTNEIGIFEMQSGGLVEVLNPSQVFLEERLDGATGSAVVVTMEGS 300

Query: 301 RPILAEVQALVTPTVFGNAKRTTTGLDFNRVSLIMAVLEKRCGLLLQNQDAYLKSAGGVK 360
           RPILAEVQ+LVTPTVFGNA+RTTTGLDFNRVSLIMAVLEKRCGLLLQNQDAYLKSAGGVK
Sbjct: 301 RPILAEVQSLVTPTVFGNARRTTTGLDFNRVSLIMAVLEKRCGLLLQNQDAYLKSAGGVK 360

Query: 361 LDEPAIDLAVAVAIASSYKEKPTNPQESFIGEIGLTGEIRRVTRIEQRINEASKLGFTKI 420
           LDEPAIDLAVAVAIASSYKEKPT+PQE+F+GEIGLTGEIRRVTRIEQRINEA+KLGFTK+
Sbjct: 361 LDEPAIDLAVAVAIASSYKEKPTSPQEAFLGEIGLTGEIRRVTRIEQRINEAAKLGFTKV 420

Query: 421 YAPKNSLAGIEIPKGIDVIGVTTVSQVLKAVFS 453
           YAPKN+L GI+IP+GI+V+GVTTV QVL AVFS
Sbjct: 421 YAPKNALQGIDIPQGIEVVGVTTVGQVLNAVFS 453
```

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

Example 2089

A DNA sequence (GBSx2204) was identified in *S.agalactiae* <SEQ ID 6467> which encodes the amino acid sequence <SEQ ID 6468>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3488(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAA97750 GB:Z73419 hypothetical protein Rv1284 [Mycobacterium
          tuberculosis]
 Identities = 69/162 (42%), Positives = 100/162 (61%), Gaps = 2/162 (1%)

Query: 3   TYFDNFLKTNQAYADLHGTAHLPIKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRNAG 62
           T  D++L  N  YA        LP+ P    +AIV CMD+RL V + LG+  G+AH++RNAG
Sbjct: 2   TVTDDYLANNVDYASGF-KGPLPMPPSKHIAIVACMDARLDVYRMLGIKEGEAHVIRNAG 60

Query: 63  GRVTDDVLRSLVISQQQLGTREIVVLHHTDCGAQTFTNEAFAAQLQRDLGVDMHGHDFLP 122
             VTDDV+RSL ISQ+ LGTREI++LHHTDCG  TFT++ F   +Q + G+
Sbjct: 61  CVVTDDVIRSLAISQRLLGTREIILLHHTDCGMLTFTDDDFKRAIQDETGIRPTWSP-ES 119

Query: 123 FNDIEESVREDVAKLHASPLIPDDVVISGAIYDVDTGRMVEV 164
           + D  E VR+ + ++  +P +     + G ++DV TG++ EV
Sbjct: 120 YPDAVEDVRQSLRRIEVNPFVTKHTSLRGFVFDVATGKLNEV 161
```

There is also homology to SEQ ID 6470:

```
Identities = 126/164 (76%), Positives = 146/164 (88%)

Query: 1   MTTYFDNFLKTNQAYADLHGTAHLPIKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN 60
           + +YF++F+  NQAY  LHGTAHLP+KPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN
Sbjct: 1   LMSYFEHFMAANQAYVALHGTAHLPLKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN 60

Query: 61  AGGRVTDDVLRSLVISQQQLGTREIVVLHHTDCGAQTFTNEAFAAQLQRDLGVDMHGHDF 120
           AGGRVT+D++RSLVISQQQ+GTREIVVLHHTDCGAQTFTNE FA  +   LGVD+ G DF
Sbjct: 61  AGGRVTEDMIRSLVISQQQMGTREIVVLHHTDCGAQTFTNEGFAKHIHEHLGVDVSGQDF 120

Query: 121 LPFNDIEESVREDVAKLHASPLIPDDVVISGAIYDVDTGRMVEV 164
           LPF D+E+SVRED+AK+ AS LI DDVVI+GA+YDVDTG+M +V
Sbjct: 121 LPFQDVEDSVREDMAKIRASSLISDDVVINGAVYDVDTGKMTQV 164
```

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

Example 2090

A DNA sequence (GBSx2205) was identified in *S.agalactiae* <SEQ ID 6471> which encodes the amino acid sequence <SEQ ID 6472>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0536(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 9473> which encodes amino acid sequence <SEQ ID 9474> was also identified.

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

```
>GP:AAC73407 GB:AE000137 putative oxidoreductase [Escherichia coli K12]
 Identities = 199/438 (45%), Positives = 286/438 (64%)

Query: 1   MKKYDVIVLGFGKAGKTLAAKLATQGKSVAMVEEDDKMYGGTCINIGCIPTKTLLVSASK 60
           M KY  +++GFGKAGKTLA  LA  G  VA++E+ + MYGGTCINIGCIPTKTL+  A +
Sbjct: 10  MNKYQAVIIGFGKAGKTLAVTLAKAGWRVALIEQSNAMYGGTCINIGCIPTKTLVHDAQQ 69

Query: 61  NHDFQEAMTTRNEVTSRLRAKNFAMLDNKDTVDVYNAKARFISNKVVELTGGADKQELTA 120
           + DF  A+  +NEV + LR KNF  L +    +DV + +A FI+N  + +         E+
```

```
Sbjct: 70  HTDFVRAIQRKNEVVNFLRNKNFHNLADMPNIDVIDGQAEFINNHSLRVHRPEGNLEIHG 129

Query: 121 DVIIINTGAKSVQLPIPGLADSQHVYDSTAIQELAHLPKRLGIIGGGNIGLEFATLYSEL 180
           + I INTGA++V  PIPG+  +  VYDST +  L  LP  LGI+GGG IG+EFA++++
Sbjct: 130 EKIFINTGAQTVVPPIPGITTTPGVYDSTGLLNLKELPGHLGILGGGYIGVEFASMFANF 189

Query: 181 GSKVTVIDSQSRIFAREEEELSEMAQDYLEEMGISFKLSADIKSVQNEDEDVVISFEDEK 240
           GSKVT++++ S    RE+ ++++     L + G+   L+A ++ + + + +  V +  E  +
Sbjct: 190 GSKVTILEAASLFLPREDRDIADNIATILRDQGVDIILNAHVERISHHENQVQVHSEHAQ 249

Query: 241 LSFDAVLYATGRKPNTEGLALENTDIKLTERGAIAVDEYCQTSVENIFAVGDVNGGPQFT 300
           L+ DA+L A+GR+P T  L  EN  I + ERGAI VD+   T+ +NI+A+GDV GG QFT
Sbjct: 250 LAVDALLIASGRQPATASLHPENAGIAVNERGAIVVDKRLHTTADNIWAMGDVTGGLQFT 309

Query: 301 YISLDDSRIVLNYLNCDKDYSLKNRGAVPTSTFTNPPLATVGLDEKTAKEKGYQVKSNSL 360
           YISLDD RIV + L  +   S  +R  VP S F  PPL+ VG+ E+ A+E G  ++  +L
Sbjct: 310 YISLDDYRIVRDELLGEGKRSTDDRKNVPYSVFMTPPLSRVGMTEEQARESGADIQVVTL 369

Query: 361 LVSAMPRAHVNNDLRGIFKVVVDTETNLILGARLFGAESHELINIITMAMDNKIPYTYFQ 420
            V+A+PRA V ND RG+ K +VD +T  +LGA L   +SHE+INI+ M MD  +PY+  +
Sbjct: 370 PVAAIPRARVMNDTRGVLKAIVDNKTQRMLGASLLCVDSHEMINIVKMVMDAGLPYSILR 429

Query: 421 KQIFTHPTMVENFNDLFN 438
           QIFTHP+M E+ NDLF+
Sbjct: 430 DQIFTHPSMSESLNDLFS 447
```

There is also homology to SEQ ID 1820.

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

Example 2091

A DNA sequence (GBSx2206) was identified in *S.agalactiae* <SEQ ID 6473> which encodes the amino acid sequence <SEQ ID 6474>. This protein is predicted to be glutamyl-tRNA synthetase (gltX). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2245(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 9475> which encodes amino acid sequence <SEQ ID 9476> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10953> which encodes amino acid sequence <SEQ ID 10954> was also identified.

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

```
>GP:AAC31971 GB:U49789 glutamyl-tRNA synthetase [Bacillus subtilis]
 Identities = 273/491 (55%), Positives = 353/491 (71%), Gaps = 19/491 (3%)

Query: 20  LANKIRVRYAPSPTGLLHIGNARTALFNYLYARHHGGDFVIRIEDTDRKRHVEDGERSQL 79
           + N++RVRYAPSPTG LHIGNARTALFNYL+AR+ GG F+IR+EDTD+KR++E GE+SQL
Sbjct: 1   MGNEVRVRYAPSPTGHLHIGNARTALFNYLFARNQGGKFIIRVEDTDKKRNIEGGEQSQL 60

Query: 80  ENLRWLGMDWDESPET---HENYRQSERLELYQRYIDQLLAEGKAYKSYVTEEELAAERE 136
             L+WLG+DWDES +     +  YRQSER ++Y+ Y ++LL +G AYK Y TEEEL  ERE
Sbjct: 61  NYLKWLGIDWDESVDVGGEYGPYRQSERNDIYKVYYEELLEKGLAYKCYCTEEELEKERE 120

Query: 137 RQELAGETPRYINEFIGMSETEKEAYIAEREAAGIIPTVRLAVNESGIYKWTDMVKGDIE 196
           Q   GE PRY  +  +++ E+E +IAE    G  P++R  V E  +  + D+VKG+I
```

```
Sbjct: 121 EQIARGEMPRYSGKHRDLTQEEQEKFIAE----GRKPSIRFRVPEGKVIAFNDIVKGEIS 176

Query: 197 FEGSNIGGDWVIQKKDGYPTYNFAVVIDDHDMQISHVIRGDDHIANTPKQLMVYEALGWE 256
           FE   IG D+VI KKDG PTYNFAV IDD+ M+++HV+RG+DHI+NTPKQ+M+Y+A GW+
Sbjct: 177 FESDGIG-DFVIVKKDGTPTYNFAVAIDDYLMKMTHVLRGEDHISNTPKQIMIYQAFGWD 235

Query: 257 APQFGHMTLIINSETGKKLSKRDTNTLQFIEDYRKKGYMSEAVFNFIALLGWNPGGEEEI 316
            PQFGHMTLI+N E+ KKLSKRD + +QFIE Y++ GY+ EA+FNFI LLGW+P GEEE+
Sbjct: 236 IPQFGHMTLIVN-ESRKKLSKRDESIIQFIEQYKELGYLPEALFNFIGLLGWSPVGEEEL 294

Query: 317 FSREQLINLFDENRLSKSPAAFDQKKMDWMSNDYLKNADFESVFALCKPFLEEAGRL--- 373
           F++EQ I +FD NRLSKSPA FD  K+ W++N Y+K  D + V  L  P L++AG++
Sbjct: 295 FTKEQFIEIFDVNRLSKSPALFDMHKLKWVNNQYVKKLDLDQVVELTLPHLQKAGKVGTE 354

Query: 374 -----TDKAEKLVELYQPQLKSADEIVPLTDLFFADFPELTEAEKEVMAAETVPTVLSAF 428
                +   KL+ LY  QL    EIV LTDLFF D  E  +  K V+  E VP VLS F
Sbjct: 355 LSAEEQEWVRKLISLYHEQLSYGAEIVELTDLFFTDEIEYNQEAKAVLEEEQVPEVLSTF 414

Query: 429 KEKLVSLSDEEFTRDTIFPQIKAVQKETGIKGKNLFMPIRIAVSGEMHGPELPDTIYLLG 488
             KL  L  EEFT D I   IKAVQKETG KGK LFMPIR+AV+G+ HGPELP +I L+G
Sbjct: 415 AAKLEEL--EEFTPDNIKASIKAVQKETGHKGKKLFMPIRVAVTGQTHGPELPQSIELIG 472

Query: 489 KEKSVQHIDNM 499
           KE ++Q + N+
Sbjct: 473 KETAIQRLKNI 483
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6475> which encodes the amino acid sequence <SEQ ID 6476>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1966(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 = 434/481 (90%), Positives = 459/481 (95%)

Query: 20  LANKIRVRYAPSPTGLLHIGNARTALFNYLYARHHGGDFVIRIEDTDRKRHVEDGERSQL 79
           ++  IRVRYAPSPTGLLHIGNARTALFNYLYAR HGG F+IRIEDTDRKRHVEDGERSQL
Sbjct: 1   MSKPIRVRYAPSPTGLLHIGNARTALFNYLYARRHGGTFIIRIEDTDRKRHVEDGERSQL 60

Query: 80  ENLRWLGMDWDESPETHENYRQSERLELYQRYIDQLLAEGKAYKSYVTEEELAAERERQE 139
           ENL+WLGMDWDESPETHENYRQSERL LYQ+YIDQLLAEGKAYKSYVTEEELAAERERQE
Sbjct: 61  ENLKWLGMDWDESPETHENYRQSERLALYQQYIDQLLAEGKAYKSYVTEEELAAERERQE 120

Query: 140 LAGETPRYINEFIGMSETEKEAYIAEREAAGIIPTVRLAVNESGIYKWTDMVKGDIEFEG 199
            AGETPRYINEFIGMS  EK  YIAEREAAGI+PTVRLAVNESGIYKWTDMVKGDIEFEG
Sbjct: 121 AAGETPRYINEFIGMSADEKAKYIAEREAAGIVPTVRLAVNESGIYKWTDMVKGDIEFEG 180

Query: 200 SNIGGDWVIQKKDGYPTYNFAVVIDDHDMQISHVIRGDDHIANTPKQLMVYEALGWEAPQ 259
            NIGGDWVIQKKDGYPTYNFAVV+DDHDMQISHVIRGDDHIANTPKQLMVYEALGWEAP+
Sbjct: 181 GNIGGDWVIQKKDGYPTYNFAVVVDDHDMQISHVIRGDDHIANTPKQLMVYEALGWEAPE 240

Query: 260 FGHMTLIINSETGKKLSKRDTNTLQFIEDYRKKGYMSEAVFNFIALLGWNPGGEEEIFSR 319
           FGHMTLIINSETGKKLSKRDTNTLQFIEDYRKKGYM EAVFNFIALLGWNPGGEEEIFSR
Sbjct: 241 FGHMTLIINSETGKKLSKRDTNTLQFIEDYRKKGYMPEAVFNFIALLGWNPGGEEEIFSR 300

Query: 320 EQLINLFDENRLSKSPAAFDQKKMDWMSNDYLKNADFESVFALCKPFLEEAGRLTDKAEK 379
           EQLI LFDENRLSKSPAAFDQKKMDWMSN+YLK+ADFE+V+ALCKPFLEEAGRLT+KAEK
Sbjct: 301 EQLIALFDENRLSKSPAAFDQKKMDWMSNEYLKHADFETVYALCKPFLEEAGRLTEKAEK 360

Query: 380 LVELYQPQLKSADEIVPLTDLFFADFPELTEAEKEVMAAETVPTVLSAFKEKLVSLSDEE 439
           LVELY+PQLKSADEI+PLTDLFF+DFPELTEAEKEVMA ETV TVL AFK KL ++SDE+
```

```
Sbjct: 361 LVELYKPQLKSADEIIPLTDLFFSDFPELTEAEKEVMAGETVSTVLQAFKAKLEAMSDED 420

Query: 440 FTRDTIFPQIKAVQKETGIKGKNLFMPIRIAVSGEMHGPELPDTIYLLGKEKSVQHIDNML 500
           F  + IFPQIKAVQKETGIKGKNLFMPIRIAVSGEMHGPELP+TIYLLG++KS++HI NML
Sbjct: 421 FKPENIFPQIKAVQKETGIKGKNLFMPIRIAVSGEMHGPELPNTIYLLGRDKSIEHIKNML 481
```

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

Example 2092

A DNA sequence (GBSx2207) was identified in *S.agalactiae* <SEQ ID 6477> which encodes the amino acid sequence <SEQ ID 6478>. This protein is predicted to be d-ribose-binding protein precursor , fragment (rbsB). 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>
```

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

```
>GP:CAB15613 GB:Z99122 ribose ABC transporter (ribose-binding
          protein) [Bacillus subtilis]
 Identities = 143/301 (47%), Positives = 205/301 (67%), Gaps = 1/301 (0%)

Query: 14  MSIVLILGACGKTGLGNSSGNSTKNVTKKSAKDLKLGVSISTTNNPYFVAMKDGIDKYAS 73
           +S++L L     T          K     + K+  +G+S+ST NNP+FV++K GI+K A
Sbjct: 5   VSVILTLSLFLLTACSLEPPQWAKPSNSGNKKEFTIGLSVSTLNNPFFVSLKKGIEKEAK 64

Query: 74  NKKISIKVADAQDDAARQADDVQNFISQNVDAILINPVDSKAIVTAIKSANNANIPVILM 133
            + + + + DAQ+D+++Q  DV++ I Q VDA+LINP DS AI TA++SAN   +PV+ +
Sbjct: 65  KRGMKVIIVDAQNDSSKQTSDVEDLIQQGVDALLINPTDSSAISTAVESANAVGVPVVTI 124

Query: 134 DRGSEGGKVLTTVASDNVAAGKMAADYAVKKLGKKAKAFELSGVPGASATVDRGKGFHSV 193
           DR +E GKV T VASDNV  G+MAA +    KLGK AK  EL GVPGASAT +RG GFH++
Sbjct: 125 DRSAEQGKVETLVASDNVKGGEMAAAFIADKLGKGAKVAELEGVPGASATRERGSGFHNI 184

Query: 194 AKSKLDILSSQSANFDRAKALNTTQNMIQGHKDVQIIFAQNDEMALGAAQAVKSAGLQNV 253
           A  KL +++ QSA+FDR K L   +N++QGH D+Q +FA NDEMALGA +A+ S+G +++
Sbjct: 185 ADQKLQVVTKQSADFDRTKGLTVMENLLQGHPDIQAVFAHNDEMALGALEAINSSG-KDI 243

Query: 254 LIVGIDGQPDAHDAIKKGDISATIAQQPAKMGEIAIQAAIDYYKGKKVEKETISPIYLVTK 314
           L++G DG  DA  +IK   +SAT+AQQP  +G++A +AA D   GKKV+K   +P+ L T+
Sbjct: 244 LVIGFDGNKDALASIKDRKLSATVAQQPELIGKLATEAADDILHGKKVQKTISAPLKLETQ 304
```

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

SEQ ID 6478 (GBS203) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 12; MW 36.8kDa).

GBS203-His was purified as shown in Figure 208, lane 8.

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

Example 2093

A DNA sequence (GBSx2208) was identified in *S.agalactiae* <SEQ ID 6479> which encodes the amino acid sequence <SEQ ID 6480>. This protein is predicted to be galactoside ABC transporter, permease protein (rbsC). Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.15   Transmembrane   63 -  79 (  52 -  85)
   INTEGRAL    Likelihood = -3.66   Transmembrane  111 - 127 ( 110 - 128)
   INTEGRAL    Likelihood = -2.71   Transmembrane  168 - 184 ( 168 - 188)
   INTEGRAL    Likelihood = -2.44   Transmembrane  189 - 205 ( 188 - 205)
   INTEGRAL    Likelihood = -0.80   Transmembrane   17 -  33 (  17 -  33)

----- 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 9287> which encodes amino acid sequence <SEQ ID 9288> was also identified.

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

```
>GP:CAB15612 GB:Z99122 ribose ABC transporter (permease) [Bacillus subtilis]
 Identities = 144/211 (68%), Positives = 182/211 (86%), Gaps = 1/211 (0%)

Query: 1   MGMLNGLFISYGKLAPFIVTLATMTIFRGATLVYSNGNPITAGLSDSFLFQFLGQGYIVG 60
           +GM+NGL I+ GK+APFI TLATMT+FRG TLVY++GNPIT GL  ++ FQ  G+GY +G
Sbjct: 113 LGMINGLLITKGKMAPFIATLATMTVFRGLTLVYTDGNPIT-GLGTNYGFQMFGRGYFLG 171

Query: 61  IPFPVILMFLTFIILYILLHKTAFGKSVYALGGNEKAAYISGIKLNKVKIIIYTISGIMA 120
           IP P I M L F+IL++LLHKT FG+  YA+GGNEKAA ISGIK+ +VK++IY+++G+++
Sbjct: 172 IPVPAITMVLAFVILWVLLHKTPFGRRTYAIGGNEKAALISGIKVTRVKVMIYSLAGLLS 231

Query: 121 SISGLIITSRLSSAQPTAGASYEMDAIAAVVLGGTSLSGGKGRIIGTLIGALIIGVLNNG 180
           +++G I+TSRL SAQPTAG SYE+DAIAAVVLGGTSLSGG+GRI+GTLIG LIIG LNNG
Sbjct: 232 ALAGAILTSRLHSAQPTAGESYELDAIAAVVLGGTSLSGGRGRIVGTLIGVLIIGTLNNG 291

Query: 181 LNIIGVSAFWQQVVKGIVILMAVLLDRFKVA 211
           LN++GVS+F+Q VVKGIVIL+AVLLDR K A
Sbjct: 292 LNLLGVSSFYQLVVKGIVILIAVLLDRKKSA 322
```

A related GBS gene <SEQ ID 8977> and protein <SEQ ID 8978> were also identified.

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

Example 2094

A DNA sequence (GBSx2209) was identified in *S.agalactiae* <SEQ ID 6481> which encodes the amino acid sequence <SEQ ID 6482>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.12   Transmembrane   75 -  91 (  74 -  91)
   INTEGRAL    Likelihood = -0.64   Transmembrane   96 - 112 (  96 - 112)

----- Final Results -----
              bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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 2095

A DNA sequence (GBSx2210) was identified in *S.agalactiae* <SEQ ID 6483> which encodes the amino acid sequence <SEQ ID 6484>. This protein is predicted to be ribose transport ATP-binding protein rbsa (rbsA). Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.00   Transmembrane  401 - 417 ( 401 - 417)

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

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

```
>GP:CAB15611 GB:Z99122 ribose ABC transporter (ATP-binding protein)
          [Bacillus subtilis]
 Identities = 297/493 (60%), Positives = 375/493 (75%), Gaps = 1/493 (0%)

Query: 1   MKIDMRNISKSFGTNKVLEKIDLELQSGQIHALMGENGAGKSTLMNILTGLFPASTGTIY 60
           M+I+M++I K+FG N+VL  +  +L  G++HALMGENGAGKSTLMNILTGL  A  G I
Sbjct: 1   MQIEMKDIHKTFGKNQVLSGVSFQLMPGEVHALMGENGAGKSTLMNILTGLHKADKGQIS 60

Query: 61  IDGEERTFSNPQEAEEFGISFIHQEMNTWPEMTVLENLFLGREIKTTFGLLNQKLMRQKA 120
           I+G E  FSNP+EAE+ GI+FIHQE+N WPEMTVLENLF+G+EI +  G+L  + M+  A
Sbjct: 61  INGNETYFSNPKEAEQHGIAFIHQELNIWPEMTVLENLFIGKEISSKLGVLQTRKMKALA 120

Query: 121 LETFKRLGVTIPLDIPIGNLSVGQQQMIEIAKSLLNQLSILVMDEPTAALTDRETENLFR 180
            E F +L V++ LD   G  SVGQQQMIEIAK+L+     +++MDEPTAALT+RE   LF
Sbjct: 121 KEQFDKLSVSLSLDQEAGECSVGQQQMIEIAKALMTNAEVIIMDEPTAALTEREISKLFE 180

Query: 181 VIRGLKQEGVGVVYISHRMEEIFKITDFVTVMRDGVIVDTKETSLINSDELVKKMVGRKL 240
           VI  LK+ GV +VYISHRMEEIF I D +T+MRDG  VDT   S T+ DE+VKKMVGR+L
Sbjct: 181 VITALKKNGVSIVYISHRMEEIFAICDRITIMRDGKTVDTTNISETDFDEVVKKMVGREL 240

Query: 241 EDYYPEKHSEIGPVAFEVSNL-CGDNFEDVSFYVRKGEILGFSGLMGAGRTEVMRTIFGI 299
            + YP++   +G   FEV N      +FEDVSFYVR GEI+G SGLMGAGRTE+MR +FG+
Sbjct: 241 TERYPKRTPSLGDKVFEVKNASVKGSFEDVSFYVRSGEIVGVSGLMGAGRTEMMRALFGV 300

Query: 300 DKKKSGKVKIDDQEITITTPSQAIKQGIGFLTENRKDEGLILDFNIKDNMTLPSTKDFSK 359
           D+  +G++ I  ++  I  P +A+K+G+GF+TENRKDEGL+LD +I++N+ LP+   FS
Sbjct: 301 DRLDTGEIWIAGKKTAIKNPQEAVKKGLGFITENRKDEGLLLDTSIRENIALPNLSSFSP 360

Query: 360 HGFFDEKTSTTFVQQLINRLYIKSGRPDLEVGNLSGGNQQKVVLAKWIGIAPKVLILDEP 419
            G  D K    FV  LI RL IK+  P+    +LSGGNQQKVV+AKWIGI PKVLILDEP
Sbjct: 361 KGLIDHKREAEFVDLLIKRLTIKTASPETHARHLSGGNQQKVVIAKWIGIGPKVLILDEP 420

Query: 420 TRGVDVGAKREIYQLMNELADRGVPIVMVSSDLPEILGVSDRIMVMHEGRISGELSRKEA 479
           TRGVDVGAKREIY LMNEL +RGV I+MVSS+LPEILG+SDRI+V+HEGRISGE+  +EA
Sbjct: 421 TRGVDVGAKREIYTLMNELTERGVAIIMVSSELPEILGMSDRIIVVHEGRISGEIHAREA 480

Query: 480 DQEKVMQLATGGK 492
           +QE++M LATGG+
Sbjct: 481 TQERIMTLATGGR 493
```

There is also homology to SEQ ID 4678.

SEQ ID 6484 (GBS407d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 2-4; MW 72kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 5 & 6; MW 47kDa).

GBS407d-His was purified as shown in Figure 235, lane 9-10.

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

Example 2096

A DNA sequence (GBSx2211) was identified in *S.agalactiae* <SEQ ID 6485> which encodes the amino acid sequence <SEQ ID 6486>. This protein is predicted to be high affinity ribose transport protein rbsd (rbsD). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2673(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:CAB15610 GB:Z99122 ribose ABC transporter (membrane protein)
           [Bacillus subtilis]
 Identities = 74/131 (56%), Positives = 95/131 (72%), Gaps = 1/131 (0%)

Query: 1   MKKTGILNSHLAKLADDLGHTDRVCIGDLGLPVPNGIPKIDLSLTSGIPSFQEVLDIYLE 60
           MKK GILNSHLAK+  DLGHTD++ I D GLPVP+G+ KIDLSL  G+P+FQ+   + E
Sbjct: 1   MKKHGILNSHLAKILADLGHTDKIVIADAGLPVPDGVLKIDLSLKPGLPAFQDIAAVLAE 60

Query: 61  NILVEKVILAEEIKEANPDQLSRLLAKLDNSVSIEYVSHNHLKQMTQDVKAVIRTGENTP 120
            + VEKVI A EIK +N +  ++ L  L +    IEY+SH   K +T+D KAVIRTGE TP
Sbjct: 61  EMAVEKVIAAAEIKASNQEN-AKFLENLFSEQEIEYLSHEEFKLLTKDAKAVIRTGEFTP 119

Query: 121 YSNIILQSGVI 131
           Y+N ILQ+GV+
Sbjct: 120 YANCILQAGVL 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 2097

A DNA sequence (GBSx2212) was identified in *S.agalactiae* <SEQ ID 6487> which encodes the amino acid sequence <SEQ ID 6488>. This protein is predicted to be ribokinase (rbsK). Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> 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:CAB15609 GB:Z99122 ribokinase [Bacillus subtilis]
 Identities = 132/293 (45%), Positives = 177/293 (60%), Gaps = 4/293 (1%)

Query: 1   MSNIVIIGSISMDLVMETNRIAKEGETVFGQRFSMVPGGKGANQAVAIGRLSQERDNITI 60
            M NI +IGS SMDLV+ +++  K GETV G  F  VPGGKGANQAVA  RL  +   + +
Sbjct: 1   MRNICVIGSCSMDLVVTSDKRPKAGETVLGTSFQTVPGGKGANQAVAAARLGAQ---VFM 57

Query: 61  LGAIGEDSFGPILLDNLNKNHVTTDFVGTIP-SSSGVAQITLYNNDNRIIYCPGANGKVD 119
            +G +G+D +G  +L+NL  N V TD++  +  + SG A I L   DN I+   GAN  +
Sbjct: 58  VGKVGDDHYGTAILNNLKANGVRTDYMEPVTHTESGTAHIVLAEGDNSIVVVKGANDDIT 117

Query: 120 TKKWSQEWSIIKEADLVVLQNEIPHQANMKIANFCKEHSIKLLYNPAPSRETDIEMLDKV 179
                   I++ D+V++Q EIP +   ++  +C  H I ++ NPAP+R    E +D
Sbjct: 118 PAYALNALEQIEKVDMVLIQQEIPEETVDEVCKYCHSHDIPIILNPAPARPLKQETIDHA 177

Query: 180 DYFTPNEHECQELFPNQKLEDILATYPEKLIVTLGTKGAIYSDGKESHLIPALETKAVDT 239
             Y TPNEHE   LFP   + + LA YP KL +T G +G  YS G +  LIP+   + VDT
Sbjct: 178 TYLTPNEHEASILFPELTISEALALYPAKLFITEGKQGVRYSAGSKEVLIPSFPVEPVDT 237

Query: 240 TGAGDTFNGAFGYAISKKFKIAKALRFATLAAHLSVQKFGAQGGMPTIKEMED 292
            TGAGDTFN AF  A+++   I  ALRFA  AA LSV  FGAQGGMPT  E+E+
Sbjct: 238 TGAGDTFNAAFAVALAEGKDIEAALRFANRAASLSVCSFGAQGGMPTRNEVEE 290
```

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 2098

A DNA sequence (GBSx2213) was identified in *S.agalactiae* <SEQ ID 6489> which encodes the amino acid sequence <SEQ ID 6490>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2272(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 9477> which encodes amino acid sequence <SEQ ID 9478> was also identified.

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

```
>GP:CAB15608 GB:Z99122 transcriptional regulator (LacI family)
            [Bacillus subtilis]
 Identities = 141/327 (43%), Positives = 204/327 (62%), Gaps = 4/327 (1%)

Query: 13  MSTIRQVAEKAGVSTSTVSRYISQNGYVSQKASQKIEQAIRELHYVPNFLAQSLKTKKNQ 72
            M+TI+ VA  AGVS +TVSR ++ NGYV ++   ++  A+ +L+Y PN +A+SL  ++++
Sbjct: 1   MATIKDVAGAAGVSVATVSRNLNDNGYVHEETRTRVIAAMAKLNYYPNEVARSLYKRESR 60

Query: 73  LVGLLLPDISNPFFPRLARGVEEFLKEQGYRVMLGNTNNKSHLEEEYLNVLLQSNAAGII 132
            L+GLLLPDI+NPFFP+LARG E+ L  +GYR++ GN++ +   E EYL    Q++ AGII
Sbjct: 61  LIGLLLPDITNPFFPQLARGAEDELNREGYRLIFGNSDEELKKELEYLQTFKQNHVAGII 120

Query: 133 --TTHDFTKNHPEIDIPVVVVDRVNQETQYGVFSDNKEGGKLAAQAIWTAGATNILLIRG 190
              T +   + +  ++ PVV +DR   E    V SD   G KLAAQAI    + I L+RG
Sbjct: 121 AATNYPDLEEYSGMNYPVVFLDR-TLEGAPSVSSDGYTGVKLAAQAIIHGKSQRITLLRG 179

Query: 191 PLDKADNLNQRFQGSQNYLLNKGACFAIEDSASFDFAEIQIEAKTLLDHHPDIDSIIAPS 250
```

```
           P         RF G+   L      F + ++ASF   + Q  AK L   +P  D +IA +
Sbjct: 180 PA-HLPTAQDRFNGALEILKQAEVDFQVIETASFSIKDAQSMAKELFASYPATDGVIASN 238

Query: 251 DIHAIAYLHEILNRGKRIPEDVQIIGYDDILMSQFIYPSLSTIHQSSYIMGQKAAELIFK 310
           DI A A LHE L RGK +PED+QIIGYDDI  S  ++P LSTI Q +Y MG++AA+L+
Sbjct: 239 DIQAAAVLHEALRRGKNVPEDIQIIGYDDIPQSGLLFPPLSTIKQPAYDMGKEAAKLLLG 298

Query: 311 ITNQLPITNKRIKLPVHYVERETLRRK 337
           I  + P+    I++PV Y+ R+T R++
Sbjct: 299 IIKKQPLAETAIQMPVTYIGRKTTRKE 325
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6491> which encodes the amino acid sequence <SEQ ID 6492>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1657(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 = 232/328 (70%), Positives = 274/328 (82%)

Query: 10  GVSMSTIRQVAEKAGVSTSTVSRYISQNGYVSQKASQKIEQAIRELHYVPNFLAQSLKTK 69
           G +M TI+QVAE+AGVS STVSRYISQ GYVS  A  KI+ AI +LHY PN LAQSLKTK
Sbjct: 14  GKAMVTIKQVAEEAGVSRSTVSRYISQKGYVSDDARHKIKAAIAKLHYTPNVLAQSLKTK 73

Query: 70  KNQLVGLLLPDISNPFFPRLARGVEEFLKEQGYRVMLGNTNNKSHLEEEYLNVLLQSNAA 129
           KNQLVGLLLPDISNPFFPRLARG EE+LKE+GYRVMLGN ++   LEEEY++VLLQSNAA
Sbjct: 74  KNQLVGLLLPDISNPFFPRLARGAEEYLKEKGYRVMLGNISDSEALEEEYVHVLLQSNAA 133

Query: 130 GIITTHDFTKNHPEIDIPVVVVDRVNQETQYGVFSDNKEGGKLAAQAIWTAGATNILLIR 189
           GIITTHDFTK +P + IPVVVVDRV+QETQYGVFSDN+ GG LAAQ +W AGA  +LLIR
Sbjct: 134 GIITTHDFTKRYPTLAIPVVVVDRVDQETQYGVFSDNRAGGLLAAQTVWQAGAKEVLLIR 193

Query: 190 GPLDKADNLNQRFQGSQNYLLNKGACFAIEDSASFDFAEIQIEAKTLLDHHPDIDSIIAP 249
           GPLD A+N+N+RF+ S +YL  +     + DS +FDF  IQ+EA   L  +P IDSIIAP
Sbjct: 194 GPLDNAENINERFEASFSYLQKQDVTMYVCDSQNFDFESIQLEASYNLKCYPTIDSIIAP 253

Query: 250 SDIHAIAYLHEILNRGKRIPEDVQIIGYDDILMSQFIYPSLSTIHQSSYIMGQKAAELIF 309
           SDIHAIAY+HE+ ++GK+IP+DVQIIGYDDILMSQFIYPSLSTIHQSSY+MG+ AAEL++
Sbjct: 254 SDIHAIAYIHELHSQGKKIPQDVQIIGYDDILMSQFIYPSLSTIHQSSYLMGRYAAELVY 313

Query: 310 KITNQLPITNKRIKLPVHYVERETLRRK 337
           K  I +QL +   RIKLPVHYVERET+R++
Sbjct: 314 KTIASQLTVKANRIKLPVHYVERETIRKR 341
```

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

Example 2099

A DNA sequence (GBSx2214) was identified in *S.agalactiae* <SEQ ID 6493> which encodes the amino acid sequence <SEQ ID 6494>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.80   Transmembrane   27 -  43 (  24 -  51)
   INTEGRAL    Likelihood =-10.61   Transmembrane  337 - 353 ( 329 - 362)
   INTEGRAL    Likelihood = -9.18   Transmembrane  257 - 273 ( 249 - 276)
   INTEGRAL    Likelihood = -8.92   Transmembrane  302 - 318 ( 291 - 326)
```

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

A related GBS nucleic acid sequence <SEQ ID 8979> which encodes amino acid sequence <SEQ ID 8980> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 6
SRCFLG: 0
McG: Length of UR:   4
     Peak Value of UR:   3.20
     Net Charge of CR: 1
McG: Discrim Score:      6.06
GvH: Signal Score (-7.5): 0.0500002
     Possible site: 46
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 47
ALOM program   count: 3 value: -10.61 threshold:  0.0
   INTEGRAL    Likelihood =-10.61   Transmembrane  326 - 342 ( 318 - 348)
   INTEGRAL    Likelihood = -9.18   Transmembrane  246 - 262 ( 238 - 265)
   INTEGRAL    Likelihood = -8.92   Transmembrane  291 - 307 ( 280 - 315)
   PERIPHERAL  Likelihood =  4.98     152
 modified ALOM score:   2.62
icm1 HYPID: 7  CFP: 0.525

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5246(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAF12525 GB:AE001863 hypothetical protein [Deinococcus radiodurans]
 Identities = 103/352 (29%), Positives = 191/352 (54%), Gaps = 9/352 (2%)

Query: 15  AWKELTFYKKKYLLIELLIIVMMFMVVFLSGLANGLGRAVSAAIENNPAQTYILNEGAEQ 74
           A +EL   K + LLI  ++ ++ FMV  L+GL  GL R  ++ + + PAQ+++  + A+
Sbjct: 4   ALRELQHQKLRSLLIGGIVALIAFMVFMLTGLTRGLSRDSASLLLDTPAQSFVTTKEADG 63

Query: 75  VITSSVLTTKDQTDLNSLNLKDSTTLNIQRSSLTRQGHEKKIDISYFAIDKDSFMAPTLS 134
           V+  S L+ +    +++L   +       ++ ++    +K++       +D   F+AP +S
Sbjct: 64  VLNRSFLSPEQ---VSALQQDNEDAAAFAQTFVSFSHGDKQLSGVLLGVDPRGFLAPDVS 120

Query: 135 EGKQLTSYKKAIILNDSLKAEGIKLGDKVIDKSSSISLTVVGFVHNSMYGHGPVAFIDKD 194
           EG+ L     A++ ++SL+ +G+K+GD +  K S   L V GF  ++   H P  ++
Sbjct: 121 EGQTLRVAGGAVV-DESLREDGVKVGDVLTLKPSGDQLRVSGFTRSARLNHQPGMYVSLA 179

Query: 195 IYTEINKKINPQYQFLPQALVMKNDKSISHLP-TQLEAVSKKDVIQHIPGYSAEQSTLNM 253
            +    +K+NP+      A+ +    +  +L   L   ++   +Q +PGY  EQ +L M
Sbjct: 180 RW----QKLNPRMHGTVNAVALPAAPAQVNLGGADLSVTNRAQTLQVLPGYKEEQGSLTM 235

Query: 254 ILWVLVVASAGILGVFFYIITLQKRHEFSVMKAIGTKMSEIALFQLSQVIILALFGIIVG 313
           I   L+  +A +L  FFY++TLQK  +F ++KAIG      +A    ++Q++IL L  + +
Sbjct: 236 IQVFLIAVAAFVLATFFYVMTLQKTAQFGLLKAIGASNRTLAGSVVAQMLILTLLAVAIA 295

Query: 314 DGLAVALSYVLPAQMPFVINWQNIILVSFVFLVIAMISSALSIVKVAKIDPV 365
             + +  +LPA MPF +   NI   S + LV+A ++S LS+ +VAK+DP+
Sbjct: 296 AAVTLGMVQLLPAGMPFHLTAANIASASGLLLVVAALASLLSVRRVAKVDPL 347
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6495> which encodes the amino acid sequence <SEQ ID 6496>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
```

```
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.31   Transmembrane  246 - 262 ( 233 - 270)
   INTEGRAL    Likelihood = -8.49   Transmembrane  327 - 343 ( 321 - 351)
   INTEGRAL    Likelihood = -1.01   Transmembrane  301 - 317 ( 301 - 317)

----- 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:AAF12525 GB:AE001863 hypothetical protein [Deinococcus radiodurans]
 Identities = 101/360 (28%), Positives = 175/360 (48%), Gaps = 11/360 (3%)

Query: 1   MFLALNEMKQSKLRYGLIAGLLCLVAYLMFFLSGLAFGLMQENRSAVDLWKADSVLLAKD 60
           M+LAL E++  KLR  LI G++ L+A+++F L+GL  GL +++ S +    A S +  K+
Sbjct: 1   MYLALRELQHQKLRSLLIGGIVALIAFMVFMLTGLTRGLSRDSASLLLDTPAQSFVTTKE 60

Query: 61  ADATLTLSQVSRAQENQITADKVAPLAQLNTVAWSVKNPKDADKVKVSLFGIDSNSFIRP 120
           AD  L  S +S  Q + +  D     A   T         K    V   L G+D   F+ P
Sbjct: 61  ADGVLNRSFLSPEQVSALQQDNEDAAAFAQTFVSFSHGDKQLSGV---LLGVDPRGFLAP 117

Query: 121 NIVKGRLFKTNKEVVLDQSLAKEEAFAIGKDFYTSSSSQALTIVGYTQNARFSVAPVVYM 180
           ++ +G+  +      V+D+SL +E+   +G        S   L + G+T++AR +  P +Y+
Sbjct: 118 DVSEGQTLRVAGGAVVDESL-REDGVKVGDVLTLKPSGDQLRVSGFTRSARLNHQPGMYV 176

Query: 181 NLEAFETLKYGEPLPKDKQVVNAFITKGS--LTDYPKKDFQKLDIKTFITKLPGYSAQLL 238
           +L  ++ L       P+     VNA      +     +  D     +  LPGY   +
Sbjct: 177 SLARWQKLN-----PRMHGTVNAVALPAAPAQVNLGGADLSVTNRAQTLQVLPGYKEEQG 231

Query: 239 TFGFMISFLVIISAIIIGIFMYILTIQKAPIFGIMKAQGISNKTITTAVLMQTFFLSFLG 298
           +   +  FL+ ++A ++  F Y++T+QK   FG++KA G SN+T+  +V+ Q    L+ L
Sbjct: 232 SLTMIQVFLIAVAAFVLATFFYVMTLQKTAQFGLLKAIGASNRTLAGSVVAQMLILTLLA 291

Query: 299 SGLGLLGTWLTSLLLPTVVPFQSNWFLYLAIFVSMICFALLGTLFSVFNIIRIDPLKAIG 358
             +   T    LLP  +PF      +   ++  A L +L SV  + ++DPL A+G
Sbjct: 292 VAIAAAVTLGMVQLLPAGMPFHLTAANIASASGLLLVVAALASLLSVRRVAKVDPLIALG 351
```

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

```
 Identities = 96/356 (26%), Positives = 178/356 (49%), Gaps = 4/356 (1%)

Query: 15  AWKELTFYKKKYLLIELLIIVMMFMVVFLSGLANGLGRAVSAAIENNPAQTYILNEGAEQ 74
           A  E+   K +Y LI  L+ ++ +++ FLSGLA GL +   +A++   A + +L + A+
Sbjct: 4   ALNEMKQSKLRYGLIAGLLCLVAYLMFFLSGLAFGLMQENRSAVDLWKADSVLLAKDADA 63

Query: 75  VITSSVLTTKDQTDLNSLNLKDSTTLNIQRSSLTRQGHEKKIDISYFAIDKDSFMAPTLS 134
           +T S ++   + + +  LN   S+      K+ +S F ID +SF+ P +
Sbjct: 64  TLTLSQVSRAQENQITADKVAPLAQLNTVAWSVKNPKDADKVKVSLFGIDSNSFIRPNIV 123

Query: 135 EGKQLTSYKKAIILNDSLKAEGIKLGDKVIDKSSSISLTVVGFVHNSMYGHGPVAFIDKD 194
           +G+   + K+ ++     K E   +G      SSS +LT+VG+  N+ +   PV +++ +
Sbjct: 124 KGRLFKTNKEVVLDQSLAKEEAFAIGKDFYTSSSSQALTIVGYTQNARFSVAPVVYMNLE 183

Query: 195 IYTEIN-KKINPQYQFLPQALVMKNDKSISHLPTQ-LEAVSKKDVIQHIPGYSAEQSTLN 252
           +  +   + P+ + A + K   S++  P +  + K  I  +PGYSA+  T
Sbjct: 184 AFETLKYGEPLPKDKQVVNAFITKG--SLTDYPKKDFQKLDIKTFITKLPGYSAQLLTFG 241

Query: 253 MILWVLVVASAGILGVFFYIITLQKRHEFSVMKAIGTKMSEIALFQLSQVIILALFGIIV 312
           ++ LV+ SA I+G+F YI+T+QK   F +MKA G     I   L Q   L+  G  +
Sbjct: 242 FMISFLVIISAIIIGIFMYILTIQKAPIFGIMKAQGISNKTITTAVLMQTFFLSFLGSGL 301

Query: 313 GDGLAVALSYVLPAQMPFVINWQNIILVSFVFLVIAMISSALSIVKVAKIDPVEVI 368
           G      S +LP  +PF  NW   + +    + A++ +  S+  + +IDP++ I
Sbjct: 302 GLLGTWLTSLLLPTVVPFQSNWFLYLAIFVSMICFALLGTLFSVFNIIRIDPLKAI 357
```

SEQ ID 8980 (GBS239) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 175 (lane 13; MW 64kDa).

GBS239-GST was purified as shown in Figure 227, lane 4.

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

Example 2100

A DNA sequence (GBSx2215) was identified in *S.agalactiae* <SEQ ID 6497> which encodes the amino acid sequence <SEQ ID 6498>. This protein is predicted to be heterocyst maturation protein (devA) (b0879). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1751(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:CAA05977 GB:AJ003195 ATP-binding subunit [Anabaena variabilis]
 Identities = 87/225 (38%), Positives = 146/225 (64%), Gaps = 1/225 (0%)

Query: 3   AILELKHISKHYPDGDELLSILDNLDLSVSAGEFVAILGPSGSGKSTLLSIAGLLLGADQ 62
           A++ +K ++ +Y  G      IL +++L +  GE V + GPSGSGK+TLLS+ G L    +
Sbjct: 5   AVIAIKSLNHYYGKGALKRQILFDINLEIYPGEIVIMIGPSGSGKTTLLSLIGGLRSVQE 64

Query: 63  GSLYVNHENVTDLSQRQRTQLRREALGFIFQSHQLLPYLTIQEQLQQEARFAKHYDKKTS 122
           G+L      ++  SQ +  Q+RR ++G+IFQ+H LL +LT ++ +Q      +H  ++ +
Sbjct: 65  GNLQFLGVELSGASQNKLVQIRR-SIGYIFQAHNLLGFLTARQNVQMAVELNEHISQEEA 123

Query: 123 LEEINKLLSDLGIEQCAHKYPNQLSGGQKQRAAIARAFINHPKVILADEPTASLDEERGR 182
           + +   +L  +G+E     YP+ LSGGQKQR AIARA +N+P ++LADEPTA+LD++ GR
Sbjct: 124 IAKAEAMLKAVGLENRVDYYPDNLSGGQKQRVAIARALVNNPPLVLADEPTAALDKQSGR 183

Query: 183 QVTELIRQEVKSHNTAAIMVTHDERVLDLVDTVYRLKDGKLVKEN 227
            V E++++  K   T+ ++VTHD R+LD+ D +  ++DG L +++
Sbjct: 184 DVVEIMQRLAKDQGTSILLVTHDNRILDIADRIVEMEDGILARDS 228
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6499> which encodes the amino acid sequence <SEQ ID 6500>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4181(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 = 103/224 (45%), Positives = 149/224 (65%), Gaps = 4/224 (1%)

Query: 3   AILELKHISKHYPDGDELLSILDNLDLSVSAGEFVAILGPSGSGKSTLLSIAGLLLGADQ 62
           ++L  K ++K + DG    ++ L    D S+ AGEFVAI+GPSGSGKST L+IAG L
Sbjct: 3   SVLTFKQVTKTFQDGHHEINALKATDFSIEAGEFVAIIGPSGSGKSTFLTIAGGLQTPSS 62

Query: 63  GSLYVNHENVTDLSQRQRTQLRREALGFIFQSHQLLPYLTIQEQLQQEARFAKHYDKKTS 122
```

```
           G L ++  + T LS+++R++LR +++GFI Q+  L+P+ T+Q+QL+       H
Sbjct: 63  GQLIIDGTDYTHLSEKERSRLRFKSVGFILQASNLIPFSTVQQQLE----LVDHLTGSKE 118

Query: 123 LEEINKLLSDLGIEQCAHKYPNQLSGGQKQRAAIARAFINHPKVILADEPTASLDEERGR 182
             + N+L  DLGI    H+ P +LSGG++QRAAIARA  + P +ILADEPTASLD E+
Sbjct: 119 KAKANQLFDDLGITGLKHQLPQELSGGERQRAAIARALYHDPALILADEPTASLDTEKAY 178

Query: 183 QVTELIRQEVKSHNTAAIMVTHDERVLDLVDTVYRLKDGKLVKE 226
           +V +L+ +E K  N A IMVTHD+R+L   D VYR++DG+L +E
Sbjct: 179 EVVKLLAKESKEKNKAIIMVTHDDRMLKYCDKVYRMQDGELCQE 222
```

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

Example 2101

A DNA sequence (GBSx2216) was identified in *S.agalactiae* <SEQ ID 6501> which encodes the amino acid sequence <SEQ ID 6502>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2645(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:CAB64972 GB:AJ012050 VicR protein [Enterococcus faecalis]
 Identities = 86/229 (37%), Positives = 132/229 (57%), Gaps = 10/229 (4%)

Query: 3   KILVVEDNIVQQKIITTKLTQEGYQFITASNGQEALNCLDTEEVQLIITDIMMPMMDGYQ 62
           KILVV+D     +I+   L +EGY+  TA +G+EAL  ++  E  LII D+M+P MDG +
Sbjct: 52  KILVVDDEKPISEIVKYNLVKEGYEVFTAYDGEEALEKVEEVEPDLIILDLMLPKMDGLE 111

Query: 63  LIQELRSAAYNVPIIVMTAKSQMEDMTKGFGLGADDYMVKPVQLQELALRIKALLRR--- 119
           + +E+R   +++PII++TAK    D   G  LGADDY+ KP   +EL  R+KA LRR
Sbjct: 112 VAREVRK-THDMPIIMVTAKDSEIDKVLGLELGADDYVTKPFSNRELVARVKANLRRGAT 170

Query: 120 ----ANIVAQHQLIIGNTCLNEDELSLKYFEQEIIFPQKEFRVLFHLLSYPNRIFTRLEL 175
               A +  Q +L IG+  ++ D   +    ++I    +EF +L++L  +  ++ TR  L
Sbjct: 171 NAKEAEVTTQSELTIGDLTIHPDAYMVSKRGEKIELTHREFELLYYLAKHIGQVMTREHL 230

Query: 176 LDSIWGMDTDLDERVVDACINKIRRKVEHLPDFK--IETVRGVGYRAKN 222
           L ++WG D   D R VD  + ++R K+E  P      + T RGVGY  +N
Sbjct: 231 LQTVWGYDYFGDVRTVDVTVRRLREKIEDSPSHPTYLVTRRGVGYYLRN 279
```

There is also homology to SEQ ID 1182.

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

Example 2102

A DNA sequence (GBSx2217) was identified in *S.agalactiae* <SEQ ID 6503> which encodes the amino acid sequence <SEQ ID 6504>. This protein is predicted to be sensor protein. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL   Likelihood = -8.97   Transmembrane   53 -  69 (  47 -  77)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAC62214 GB:AF049873 sensor protein [Lactococcus lactis]
 Identities = 97/307 (31%), Positives = 169/307 (54%), Gaps = 16/307 (5%)

Query: 57  SALAVVFLSLVIASISMWYGSYHLTKPILDISHIVSNVADGDFEGHIYRNSNRRKSYEYY 116
           + LAV+  +L++ + S++Y +  +T+P+L I       +A GD     +  N+
Sbjct: 170 AVLAVI--TLIVTAFSIFYITRTVTRPLLKIKLGTDKIAQGDLSIQLNVNTE-------- 219

Query: 117 NELDELSESINQMIVSLSHMDHMRKDFITNVSHELKTPIAAVANIVELLQDPELDEETQS 176
           +EL EL++SI  +   L  M   R +F+++V+HEL+TP+  +    ++         E ++
Sbjct: 220 DELGELAKSIEDLAEKLDFMKRERNEFLSSVAHELRTPLTFIKGYADIANRSTTSLEDKT 279

Query: 177 ELLGLVKTESLRLTRLCDTMLQMSRVDNQETIGELSSVRVDEQIRQAMISLTERWQAKRI 236
           + L +++ ES  LT+L + ++ +++++      E   V + E I + +  ++  +  KRI
Sbjct: 280 QYLRIIREESRHLTQLMEDLMNLAQLEENGFKVEKHQVLIQELINEVVSKVSGVFSEKRI 339

Query: 237 NFQLDSKPYTVYSNSDLLM--QVWINLLDNAIKYSEDIVDLSVRMEETNNHYLRVIISDK 294
           NF L S      Y+N D +   QV +NLL NA KYS D  D+ +        ++ +++ISDK
Sbjct: 340 NF-LISGEGNFYANIDFMRIEQVLVNLLMNAYKYSADESDIKLAFIPEKENF-KIVISDK 397

Query: 295 GRGISQYDVQHIFDKFYQADQSHNQQ--GNGLGLAIVKRIIVLCKGRISVSSQLEIGTEF 352
           G GI + D+ +IF++FY+ D+S  +   G GLGLAIV+ I+    G+I V S    GT F
Sbjct: 398 GEGIPEQDLPYIFERFYRVDKSRTRTTGGVGLGLAIVQDIVKKHNGKIIVESIQNQGTTF 457

Query: 353 CVELPLS 359
            +ELP S
Sbjct: 458 IIELPYS 464
```

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 8981> and protein <SEQ ID 8982> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:       4.84
GvH: Signal Score (-7.5): 0.179999
     Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 1 value:  -8.97 threshold:  0.0
   INTEGRAL    Likelihood = -8.97   Transmembrane   50 -  66 (  47 -  77)
   PERIPHERAL  Likelihood =  1.27     324
 modified ALOM score:   2.29

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
31.9/57.3% over 293aa
                                                                Lactococcus lactis
  GP|3687664| sensor protein Insert characterized

ORF01881(478 - 1377 of 1677)
GP|3687664|gb|AAC62214.1||AF049873(171 - 464 of 464) sensor protein (Lactococcus lactis)
%Match = 12.9
```

```
%Identity = 31.9  %Similarity = 57.3
Matches = 94  Mismatches = 121  Conservative Sub.s = 75

339       369       399       429       459       489       519       549
MTKLRRFRFPLRFYFTLMFVLTMLFSVLASLLLVAAIVFTFFQGVLTTHVLQVSALAVVFLSLVIASISMWYGSYHLTKP
  | ::  : : :      |          ::    : ::         :      :  ::|:: : |::| :  :|:|
EKKNKKESLHFHWLGDKYIVSKSRIQSNGKIVGSVYMFLSTRPIQKMVFNFTGIFAVLAVITLIVTAFSIFYITRTVTRP
              130       140       150       160       170       180       190

579       609       639       669       699       729       759       789
ILDISHIVSNVADGDFEGHIYRNSNRRKSYEYYNELDELSESINQMIVSLSHMDHMRKDFITNVSHELKTPIAAVANIVE
:| |      :| ||:  ::  |:         :|| ||::||  :   |  |   | :|:::|:|||:||:  :    :
LLKIKLGTDKIAQGDLSIQLNVNTE--------DELGELAKSIEDLAEKLDFMKRERNEFLSSVAHELRTPLTFIKGYAD
              210       '       220       230       240       250       260

819       849       879       909       939       969       999       1029
LLQDPELDEETQSELLGLVKTESLRLTRLCDTMLQMSRVDNQETIGELSSVRVDEQIRQAMISLTERWQAKRINFQLDSK
:       | ::: | ::: ||  ||:| : :: :::::      |   | : | | : :  ::  :  |||||  | |
IANRSTTSLEDKTQYLRIIREESRHLTQLMEDLMNLAQLEENGFKVEKHQVLIQELINEVVSKVSGVFSEKRINF-LISG
          280       290       300       310       320       330       340

1059      1083      1113      1143      1173      1203      1233
PYTVYSNSDLL--MQVWINLLDNAIKYSEDIVDLSVRMEETNNHYLRVIISDKGRGISQYDVQHIFDKFYQADQSHNQQ-
    |:| |::   || :||| || ||| |  |: :       :: :::|||||  || : |: :||::||: |:|  :
EGNFYANIDFMRIEQVLVNLLMNAYKYSADESDIKLAFIPEKENF-KIVISDKGEGIPEQDLPYIFERFYRVDKSRTRTT
          360       370       380       390       400       410       420

1287      1317      1347      1377      1407      1437      1467      1497
-GNGLGLAIVKRIIVLCKGRISVSSQLEIGTEFCVELPLS*LFKTITANWQLLFYLFRNKYTKNRQKL*KYLTINIASV*
 | |||||||: |:    |:| | |    || | :||| |
GGVGLGLAIVQDIVKKHNGKIIVESIQNQGTTFIIELPYS
          440       450       460
```

SEQ ID 8982 (GBS170d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 181 (lane 4; MW 35kDa) and in Figure 123 (lane 5-7; MW 35kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 123 (lane 2-4; MW 60kDa) and in Figure 184 (lane 3; MW 60kDa). Purified GBS170d-GST is shown in Figure 243, lane 7; purified GBS170d-His is shown in Figure 234, lanes 5-6.

Example 2103

A DNA sequence (GBSx2218) was identified in *S.agalactiae* <SEQ ID 6505> which encodes the amino acid sequence <SEQ ID 6506>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0502(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:BAB06906 GB:AP001518 argininosuccinate synthase
          (citrulline-asparate ligase) [Bacillus halodurans]
 Identities = 262/396 (66%), Positives = 321/396 (80%), Gaps = 1/396 (0%)

Query: 1   MGKEKLILAYSGGLDTSVAIAWLK-KDYDVIAVCMDVGEGKDLDFIHDKALTIGAIESYI 59
            M K+K++LAYSGGLDTSVAI WL  K YDVIAV +DVGEGKDL+F+ +KAL +GAIESY
Sbjct: 1   MSKKKVVLAYSGGLDTSVAIKWLSDKGYDVIAVGLDVGEGKDLEFVKEKALKVGAIESYT 60

Query: 60  LDVKDEFAEHFVLPALQAHAMYEQKYPLVSALSRPIIAQKLVEMAHQTGATTIAHGCTGK 119
            +D K EFAE FVLPALQAHA+YEQKYPLVSALSRP+I++KLVE+A QTGA  +AHGCTGK
Sbjct: 61  IDAKKEFAEEFVLPALQAHALYEQKYPLVSALSRPLISKKLVEIAEQTGAQAVAHGCTGK 120
```

```
Query: 120 GNDQVRFEVAIAALDPELKVIAPVREWKWHREEEITFAKANGVPIPADLDNPYSIDQNLW 179
           GNDQVRFEV+I AL+P L+V+APVREW W R+EEI +AK N +PIP DLDNPYS+DQNLW
Sbjct: 121 GNDQVRFEVSIQALNPNLEVLAPVREWAWSRDEEIEYAKKNNIPIPIDLDNPYSVDQNLW 180

Query: 180 GRANECGVLENPWNQAPEEAFGITKSPEEAPDCAEYIDITFQNGKPIAINNQEMTLADLI 239
           GR+NECG+LE+PW   PE A+ +T + E+APD  E ++I F+ G P+ +N +   + +LI
Sbjct: 181 GRSNECGILEDPWATPPEGAYELTVAIEDAPDQPEIVEIGFEKGIPVTLNGKSYPVHELI 240

Query: 240 LSLNEIAGKHGIGRIDHVENRLVGIKSREIYECPAAMVLLAAHKEIEDLTLVREVSHFKP 299
           L LN+IAGKHG+GRIDHVENRLVGIKSRE+YECP AM L+ AHKE+EDLTL +EV+HFKP
Sbjct: 241 LELNQIAGKHGVGRIDHVENRLVGIKSREVYECPGAMTLIKAHKELEDLTLTKEVAHFKP 300

Query: 300 ILENELSNLIYNALWFSPATKAIIAYVKETQKVVNGTTKVKLYKGSAQVVARHSSNSLYD 359
           ++E +++ LIY  LWFSP   A+ A++KETQ  V G  +VKL+KG A V  R S  SLY+
Sbjct: 301 VVEKKIAELIYEGLWFSPLQPALSAFLKETQSTVTGVVRVKLFKGHAIVEGRKSEYSLYN 360

Query: 360 ENLATYTAADSFDQDAAVGFIKLWGLPTQVNAQVNK 395
           E LATYT  D FD +AAVGFI LWGLPT+V + VNK
Sbjct: 361 EKLATYTPDDEFDHNAAVGFISLWGLPTKVYSMVNK 396
```

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 2104

A DNA sequence (GBSx2219) was identified in *S.agalactiae* <SEQ ID 6507> which encodes the amino acid sequence <SEQ ID 6508>. This protein is predicted to be argininosuccinate lyase (argH). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2131(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:BAB06905 GB:AP001518 argininosuccinate lyase [Bacillus halodurans]
 Identities = 284/454 (62%), Positives = 350/454 (76%)

Query: 6   KLWGGRFESSLEKWVEEFGASISFDQKLAPYDMKASMAHVTMLGKTDIISQEEAGLIKDG 65
           KLWGGRF  + E WV+EFGASI FDQ+L   D++ S+AHVTML K+ I++ EE   IK G
Sbjct: 3   KLWGGRFTKTAEAWVDEFGASIGFDQQLVEEDIEGSLAHVTMLEKSGILANEEVEQIKKG 62

Query: 66  LKILQDKYRAGQLTFSISNEDIHMNIESLLTAEIGEVAGKLHTARSRNDQVATDMHLYLK 125
           L IL +K + G+L +S++NEDIH+NIE LL  EIG V GKLHT RSRNDQVATDMHLYL+
Sbjct: 63  LHILLEKAKKGELNYSVANEDIHLNIEKLLIDEIGPVGGKLHTGRSRNDQVATDMHLYLR 122

Query: 126 DKLQEMMKKLLHLRTTLVNLAENHIYTVMPGYTHLQHAQPISFGHHLMAYYNMFTRDTER 185
            + +E+++ + +++  LV  A+ H+ T++PGYTHLQ AQPISF HHL+AY+ M  RD  R
Sbjct: 123 KQTKEILQLVKNVQAALVEQAKQHVETLIPGYTHLQRAQPISFAHHLLAYFWMLERDYGR 182

Query: 186 LEFNMKHTNLSPLGAAALAGTTFPIDRHMTTRLLDFEKPYSNSLDAVSDRDFIIEFLSNA 245
            E ++K  N+SPLGA ALAGTTFPID   T  LL F+  Y NSLDAVSDRDFI+EFLS +
Sbjct: 183 YEDSLKRLNVSPLGAGALAGTTFPIDREYTAELLGFDGIYENSLDAVSDRDFIVEFLSAS 242

Query: 246 SILMMHLSRFCEEIINWCSYEYQFITLSDTFSTGSSIMPQKKNPDMAELIRGKTGRVYGN 305
           S+LM HLSR CEE+I W S E+QF+ + D F+TGSSIMPQKKNPDMAELIRGKTGRVYG+
Sbjct: 243 SLLMTHLSRLCEELILWSSQEFQFVEMDDAFATGSSIMPQKKNPDMAELIRGKTGRVYGS 302
```

```
Query: 306 LFSLLTVMKSLPLAYNKDLQEDKEGMFDSVETVSIAIEIMANMLETMTVNEHIMMTSTET 365
           LFSLLTV+K LPLAYNKD+QEDKEGMFD+V+TV  ++ I A M++TM V E  M  +
Sbjct: 303 LFSLLTVLKGLPLAYNKDMQEDKEGMFDAVKTVKGSLAIFAGMIQTMKVKEETMTKAVHQ 362

Query: 366 DFSNATELADYLASKGVPFRKAHEIVGKLVLECSKNGSYLQDIPLKYYQEISELIENDIY 425
           DFSNATELADYLA+KG+PFR+AHE+VGKLVL C + G YL D+PL  Y+  S+L + DIY
Sbjct: 363 DFSNATELADYLATKGMPFREAHEVVGKLVLLCIQKGIYLLDLPLSDYKAASDLFDEDIY 422

Query: 426 EILTAKTAVKRRNSLGGTGFDQVKKQILLARKEL 459
           ++L  KT V RR S GGTGF +VKK I  A K L
Sbjct: 423 DVLQPKTVVARRTSAGGTGFTEVKKAIAKAEKIL 456
```

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

Example 2105

A DNA sequence (GBSx2220) was identified in *S.agalactiae* <SEQ ID 6509> which encodes the amino acid sequence <SEQ ID 6510>. This protein is predicted to be class-II aldolase (fba). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2930(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 9289> which encodes amino acid sequence <SEQ ID 9290> was also identified. Analysis of this sequence reveals:

```
GvH: Signal Score (-7.5): -2.92
     Possible site: 42
>>> Seems to have no N-terminal signal seq.
ALOM program   count: 0 value:   0.37 threshold:  0.0
   PERIPHERAL  Likelihood =  0.37      66
 modified ALOM score:  -0.57

*** Reasoning Step: 3

----- Final Results -----

            bacterial cytoplasm --- Certainty=0.2930(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:BAB16889 GB:AB050113 class-II aldolase [Streptococcus bovis]
 Identities = 221/242 (91%), Positives = 234/242 (96%)

Query: 1   MAIVSAEKFVQAARDNGYAVGGFNTNNLEWTQAILRAAEAKKAPVLIQTSMGAAKYMGGY 60
           MAIVSAEKF++AAR+NGYAVGGFNTNNLEWTQAILRAAEAKKAP+LIQTSMGAAKYMGGY
Sbjct: 1   MAIVSAEKFIKAARENGYAVGGFNTNNLEWTQAILRAAEAKKAPILIQTSMGAAKYMGGY 60

Query: 61  KLCKQLIETLVESMGITVPVAIHLDHGHYDDALECIEVGYTSIMFDGSHLPVEENLEKAR 120
           KLCK LIE LVESMGITVPVAIHLDHGH++DALECIEVGYTS+MFDGSHLPVEENLEKA+
Sbjct: 61  KLCKTLIENLVESMGITVPVAIHLDHGHFEDALECIEVGYTSVMFDGSHLPVEENLEKAK 120

Query: 121 EVVAKAHAKGISVEAEVGTIGGEEDGIVGKGELAPIEDAKAMVETGIDFLAAGIGNIHGP 180
           EVVAKAHAKG+SVEAEVGTIGGEEDGIVG GELAPIEDAKAMV TGIDFLAAGIGNIHGP
Sbjct: 121 EVVAKAHAKGVSVEAEVGTIGGEEDGIVGGGELAPIEDAKAMVATGIDFLAAGIGNIHGP 180
```

```
Query: 181 YPANWEGLDLDHLKKLTEAVPGFPIVLHGGSGIPDDQIQEAIKLGVAKVNVNTECQLAFC 240
           YPANW+GL LDHLKKLT AVPGFPIVLHGGSGIPDDQI+ AIKLGVAKVNVNTECQ+AF
Sbjct: 181 YPANWQGLHLDHLKKLTAAVPGFPIVLHGGSGIPDDQIKAAIKLGVAKVNVNTECQIAFA 240

Query: 241 QA 242
           +A
Sbjct: 241 KA 242
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6511> which encodes the amino acid sequence <SEQ ID 6512>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2930(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 = 217/242 (89%), Positives = 228/242 (93%)

Query: 1   MAIVSAEKFVQAARDNGYAVGGFNTNNLEWTQAILRAAEAKKAPVLIQTSMGAAKYMGGY 60
           MAIVSAEKFVQAAR+NGYAVGGFNTNNLEWTQAILRAAEAK+APVLIQTSMGAAKYMGGY
Sbjct: 1   MAIVSAEKFVQAARENGYAVGGFNTNNLEWTQAILRAAEAKQAPVLIQTSMGAAKYMGGY 60

Query: 61  KLCKQLIETLVESMGITVPVAIHLDHGHYDDALECIEVGYTSIMFDGSHLPVEENLEKAR 120
           K+C+ LI  LVESMGITVPVAIHLDHGHY+DALECIEVGYTSIMFDGSHLPVEENL K
Sbjct: 61  KVCQSLITNLVESMGITVPVAIHLDHGHYEDALECIEVGYTSIMFDGSHLPVEENLAKTA 120

Query: 121 EVVAKAHAKGISVEAEVGTIGGEEDGIVGKGELAPIEDAKAMVETGIDFLAAGIGNIHGP 180
           EVV  AHAKG+SVEAEVGTIGGEEDGI+GKGELAPIEDAKAMVETGIDFLAAGIGNIHGP
Sbjct: 121 EVVKIAHAKGVSVEAEVGTIGGEEDGIIGKGELAPIEDAKAMVETGIDFLAAGIGNIHGP 180

Query: 181 YPANWEGLDLDHLKKLTEAVPGFPIVLHGGSGIPDDQIQEAIKLGVAKVNVNTECQLAFC 240
           YP NWEGL LDHL+KLT AVPGFPIVLHGGSGIPDDQI+EAI+LGVAKVNVNTE Q+AF
Sbjct: 181 YPENWEGLALDHLEKLTAAVPGFPIVLHGGSGIPDDQIKEAIRLGVAKVNVNTESQIAFS 240

Query: 241 QA 242
           A
Sbjct: 241 NA 242
```

SEQ ID 9290 (GBS683) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 150 (lane 8 & 10; MW 55kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 150 (lane 11-13; MW 30kDa) and in Figure 184 (lane 11; MW 30kDa).

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

Example 2106

A DNA sequence (GBSx2221) was identified in *S.agalactiae* <SEQ ID 6513> which encodes the amino acid sequence <SEQ ID 6514>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2775(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:AAA88585 GB:M18954 unknown protein [Streptococcus mutans]
 Identities = 109/229 (47%), Positives = 156/229 (67%), Gaps = 1/229 (0%)

Query: 1   MFSGKRLKKRRITLGYSQSELADKLHINRSSYFNWENEKTKPNQSNLKQLAILLDVPETY 60
            MFS ++LK+RR  LG SQ++ ADKL I+R SYFNWE  KTKPNQ NL +LA LL V   Y
Sbjct: 1   MFSSQKLKERRKKLGLSQAQTADKLGISRPSYFNWEIGKTKPNQKNLDKLAHLLKVDSAY 60

Query: 61  FESEYKIVNTYLQLSLQNQEKVEKYAEELLQTQKVHEKIVPLFAVEVLSEIQLSAGPGEG 120
            F S++ IV  Y +L+  N+ K  KY++ LL+ Q      ++             +LSAG G
Sbjct: 61  FLSQHDIVEIYTRLNESNKTKTLKYSQHLLEQQDKKRNLMKNKRYPYRVYEKLSAGTGYS 120

Query: 121 LYDEFETETVYSEDEYTGFDIATWISGNSMEPVYKDGEVALIRSTGFDHDGAVYALNWNG 180
             + +   +TV+ ++E    D A+WI G+SMEP++ +GEVALI+ TGFD+DGA+YA++W+G
Sbjct: 121 YFGDGNFDTVFYDEEID-HDFASWIFGDSMEPIFLNGEVALIKQTGFDYDGAIYAIDWDG 179

Query: 181 SLYIKKLYREEDGFRMVSINPDVAERFIPFEDEIRIVGKIVGHFMPVIG 229
              YIKK+YREE G R+VS+N   A++F P+++  RI+G IVG+F+P+ G
Sbjct: 180 QTYIKKVYREETGLRLVSLNKKYADKFAPYDENPRIIGLIVGNFIPLEG 228
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6515> which encodes the amino acid sequence <SEQ ID 6516>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4340(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 = 84/209 (40%), Positives = 130/209 (62%), Gaps = 9/209 (4%)

Query: 25  LHINRSSYFNWENEKTKPNQSNLKQLAILLDVPETYFESEYKIVNTYLQLSLQNQEKVEK 84
            LH+N+ +  NWE  K  PN+ +L  L  L +V   YF+  Y+++  Y QL++ N+EKV
Sbjct: 5   LHVNKMTISNWEKGKNIPNEKHLNALLHLFNVTSDYFDPNYRLLTPYNQLTISNKEKVIG 64

Query: 85  YAEELLQTQ------KVHEKIVPLFAVEVLSEIQLSAGPGEGLYDEFETETVYSEDEYTG 138
            Y+E LL  Q       + +K   L+A  V     LSAG G   + +   + V+  DE
Sbjct: 65  YSERLLNHQIDKKSKDLIDKPSQLYAYRVYES--LSAGTGYSYFGDGNFDVVFY-DEQLE 121

Query: 139 FDIATWISGNSMEPVYKDGEVALIRSTGFDHDGAVYALNWNGSLYIKKLYREEDGFRMVS 198
            +D A+W+ G+SMEP Y +GEV LI+   FD+DGA+YA+ W+G  YIKK++RE++G R+VS
Sbjct: 122 YDFASWVFGDSMEPTYLNGEVVLIKQNSFDYDGAIYAVEWDGQTYIKKVFREDEGLRLVS 181

Query: 199 INPDVAERFIPFEDEIRIVGKIVGHFMPV 227
            +N   +++F P+ +E RI+GKI+ +F P+
Sbjct: 182 LNKKYSDKFAPYSEEPRIIGKIIANFRPL 210
```

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

Example 2107

A DNA sequence (GBSx2222) was identified in *S.agalactiae* <SEQ ID 6517> which encodes the amino acid sequence <SEQ ID 6518>. Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2387(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 2108

A DNA sequence (GBSx2223) was identified in *S.agalactiae* <SEQ ID 6519> which encodes the amino acid sequence <SEQ ID 6520>. This protein is predicted to be UmuC MucB homolog (uvrX). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2195(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 9925> which encodes amino acid sequence <SEQ ID 9926> was also identified.

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

```
>GP:AAC98439 GB:L29324 UmuC MucB homolog [Streptococcus pneumoniae]
 Identities = 303/436 (69%), Positives = 360/436 (82%)

Query: 39  LHTSLCVMSRADNSAGLILASSPMFKKVFGKGNVGRAYDLPFDVHTRKFNYYRAKISGLP 98
           L   LCVMSRADNSAGLILASSPMFKKVFGK NVGR+YDLPFDV TRKF+YY AK  GLP
Sbjct: 5   LRLRLCVMSRADNSAGLILASSPMFKKVFGKSNVGRSYDLPFDVKTRKFSYYNAKKQGLP 64

Query: 99  TDAKFVSFIENWAKRTFIVPPRMDLYIQKNLEIQKVFQNYADPTDILPYSIDEGFIDLTS 158
           T   +V +IE WAK T IVP    L I  N+EIQK+FQ++A P DI PYSIDEGFIDLTS
Sbjct: 65  TTIDYVRYIEEWAKSTVIVPREWILTIAVNMEIQKIFQDFAAPDDIYPYSIDEGFIDLTS 124

Query: 159 SLNYFVEDKSLSRKDKLDVVSAKIQHDIWEKTGVYSTVGMSNANPLLAKLALDNEAKTTA 218
           SLNYFV DKS+SRKDKLD++SA IQ  IW KTG+YSTVGMSNANPLLAKLALDNEAK T
Sbjct: 125 SLNYFVPDKSISRKDKLDIISAAIQKKIWRKTGIYSTVGMSNANPLLAKLALDNEAKKTP 184

Query: 219 TMRANWSYEDVETKVWNIPKMTDFWGIGSRTEKRLNKLGIYSIKELANCDPTILKKEFGV 278
           TMRANWSYEDVE KVW IPKMTDFWGIG+R EKRL+ LGI+SIKELA  +P ++KKE G+
Sbjct: 185 TMRANWSYEDVEKKVWTIPKMTDFWGIGNRMEKRLHNLGIFSIKELAQANPDLIKKELGI 244

Query: 279 IGVQHWFHANGIDESNVHEPYRPKAVGIGNSQVLHKDYTRQSDIELVLREMAEQVAIRLR 338
           +G++ WFHANGIDESNVH+PY+PK+ GIGNSQVL KDY +Q DIE++LREMAEQVA+RLR
Sbjct: 245 MGLELWFHANGIDESNVHKPYKPKSKGIGNSQVLPKDYIKQRDIEIILREMAEQVAVRLR 304

Query: 339 RRHKKATVVAINVGYSNFENKKSINVQRKINPNNRTLVFQDEVVSLFRSKYDGGAVRSIA 398
           R  KKATVV+I++GYS  E K+SIN Q KI P N+T +  + V+ LF +KY  GA+R++A
Sbjct: 305 RSGKKATVVSIHLGYSKVEQKRSINTQMKIEPTNQTALLTNYVLKLFHTKYTSGAIRNVA 364

Query: 399 VRYDGLVDENFAVISLFDDFEESEKEEKLETTIDSIRDRFGFLAVQKASSLLENSRAISR 458
           V Y GLVDE+F +ISLFDD E+ EKEE+L++ ID+IR  FGF ++ K ++L + SR I+R
Sbjct: 365 VNYSGLVDESFGLISLFDDIEKIEKEERLQSAIDAIRTEFGFTSLLKGNALDQASRTIAR 424

Query: 459 SRLVGGHSAGGLEGLK 474
           S+L+GGHSAGGL+GLK
```

```
Sbjct: 425 SKLIGGHSAGGLDGLK 440
```

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

Example 2109

A DNA sequence (GBSx2224) was identified in *S.agalactiae* <SEQ ID 6521> which encodes the amino acid sequence <SEQ ID 6522>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4016(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 2110

A DNA sequence (GBSx2225) was identified in *S.agalactiae* <SEQ ID 6523> which encodes the amino acid sequence <SEQ ID 6524>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2088(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:AAG13001 GB:AF227520 unknown [Streptococcus pneumoniae]
 Identities = 68/122 (55%), Positives = 89/122 (72%), Gaps = 6/122 (4%)

Query: 1   MIDRSYLPFKVAREYQDRKMAKWMGFFLSEHTAGLDSELNKVDYTSELSISDKLLLLNQL 60
           MIDRSYLPF+ AREYQD KM KWMGFFLSEHT+ L  + NKV Y S+LS+  KLLLL+Q+
Sbjct: 1   MIDRSYLPFQSAREYQDTKMQKWMGFFLSEHTSALTDDANKVTYMSDLSLEKKLLLLSQV 60

Query: 61  YSNQLNGIIAVPGQ----YYSGKVDNLTFNHVSLKTKTGFVSIPIKDILSIDL--EVEYE 114
           Y+ QLN  I V  +     Y+G + +LT + + +KT TG +++ +KDI+SI+L  EV YE
Sbjct: 61  YAGQLNTRIHVVKKNNQVSYTGTIPSLTKDFILIKTTTGHINLKLKDIVSIELVEEVLYE 120

Query: 115 SA 116
           SA
Sbjct: 121 SA 122
```

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 2111

A DNA sequence (GBSx2226) was identified in *S.agalactiae* <SEQ ID 6525> which encodes the amino acid sequence <SEQ ID 6526>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4025(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 9927> which encodes amino acid sequence <SEQ ID 9928> was also identified.

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

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

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

Example 2112

A DNA sequence (GBSx2227) was identified in *S.agalactiae* <SEQ ID 6527> which encodes the amino acid sequence <SEQ ID 6528>. This protein is predicted to be soluble transducer HtrXIII. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.5246(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 2113

A DNA sequence (GBSx2228) was identified in *S.agalactiae* <SEQ ID 6529> which encodes the amino acid sequence <SEQ ID 6530>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.5131(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 2114

A DNA sequence (GBSx2229) was identified in *S.agalactiae* <SEQ ID 6531> which encodes the amino acid sequence <SEQ ID 6532>. This protein is predicted to be pXO2-78. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2105(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:AAF13682 GB:AF188935 pXO2-78 [Bacillus anthracis]
 Identities = 101/314 (32%), Positives = 147/314 (46%), Gaps = 46/314 (14%)

Query: 27  SGQIYEHPDHDSFRIFADTNTFKWFSRDIQGDVIDFVQLVAGVSFKKALSYLETG--GFE 84
           S + Y   +HDS  I    N F W SR + G++I FVQ V   SF  A+  L  G   +E
Sbjct: 39  SERYYRLTEHDSLIIDRKKNQFYWNSRGVNGNIIKFVQEVEDASFPGAMQRLLDGEQDYE 98

Query: 85  EAKVIEETYQPFQYYLREEP----FQQARTYLKDIRGLSNQTINSFGRQGLLAQATYQAE 140
           +A  I    +P+ Y   E+     F +AR YL + R +  Q +++   +GL+ Q  Y
Sbjct: 99  KASEITFVSEPYDYEHFEQKEVSRFDRAREYLIEERKIDPQVVDALHNKGLIKQDKYN-- 156

Query: 141 SVLVFKSFDHNGTLQAASLQGLVKNEEKYDRGYLKKIMKGSHGHVGISFDIGNPKRLIFC 200
           +VL        G +   S QG+VK++ KY RG  K I K S  + G +   G P+ L F
Sbjct: 157 NVLFLWKDRETGAVMGGSEQGVVKSD-KYKRGAWKSIQKNSTANYGFNVLNGEPRNLKFY 215

Query: 201 ESVIDMMSYYQLHQKQLSDVRLISMEGLKLSVIAYQTLRLAAEEQGKLAFLDTVKPIRLS 260
           ES ID++SY  LH+  L D  LISMEGLK  VI                        +
Sbjct: 216 ESDIDLLSYATLHKHNLKDTHLISMEGLKPQVI------------------------FN 250

Query: 261 HYLQAIQETTTFFQTHSNVITMAVDNDEAGREFYQKL-------SDKGFPIFQ-DLPPLQ 312
           +Y++A +          + +++ VDND+AG+ F ++L       +D     F+ + P
Sbjct: 251 YYMKACERIGDV----PDSLSLCVDNDKAGKAFVERLIHFRYEKNDGSIVAFKPEYPQAP 306

Query: 313 RLETKSDWNDIVKR 326
             E K DWND  KR
Sbjct: 307 SEEKKWDWNDECKR 320
```

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 2115

A DNA sequence (GBSx2230) was identified in *S.agalactiae* <SEQ ID 6533> which encodes the amino acid sequence <SEQ ID 6534>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.7013(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 2116

A DNA sequence (GBSx2231) was identified in *S.agalactiae* <SEQ ID 6535> which encodes the amino acid sequence <SEQ ID 6536>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1310(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 2117

A DNA sequence (GBSx2232) was identified in *S.agalactiae* <SEQ ID 6537> which encodes the amino acid sequence <SEQ ID 6538>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.6726(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 9373> which encodes amino acid sequence <SEQ ID 9374> was also identified.

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

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

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

Example 2118

A DNA sequence (GBSx2233) was identified in *S.agalactiae* <SEQ ID 6539> which encodes the amino acid sequence <SEQ ID 6540>. This protein is predicted to be phosphoglucomutase (manB). Analysis of this protein sequence reveals the following:

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

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2147(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 9355> which encodes amino acid sequence <SEQ ID 9356> was also identified.

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

```
>GP:CAB96418 GB:AJ243290 phosphoglucomutase [Streptococcus thermophilus]
 Identities = 391/465 (84%), Positives = 424/465 (91%), Gaps = 1/465 (0%)

Query: 1   MAQHGIKSYVFEALRPTPELSFAVRHLNAYAGIMVTASHNPAPFNGYKVYGQDGGQLPPA 60
            +A HGIKSYVFE+LRPTPELSFAVRHL+ +AGIM+TASHNPAPFNGYKVYG+DGGQ+PPA
Sbjct: 107 LAAHGIKSYVFESLRPTPELSFAVRHLHTFAGIMITASHNPAPFNGYKVYGEDGGQMPPA 166

Query: 61  DADALTDFIRAIENPFAVELADLDESKSSGLIQVIGEDVDIEYLREVKDVNINQDLINNF 120
            DADALTD+IRAI+NPF V+LADL++SK+SGLI++IGE+VD EYL+EVKDVNINQDLIN +
Sbjct: 167 DADALTDYIRAIDNPFTVKLADLEDSKASGLIEIIGENVDAEYLKEVKDVNINQDLINEY 226

Query: 121 GKDMKIVYTPLHGTGEMLTRRALAQAGFESVVVVESQAKADPDFSTVKSPNPESQAAFAL 180
            G+DMKIVYT LHGTGEML RRALAQAGF++V VVE+QA     DF TVKSPNPE+Q AFAL
Sbjct: 227 GRDMKIVYTSLHGTGEMLVRRALAQAGFDAVQVVEAQAVPHADFLTVKSPNPENQDAFAL 286

Query: 181 AEELGREVDADVLVATDPDADRLGVEIRQPDGSYKNLSGNQIGAIIAKYILEAHKTAGTL 240
            AEELGR VDADVLVATDPDADRLGVEIRQPDGSY NLSGNQIGAIIAKYILEAHKTAGTL
Sbjct: 287 AEELGRNVDADVLVATDPDADRLGVEIRQPDGSYLNLSGNQIGAIIAKYILEAHKTAGTL 346

Query: 241 PENAALAKSIVSTELVTKIAESYGATMFNVLTGFKFIAEKIQEFEEKHNHTYMFGFEESF 300
            P NAAL KSIVSTELVTKIAESYGATMFNVLTGFKFI EKI EFE +HN+TYMFGFEESF
Sbjct: 347 PANAALCKSIVSTELVTKIAESYGATMFNVLTGFKFIGEKIHEFETQHNYTYMFGFEESF 406

Query: 301 GYLIKPFVRDKDAIQAVLLVAEIAAYYRSRGLTLADGIDEIYKEYGYFAEKTISVTLSGV 360
            GYLIKPFVRDKDAIQAVL+VAEIAAYYRSRG+TLADGI+EIYK+YGYF+EKTISVTLSGV
Sbjct: 407 GYLIKPFVRDKDAIQAVLIVAEIAAYYRSRGMTLADGIEEIYKQYGYFSEKTISVTLSGV 466

Query: 361 DGAAEIKKIMDKFRENGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLKYTLA 420
            DGAAEIKKIMDKFR N PKQFNNTDI   EDF +QTAT  DG +  LTTPPSNVLKY LA
Sbjct: 467 DGAAEIKKIMDKFRRNAPKQFNNTDIAKTEDFLEQTATTADG-VEKLTTPPSNVLKYILA 525

Query: 421 DDSWIAVRPSGTEPKIKFYIATVGNDLADAETKIANIEKEITTFV 465
            DDSW AVRPSGTEPKIKFYIATVG   ADA+ KIANIE EI  FV
Sbjct: 526 DDSWFAVRPSGTEPKIKFYIATVGETEADAKEKIANIEAEINAFV 570
```

There is also homology to SEQ ID 6156:

```
Query: 1   MAQHGIKSYVFEALRPTPELSFAVRHLNAYAGIMVTASHNPAPFNGYKVYGQDGGQLPPA 60
            +AQHGIKSYVFEALRPTPELSFAVRHLNAYAGIMVTASHNPAPFNGYKVYGQDGGQLPPA
Sbjct: 107 LAQHGIKSYVFEALRPTPELSFAVRHLNAYAGIMVTASHNPAPFNGYKVYGQDGGQLPPA 166

Query: 61  DADALTDFIRAIENPFAVELADLDESKSSGLIQVIGEDVDIEYLREVKDVNINQDLINNF 120
            DADALTDFIRAIENPFAVELADLDE+KSSGLIQVIGEDVD+EYLREVKDVNINQDLINNF
Sbjct: 167 DADALTDFIRAIENPFAVELADLDENKSSGLIQVIGEDVDMEYLREVKDVNINQDLINNF 226

Query: 121 GKDMKIVYTPLHGTGEMLTRRALAQAGFESVVVVESQAKADPDFSTVKSPNPESQAAFAL 180
            GKDMKIVYTPLHGTGEMLTRRALAQAGFESVVVVESQAKADPDFSTVKSPNPESQAAFAL
Sbjct: 227 GKDMKIVYTPLHGTGEMLTRRALAQAGFESVVVVESQAKADPDFSTVKSPNPESQAAFAL 286

Query: 181 AEELGREVDADVLVATDPDADRLGVEIRQPDGSYKNLSGNQIGAIIAKYILEAHKTAGTL 240
            AEELGREV+ADVLVATDPDADRLGVEIRQPDGSYKNLSGNQIGAIIAKYILEAHKTAGTL
Sbjct: 287 AEELGREVEADVLVATDPDADRLGVEIRQPDGSYKNLSGNQIGAIIAKYILEAHKTAGTL 346

Query: 241 PENAALAKSIVSTELVTKIAESYGATMFNVLTGFKFIAEKIQEFEEKHNHTYMFGFEESF 300
            PENAALAKSIVSTELVTKIAESYGATMFNVLTGFKFIAEKIQEFEEKHNHTYMFGFEESF
Sbjct: 347 PENAALAKSIVSTELVTKIAESYGATMFNVLTGFKFIAEKIQEFEEKHNHTYMFGFEESF 406
```

```
Query: 301 GYLIKPFVRDKDAIQAVLLVAEIAAYYRSRGLTLADGIDEIYKEYGYFAEKTISVTLSGV 360
           GYLIKPFVRDKDAIQAVLLVAEIAAYYRSRGLTLADGIDEIYKEYGYFAEKTISVTLSGV
Sbjct: 407 GYLIKPFVRDKDAIQAVLLVAEIAAYYRSRGLTLADGIDEIYKEYGYFAEKTISVTLSGV 466

Query: 361 DGAAEIKKIMDKFRENGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLKYTLA 420
           DGAAEIKKIMDKFRENGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLKYTLA
Sbjct: 467 DGAAEIKKIMDKFRENGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLKYTLA 526

Query: 421 DDSWIAVRPSGTEPKIKFYIATVGNDLADAETKIANIEKEITTFV 465
           DDSWIAVRPSGTEPKIKFYIAT+G+ L  A+ KIANIE EI TFV
Sbjct: 527 DDSWIAVRPSGTEPKIKFYIATIGDTLDIAQEKIANIETEINTFV 571
```

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

Example 2119

A DNA sequence (GBSx2235) was identified in *S.agalactiae* <SEQ ID 6541> which encodes the amino acid sequence <SEQ ID 6542>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1564(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 9905> which encodes amino acid sequence <SEQ ID 9906> was also identified. There is also homology to SEQ ID 32.

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

Example 2120

A DNA sequence (GBSx2236) was identified in *S.agalactiae* <SEQ ID 6543> which encodes the amino acid sequence <SEQ ID 6544>. This protein is predicted to be ABC transporter, ATP-binding protein (msbA). Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.92   Transmembrane  162 - 178 ( 135 - 184)
   INTEGRAL    Likelihood = -7.11   Transmembrane   58 -  74 (  56 -  78)
   INTEGRAL    Likelihood = -6.42   Transmembrane  136 - 152 ( 135 - 161)
   INTEGRAL    Likelihood = -5.20   Transmembrane   23 -  39 (  21 -  49)
   INTEGRAL    Likelihood = -1.75   Transmembrane  485 - 501 ( 485 - 501)

----- Final Results -----
            bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAD35376 GB:AE001710 ABC transporter, ATP-binding protein
          [Thermotoga maritima]
 Identities = 216/552 (39%), Positives = 336/552 (60%), Gaps = 3/552 (0%)

Query: 26  MALLGTVVQVCLTVYLPVLIGQAVDVVLSPHSMILLLPIMWKMIAVILANTIIQWINPLL 85
           M +   V   L V  P LIG+ +DVV  P    LL   M  +  +    +++ W+   +
```

```
Sbjct: 41  MVFVFVTVSSILGVLSPYLIGKTIDVVFVPRRFDLLPRYMLILGTIYALTSLLFWLQGKI 100

Query: 86  YNRLIFHYVASLRKAVMEKLNLLPIAYLDKRGIGDLISRVTTDTEQLSNGLLMVFNQFFV 145
              L    V  LRK + EKL  +P+ + D+   GD+ISRV  D + ++N L     QFF
Sbjct: 101 MLTLSQDVVFRLRKELFEKLQRVPVGFFDRTPHGDIISRVINDVDNINNVLGNSIIQFFS 160

Query: 146 GLLTIIVTIFSMAKIDLLMLFLVLFLIPLSLFLARFIAKKSY-HLYQNQTASRGRQTQFI 204
           G++T+   + M ++++++  + L + PL++ + + ++ ++  + Y+NQ    G+    I
Sbjct: 161 GIVTLAGAVIMMFRVNVILSLVTLSIVPLTVLITQIVSSQTRKYFYENQRVL-GQLNGII 219

Query: 205 EEMVSQESLIQAFSAQEESSDHFRTINQEYANFSQSAIFYSSTVNPSTRFINSLIYGFLA 264
           EE +S  ++I+ F+ +E+  + F  +N+         A  +S  + P     +N+L +  ++
Sbjct: 220 EEDISGLTVIKLFTREEKEMEKFDRVNESLRKVGTKAQIFSGVLPPLMNMVNNLGFALIS 279

Query: 265 GIGALRIMSGAFSVGQLITFLNYVNQYTKPFNDISSVLSEMQSALACAERLYSILEESSP 324
           G G    +    +VG + TF+ Y  Q+T+P N++S+  + +Q ALA AER++ IL+
Sbjct: 280 GFGGWLALKDIITVGTIATFIGYSRQFTRPLNELSNQFNMIQMALASAERIFEILDLEEE 339

Query: 325 NITGTEKLDSSTVKGQIDFKNVVFGYNKSKLLLNGINLHIPAGAKVAIVGPTGAGKSTLI 384
                + ++   V+G+I+FKNV F Y+K K +L  I  HI  G KVA+VGPTG+GK+T++
Sbjct: 340 K-DDPDAVELREVRGEIEFKNVWFSYDKKKPVLKDITFHIKPGQKVALVGPTGSGKTTIV 398

Query: 385 NLIMRFYEVDGGNILLDCKPITDYEPSQLRQEIGMVLQETWLKSATIHDNIAYANPKASR 444
           NL+MRFY+VD G IL+D   I   + S LR  IG+VLQ+T L S T+ +N+ Y NP A+
Sbjct: 399 NLLMRFYDVDRGQILVDGIDIRKIKRSSLRSSIGIVLQDTILFSTTVKENLKYGNPGATD 458

Query: 445 EEVIEAAKAANADFFIKQLPNGYDTYLEDAGDSLSQGQCQLLTIARIFLKLPRILILDEA 504
           EE+ EAAK  ++D FIK LP GY+T L D G+ LSQGQ QLL I R FL  P+ILILDEA
Sbjct: 459 EEIKEAAKLTHSDHFIKHLPEGYETVLTDNGEDLSQGQRQLLAITRAFLANPKILILDEA 518

Query: 505 TSSIDTRTEVLVQEAFQMLMKGRTSFIIAHRLSTIQTADIILVMVSGEIVEVGNHSELMA 564
           TS++DT+TE  +Q A   LM+G+TS IIAHRL+TI+ AD+I+V+  GEIVE+G H EL+
Sbjct: 519 TSNVDTKTEKSIQAAMWKLMEGKTSIIIAHRLNTIKNADLIIVLRDGEIVEMGKHDELIQ 578

Query: 565 QKGIYYQMQNAQ 576
           ++G YY++  +Q
Sbjct: 579 KRGFYYELFTSQ 590
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6545> which encodes the amino acid sequence <SEQ ID 6546>. Analysis of this protein sequence reveals the following:

```
     Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.07   Transmembrane  162 - 178 ( 159 - 182)
   INTEGRAL    Likelihood = -7.17   Transmembrane  143 - 159 ( 137 - 161)
   INTEGRAL    Likelihood = -5.84   Transmembrane   23 -  39 (  19 -  45)
   INTEGRAL    Likelihood = -5.68   Transmembrane   68 -  84 (  60 -  86)
   INTEGRAL    Likelihood = -2.55   Transmembrane  261 - 277 ( 256 - 278)

----- Final Results -----
              bacterial membrane --- Certainty=0.4227(Affirmative) < succ>
               bacterial outside --- Certainty=0,0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAD35376 GB:AE001710 ABC transporter, ATP-binding protein
          [Thermotoga maritima]
 Identities = 206/572 (36%), Positives = 342/572 (59%), Gaps = 5/572 (0%)

Query: 2   IKTDHHLLKRVLQDLLKKPLPVCILVIASFVQVG--LSVYLPVLIGKAVDMSLSVNSWQT 59
           +K     L+R+L  L  +P     ++++  FV V   L V  P LIGK +D+      +
Sbjct: 18  LKNPTATLRRLLGYL--RPHTFTLIMVFVFVTVSSILGVLSPYLIGKTIDVVFVPRRFDL 75

Query: 60  LKWLLGQMLVIIVVNTLIQWVMPLVYSRLLYQYSQQLKDKLLEKIHRLPFAYLDRQTIGD 119
           L   +  +  I  + +L+ W+    +    L       +L+ +L EK+ R+P  + DR   GD
Sbjct: 76  LPRYMLILGTIYALTSLLFWLQGKIMLTLSQDVVFRLRKELFEKLQRVPVGFFDRTPHGD 135
```

```
Query: 120 LVSRVITDTEQLINGLQMVFNQFILGLLTILCTIIAMAQIDWLMLILVLVLTPSSLFLAR 179
           ++SRVI D + + N L     QF  G++T+   +I M +++ ++ ++ L + P ++ + +
Sbjct: 136 IISRVINDVDNINNVLGNSIIQFFSGIVTLAGAVIMMFRVNVILSLVTLSIVPLTVLITQ 195

Query: 180 FIAQKSFHYAQAQTKSRGNLAQFTEEILRQEGLVQLFNAQEQSICDYHVLNKTYCEASQK 239
            ++ ++  Y     +  G L    EE +    +++LF  +E+ +  +  +N++  +   K
Sbjct: 196 IVSSQTRKYFYENQRVLGQLNGIIEEDISGLTVIKLFTREEKEMEKFDRVNESLRKVGTK 255

Query: 240 AIFYASTVNPATRFINSVIYALLAGLGAVRIMAGLFSVGQLTTFLNVVVQYTKPFNDISS 299
           A  ++  + P    +N++ +AL++G G     +  +VG + TF+    Q+T+P N++S+
Sbjct: 256 AQIFSGVLPPLMNMVNNLGFALISGFGGWLALKDIITVGTIATFIGYSRQFTRPLNELSN 315

Query: 300 VLAEIQSSLACAQRLYDLLDIEIKEQEHFLTFKASAVKGQIDFEEVSFSYQKDRPLLKDI 359
               IQ +LA A+R++++LD+E +E++     +   V+G+I+F+ V FSY K +P+LKDI
Sbjct: 316 QFNMIQMALASAERIFEILDLE-EEKDDPDAVELREVRGEIEFKNVWFSYDKKKPVLKDI 374

Query: 360 NFSVPAGSKVAIVGPTGAGKSTLINLLMRFYELDAGSIKLDKVPIKCYAKEELRSITGIV 419
            F +  G KVA+VGPTG+GK+T++NLLMRFY++D G I +D + I+   +  LRS  GIV
Sbjct: 375 TFHIKPGQKVALVGPTGSGKTTIVNLLMRFYDVDRGQILVDGIDIRKIKRSSLRSSIGIV 434

Query: 420 LQETWLKDATVHELIAYGSEEASRDEVVAAAKAAHAHFFIMQLPKTYDTYLSASDDALSQ 479
           LQ+T L   TV E + YG+  A+ +E+  AAK  H+  FI  LP+ Y+T L+ + + LSQ
Sbjct: 435 LQDTILFSTTVKENLKYGNPGATDEEIKEAAKLTHSDHFIKHLPEGYETVLTDNGEDLSQ 494

Query: 480 GQLQLLAIARMFLKKPKVLVLDEATSSIDIRTEAVIQEALKELMRGRTSFIIAHRLSTIQ 539
           GQ QLLAI R FL  PK+L+LDEATS++D +TE  IQ A+ +LM G+TS IIAHRL+TI+
Sbjct: 495 GQRQLLAITRAFLANPKILILDEATSNVDTKTEKSIQAAMWKLMEGKTSIIIAHRLNTIK 554

Query: 540 SADLILVMDQGRLVEWGTHASLMSKNGCYVRL 571
           +ADLI+V+  G +VE G H  L+ K G Y  L
Sbjct: 555 NADLIIVLRDGEIVEMGKHDELIQKRGFYYEL 586
```

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

```
Identities = 340/566 (60%), Positives = 433/566 (76%)

Query: 11  KKLVQDLLSKKSLVGMALLGTVVQVCLTVYLPVLIGQAVDVVLSPHSMILLLPIMWKMIA 70
           K+++QDLL K   V + ++ + VQV L+VYLPVLIG+AVD+ LS +S   L  ++ +M+
Sbjct: 10  KRVLQDLLKKPLPVCILVIASFVQVGLSVYLPVLIGKAVDMSLSVNSWQTLKWLLGQMLV 69

Query: 71  VILANTIIQWINPLLYNRLIFHYVASLRKAVMEKLNLLPIAYLDKRGIGDLISRVTTDTE 130
           +I+ NT+IQW+ PL+Y+RL++ Y   L+  ++EK++ LP AYLD++ IGDL+SRV TDTE
Sbjct: 70  IIVVNTLIQWVMPLVYSRLLYQYSQQLKDKLLEKIHRLPFAYLDRQTIGDLVSRVITDTE 129

Query: 131 QLSNGLLMVFNQFFVGLLTIIVTIFSMAKIDLLMLFLVLFLTPLSLFLARFIAKKSYHLY 190
           QL NGL MVFNQF +GLLTI+ TI +MA+ID LML LVL LTP SLFLARFIA+KS+H
Sbjct: 130 QLINGLQMVFNQFILGLLTILCTIIAMAQIDWLMLILVLVLTPSSLFLARFIAQKSFHYA 189

Query: 191 QNQTASRGRQTQFIEEMVSQESLIQAFSAQEESSDHFRTINQEYANFSQSAIFYSSTVNP 250
           Q QT SRG   QF EE++ QE L+Q F+AQE+S   +  +N+ Y   SQ AIFY+STVNP
Sbjct: 190 QAQTKSRGNLAQFTEEILRQEGLVQLFNAQEQSICDYHVLNKTYCEASQKAIFYASTVNP 249

Query: 251 STRFINSLIYGFLAGIGALRIMSGAFSVGQLITFLNYVNQYTKPFNDISSVLSEMQSALA 310
           +TRFINS+IY  LAG+GA+RIM+G FSVGQL TFLN V QYTKPFNDISSVL+E+QS+LA
Sbjct: 250 ATRFINSVIYALLAGLGAVRIMAGLFSVGQLTTFLNVVVQYTKPFNDISSVLAEIQSSLA 309

Query: 311 CAERLYSILEESSPNITGTEKLDSSTVKGQIDFKNVVFGYNKSKLLLNGINLHIPAGAKV 370
           CA+RLY +L+              +S VKGQIDF+ V F Y K + LL  IN  +PAG+KV
Sbjct: 310 CAQRLYDLLDIEIKEQEHFLTFKASAVKGQIDFEEVSFSYQKDRPLLKDINFSVPAGSKV 369

Query: 371 AIVGPTGAGKSTLINLIMRFYEVDGGNILLDCKPITDYEPSQLRQEIGMVLQETWLKSAT 430
           AIVGPTGAGKSTLINL+MRFYE+D G+I LD  PI  Y   +LR   G+VLQETWLK AT
Sbjct: 370 AIVGPTGAGKSTLINLLMRFYELDAGSIKLDKVPIKCYAKEELRSITGIVLQETWLKDAT 429

Query: 431 IHDNIAYANPKASREEVIEAAKAANADFFIKQLPNGYDTYLEDAGDSLSQGQCQLLTIAR 490
           +H+ IAY + +ASR+EV+ AAKAA+A FFI QLP  YDTYL  + D+LSQGQ QLL IAR
Sbjct: 430 VHELIAYGSEEASRDEVVAAAKAAHAHFFIMQLPKTYDTYLSASDDALSQGQLQLLAIAR 489
```

```
Query: 491 IFLKLPRILILDEATSSIDTRTEVLVQEAFQMLMKGRTSFIIAHRLSTIQTADIILVMVS 550
          +FLK P++L+LDEATSSID RTE ++QEA + LM+GRTSFIIAHRLSTIQ+AD+ILVM
Sbjct: 490 MFLKKPKVLVLDEATSSIDIRTEAVIQEALKELMRGRTSFIIAHRLSTIQSADLILVMDQ 549

Query: 551 GEIVEVGNHSELMAQKGIYYQMQNAQ 576
           G +VE G H+ LM++ G Y ++Q  +
Sbjct: 550 GRLVEWGTHASLMSKNGCYVRLQKIE 575
```

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

Example 2121

A DNA sequence (GBSx2237) was identified in *S.agalactiae* <SEQ ID 6547> which encodes the amino acid sequence <SEQ ID 6548>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1099(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 2122

A DNA sequence (GBSx2238) was identified in *S.agalactiae* <SEQ ID 6549> which encodes the amino acid sequence <SEQ ID 6550>. This protein is predicted to be ABC transporter, ATP-binding protein (msbA). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.69   Transmembrane  157 - 173 ( 130 - 182)
   INTEGRAL    Likelihood =-10.88   Transmembrane   56 -  72 (  49 -  77)
   INTEGRAL    Likelihood = -7.75   Transmembrane  239 - 255 ( 235 - 258)
   INTEGRAL    Likelihood = -6.42   Transmembrane  133 - 149 ( 130 - 156)
   INTEGRAL    Likelihood = -4.78   Transmembrane  271 - 287 ( 270 - 289)
   INTEGRAL    Likelihood = -1.91   Transmembrane   20 -  36 (  20 -  37)

----- Final Results -----
             bacterial membrane --- Certainty=0.6477(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAD35375 GB:AE001710 ABC transporter, ATP-binding protein
          [Thermotoga maritima]
 Identities = 196/570 (34%), Positives = 327/570 (56%), Gaps = 5/570 (0%)

Query: 1   MKRLTYYFKGYIKETIFGPLFKLLEASFELLVPIVIAKMIDETIPRGDRSGLLLQIGLIF 60
           MK L  Y K Y    +  PLF ++E   +L  P ++A+++DE I RGD S L+L+ G++
Sbjct: 1   MKTLARYLKPYWIFAVLAPLFMVVEVICDLSQPTLLARIVDEGIARGDFS-LVLKTGILM 59

Query: 61  FLAA-VGVVVAITAQYYSSKAAVGYTRQLTEDLYQKVMSLGKKDRDELGTASLITRLTAD 119
```

```
            + A +G V  I    ++S A+  +   L  DL++KV+S    + +   T+SLITRLT D
Sbjct: 60   LIVALIGAVGGIGCTVFASYASQNFGADLRRDLFRKVLSFSISNVNRFHTSSLITRLTND 119

Query: 120 TFQIQTGLNQFLRLFLRAPIIVFGAIIMAFSISPSLTIWFLVMVVTLFIIVFVMSRLLNP 179
             Q+Q  +   LR+ +RAP++  G I+MA SI+  L+   + ++  + ++   +++  NP
Sbjct: 120 VTQLQNLVMMLLRIVVRAPLLFVGGIVMAVSINVKLSSVLIFLIPPIVLLFVWLTKKGNP 179

Query: 180 IYLKIRTSTDYLVKLTRQQLQGVRVIRAFNQVDRESEAFNDINYHYTNLQLKAGRLSSLV 239
            ++ KI+ STD + ++ R+ L GVRV+RAF + + E+E F   N       + A  L
Sbjct: 180 LFRKIQESTDEVNRVVRENLLGVRVVRAFRREEYENENFRKANESLRRSIISAFSLIVFA 239

Query: 240 TPLTFLVVNITLVVIIWRGNLNIANHLLSQGMLVALINYLLQILVELLKMTMLVTSLNQS 299
             PL   +VN+ ++ ++W G + + N+ +  G ++A  NYL+QI+  L+ +  ++  + ++
Sbjct: 240 LPLFIFIVNMGMIAVLWFGGVLVRNNQMEIGSIMAYTNYLMQIMFSLMMIGNILNFIVRA 299

Query: 300 YISAKRIIAVF-ERPS-EIIDDKLEPKYSNKALEVQEMAFSYPNSSEKALSDITFSMNVG 357
             SAKR++ V  E+P+ E  D+ L       ++  + + F Y  +++  LS + FS+  G
Sbjct: 300 SASAKRVLEVLNEKPAIEEADNALALPNVEGSVSFENVEFRYFENTDPVLSGVNFSVKPG 359

Query: 358 ETLGIIGGTGSGKSTLINLLLHIYKVQEGDIDIYHQGKSPDTISNWRTLVRVVPQNAQLF 417
             + ++G TGSGKSTL+NL+  +   + G +++          + + R  +  VPQ   LF
Sbjct: 360 SLVAVLGETGSGKSTLMNLIPRLIDPERGRVEVDELDVRTVKLKDLRGHISAVPQETVLF 419

Query: 418 KGTIRSNLSLGLGKVSEEKLWTALEIAQASDFVKEKDGQLDAPVESFGRNFSGGQRQRLT 477
             GTI+ NL  G    +++++  A +IAQ  DF+       D+ VE  GRNFSGGQ+QRL+
Sbjct: 420 SGTIKENLKWGREDATDDEIVEAAKIAQIHDFIISLPEGYDSRVERGGRNFSGGQKQRLS 479

Query: 478 IARALVQDKIPFLILDDATSALDYLTEARLFKAITKHFNQTNLIIVSQRINSIQNADRIL 537
            IARALV+ K   LILDD TS++D +TE R+   + ++       I++Q+I +   AD+IL
Sbjct: 480 IARALVK-KPKVLILDDCTSSVDPITEKRILDGLKRYTKGCTTFIITQKIPTALLADKIL 538

Query: 538 LLDKGKQVGFDNHQSLLAHNKVYKSIYHSQ 567
            +L +GK  GF  H+ LL H K Y+ IY SQ
Sbjct: 539 VLHEGKVAGFGTHKELLEHCKPYREIYESQ 568
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6551> which encodes the amino acid sequence <SEQ ID 6552>. Analysis of this protein sequence reveals the following:

```
     Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.47   Transmembrane  157 - 173 ( 149 - 185)
   INTEGRAL    Likelihood = -7.75   Transmembrane   55 -  71 (  51 -  74)
   INTEGRAL    Likelihood = -4.25   Transmembrane  239 - 255 ( 237 - 260)
   INTEGRAL    Likelihood = -3.77   Transmembrane   20 -  36 (  19 -  37)
   INTEGRAL    Likelihood = -3.50   Transmembrane  271 - 287 ( 270 - 288)
   INTEGRAL    Likelihood = -2.55   Transmembrane  133 - 149 ( 130 - 151)

----- Final Results -----
              bacterial membrane --- Certainty=0.5989(Affirmative) < succ>
               bacterial outside --- 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:AL137187 putative ABC transporter [Streptomyces ...    296  6e-79

>GP:CAB69751 GB:AL137187 putative ABC transporter [Streptomyces
          coelicolor A3(2)]
 Identities = 185/569 (32%), Positives = 306/569 (53%), Gaps = 8/569 (1%)

Query: 1   MKRLRPYVKGYLKESILGPLFKLLEALFELLVPLLIANMIDISISQHNSQGILRVVLTLF 60
           ++ LR Y++ Y K   L    + L+    L +P L A++ID  + + +S  IL     +
Sbjct: 3   IRLLRTYLRPYKKPIALLVALQFLQTCASLYLPTLNAHIIDEGVVKGDSGYILSYGALMI 62

Query: 61  GLATIGLLLSVTAQYFSSKAAVGFTRQMTDDLFKKIMFLSKEDQDHLGYASLLSRLTSDS 120
           G++   ++ ++ A ++ ++ A    R +   +F ++   S  +  H G  SL++R T+D
Sbjct: 63  GISLAQVVCNIGAVFYGARTAAALGRDVRGAVFDRVQSFSAREVGHFGAPSLITRTTNDV 122
```

```
Query: 121 FQIQTGINQFLRLFLRAPIIVCGAMVMAYWISPSLTLWFVMMVIVLLTLVFVMSHLLGPL 180
           Q+Q        L + API+  G +VMA  +    L+   + +V VL   V ++   L PL
Sbjct: 123 QQVQMLALMTFTLMVSAPIMCVGGIVMALGLDVPLSGVLLGVVPVLAICVTLIVRKLRPL 182

Query: 181 YLLIRRETDHLVRLTSQQLQGIRVIKAFNQTQKELQAFKQQNMLLSRHQYQAATLANVLN 240
           +  ++   D + R+  +Q+ G RVI+AF + + E Q F++ N  L+         L  ++
Sbjct: 183 FRKMQVRLDTVNRVLREQITGNRVIRAFVRDEYEQQRFRKANTELTEVALGTGNLLALMF 242

Query: 241 PMTFLVVNLTLLILIWQGSWQVAHRSLSQGMLVALINYLLQILAELLKMTMLMGTINQSV 300
           P+   VVNL+ + ++W G+ ++    + G L A + YL+QI+  ++  T +   + ++
Sbjct: 243 PVVMTVVNLSSIAVVWFGAHRIDSGGMQIGDLTAFLAYLMQIVMSVMMATFMFMMVPRAE 302

Query: 301 TAAKRINQVFVLADEAPLPLLKDGPISTH-LLTIRHLTFTYPGAAEPSLYDIQLSADQGE 359
             A+RI +V         P+    + H  L IR   F YPGA EP L  I L A  GE
Sbjct: 303 VCAERIQEVLETESSVVPPVAPVTELRRHGHLEIREAGFRYPGAEEPVLRHIDLVARPGE 362

Query: 360 WIGIIGGTGAGKTTLIDLICQTYSQYSGEISLNW---QGEVPKTLTEWRNVIALVPQKAQ 416
             +IG TG+GK+TL+ L+ + +    GE+ +N    +   PKTL +   V++LVPQK
Sbjct: 363 TTAVIGSTGSGKSTLLGLVPRLFDATDGEVLVNGVDVRTVDPKTLAK---VVSLVPQKPY 419

Query: 417 LFKGTIRSNLLLGQSMPISDEELWRALELAQAKEFVAALPEQLEAPVEAFGRHFSGGQRQ 476
           LF GT+ +NL  G +   +DEELW AL +AQAKEFV+ L  L+AP+   G + SGGQRQ
Sbjct: 420 LFAGTVATNLRYG-NPDATDEELWHALAVAQAKEFVSELEGGLDAPIAQGGTNVSGGQRQ 478

Query: 477 RLAIARALLKPKPILILDDASSALDNETRGRLFKALKEELSDVLVILVTQSIKNLQFADK 536
           RLAIAR L++   I + DD+ SALD  T   L   L +E ++  V++V Q +  ++ AD+
Sbjct: 479 RLAIARTLVQRPEIYLFDDSFSALDYATDAALRAELAQETAEATVVIVAQRVATIRDADR 538

Query: 537 ILVLEQGHQLDFASHDQLKVSNALYQEML 565
           I+VL++G +   H +L  N  Y+E++
Sbjct: 539 IVVLDEGRVVGVGRHHELMADNETYREIV 567
```

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

```
Identities = 313/568 (55%), Positives = 428/568 (75%), Gaps = 9/568 (1%)

Query: 1   MKRLTYYFKGYIKETIFGPLFKLLEASFELLVPIVIAKMIDETIPRGDRSGLLLQIGLIF 60
           MKRL  Y KGY+KE+I GPLFKLLEA FELLVP++IA MID +I + +  G+L  +  +F
Sbjct: 1   MKRLRPYVKGYLKESILGPLFKLLEALFELLVPLLIANMIDISISQHNSQGILRVVLTLF 60

Query: 61  FLAAVGVVVAITAQYYSSKAAVGYTRQLTEDLYQKVMSLGKKDRDELGTASLITRLTADT 120
            LA +G+++++TAQY+SSKAAVG+TRQ+T+DL++K+M L K+D+D LG ASL++RLT+D+
Sbjct: 61  GLATIGLLLSVTAQYFSSKAAVGFTRQMTDDLFKKIMFLSKEDQDHLGYASLLSRLTSDS 120

Query: 121 FQIQTGLNQFLRLFLRAPIIVFGAIIMAFSISPSLTIWFLVMVVTLFIIVFVMSRLLNPI 180
           FQIQTG+NQFLRLFLRAPIIV GA++MA+ ISPSLT+WF++MV+ L  +VFVMS LL P+
Sbjct: 121 FQIQTGINQFLRLFLRAPIIVCGAMVMAYWISPSLTLWFVMMVIVLLTLVFVMSHLLGPL 180

Query: 181 YLKIRTSTDYLVKLTRQQLQGVRVIRAFNQVDRESEAFNDINYHYTNLQLKAGRLSSLVT 240
           YL IR  TD+LV+LT QQLQG+RVI+AFNQ  +E +AF   N  +  Q +A  L++++
Sbjct: 181 YLLIRRETDHLVRLTSQQLQGIRVIKAFNQTQKELQAFKQQNMLLSRHQYQAATLANVLN 240

Query: 241 PLTFLVVNITLVVIIWRGNLNIANHLLSQGMLVALINYLLQILVELLKMTMLVTSLNQSY 300
           P+TFLVVN+TL+++IW+G+  +A+  LSQGMLVALINYLLQIL ELLKMTML+ ++NQS
Sbjct: 241 PMTFLVVNLTLLILIWQGSWQVAHRSLSQGMLVALINYLLQILAELLKMTMLMGTINQSV 300

Query: 301 ISAKRIIAVF----ERPSEIIDDKLEPKYSNKALEVQEMAFSYPNSSEKALSDITFSMNV 356
           .+AKRI  VF    E P  ++ D      S   L ++ + F+YP  ++E +L DI  S +
Sbjct: 301 TAAKRINQVFVLADEAPLPLLKD---GPISTHLLTIRHLTFTYPGAAEPSLYDIQLSADQ 357

Query: 357 GETLGIIGGTGSGKSTLINLLLHIYKVQEGDIDIYHQGKSPDTISNWRTLVRVVPQNAQL 416
           GE +GIIGGTG+GK+TLI+L+   Y   G+I +  QG+ P T++ WR ++ +VPQ AQL
Sbjct: 358 GEWIGIIGGTGAGKTTLIDLICQTYSQYSGEISLNWQGEVPKTLTEWRNVIALVPQKAQL 417

Query: 417 FKGTIRSNLSLGLG-KVSEEKLWTALEIAQASDFVKEKDGQLDAPVESFGRNFSGGQRQR 475
           FKGTIRSNL LG    +S+E+LW ALE+AQA +FV      QL+APVE+FGR+FSGGQRQR
Sbjct: 418 FKGTIRSNLLLGQSMPISDEELWRALELAQAKEFVAALPEQLEAPVEAFGRHFSGGQRQR 477
```

```
Query: 476 LTIARALVQDKIPFLILDDATSALDYLTEARLFKAITKHFNQTNLIIVSQRINSIQNADR 535
          L IARAL++ K P LILDDA+SALD  T  RLFKA+ +  +   +I+V+Q I ++Q AD+
Sbjct: 478 LAIARALLKPK-PILILDDASSALDNETRGRLFKALKEELSDVLVILVTQSIKNLQFADK 536

Query: 536 ILLLDKGKQVGFDNHQSLLAHNKVYKSI 563
          IL+L++G Q+ F +H  L   N +Y+ +
Sbjct: 537 ILVLEQGHQLDFASHDQLKVSNALYQEM 564
```

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

Example 2123

A DNA sequence (GBSx2239) was identified in *S.agalactiae* <SEQ ID 6553> which encodes the amino acid sequence <SEQ ID 6554>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.26   Transmembrane    8 -  24 (   1 -  28)

----- Final Results -----
               bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

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

```
>GP:AAB84433 GB:AF027868 RAS-related protein [Bacillus subtilis]
 Identities = 53/140 (37%), Positives = 78/140 (54%), Gaps = 2/140 (1%)

Query: 28  VKKVLQYHDLVQNTLAENGSEANVHLVLSMIYTETKGDAIDVMQSSESISGTTNSITDSH 87
           ++++  Y  LV+  L   G       L+L M+Y E+KG    D MQSSES+     N ITD
Sbjct: 49  LERLTDYKPLVEEELESQGLSNYTSLILGMMYQESKGKGNDPMQSSESLGLKRNEITDPQ 108

Query: 88  TSIKHGVTLLSQNISQAKKAKVDVWTAVQAYNFGSSYIDYVADHGGENSIELAKNYSKNV 147
            S+K G+    +       K+  VD+ T +Q+YN G+ YID+VA+HGG ++ ELAK YS+
Sbjct: 109 LSVKQGIKQFTLMYKTGKEKGVDLDTIIQSYNMGAGYIDFVAEHGGTHTEELAKQYSEQQ 168

Query: 148 VA--PSLGNYNGDTYFYYHP 165
           V   P L    G+   + +P
Sbjct: 169 VKKNPDLYTCGGNAKNFRYP 188
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4143> which encodes the amino acid sequence <SEQ ID 4144>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.66   Transmembrane    8 -  24 (   7 -  25)

----- 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>
```

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

```
 Identities = 134/200 (67%), Positives = 165/200 (82%), Gaps = 1/200 (0%)

Query: 1   MFKFLKRLIALIIIIFIGYRLVIIHENVKKVLQYHDLVQNTLAENGSEANVHLVLSMIYT 60
           MF+ LKR  + +++ F+ Y+  +IH NV++VL Y  +V+ TLAEN ++ANV LVL+MIYT
Sbjct: 1   MFRLLKRACSFLLL-FVIYQSFVIHHNVQRVLAYKPMVEKTLAENDTKANVDLVLAMIYT 59

Query: 61  ETKGDAIDVMQSSESISGTTNSITDSHTSIKHGVTLLSQNISQAKKAKVDVWTAVQAYNF 120
           ETKG   DVMQSSES SG  NSITDS  SI+HGV LLS N++ A++A VD WTAVQAYNF
```

```
Sbjct: 60  ETKGGEADVMQSSESSSGQKNSITDSQASIEHGVNLLSHNLALAEEAGVDSWTAVQAYNF 119

Query: 121 GSSYIDYVADHGGENSIELAKNYSKNVVAPSLGNYNGDTYFYYHPLALISGGKLYKNGGN 180
           G++YIDY+A+HGG+N+++LA  YSK VVAPSLGN +G TYFYYHPLALISGGKLYKNGGN
Sbjct: 120 GTAYIDYIAEHGGQNTVDLATTYSKTVVAPSLGNTSGQTYFYYHPLALISGGKLYKNGGN 179

Query: 181 IYYSREVQFNLYLIKIMELF 200
           IYYSREV FNLYLI++M LF
Sbjct: 180 IYYSREVHFNLYLIELMSLF 199
```

SEQ ID 6554 (GBS244) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 4; MW 23.1kDa). 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 2; MW 48kDa).

GBS244-GST was purified as shown in Figure 211, lane 5.

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

Example 2124

A DNA sequence (GBSx2240) was identified in *S.agalactiae* <SEQ ID 6555> which encodes the amino acid sequence <SEQ ID 6556>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2401(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 9837> which encodes amino acid sequence <SEQ ID 9838> was also identified.

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

```
>GP:CAB71302 GB:AJ130879 hypothetical protein [Clostridium
          sticklandii]
 Identities = 32/95 (33%), Positives = 53/95 (55%), Gaps = 1/95 (1%)

Query: 235 LSPEKLADQLFDDNLTARLTFVDELKDAIPGPVQVSDIDHSRQIKKLENQKLSLSNGIEL 294
           LS EK  +  F++    +  + + L  A     Q+ ++  +   +K E QK+   +GIE+
Sbjct: 2   LSVEKALETAFEETDEIKAIYKEALSKAGIENEQI-EVSETALKRKFEIQKIITESGIEV 60

Query: 295 IVPNNVYQDAESVEFIQNPDGTYSILIKNIQDIQN 329
            +P N Y D   +EF+ N DGT S++IKNI +IQ+
Sbjct: 61  KIPVNYYGDPSKLEFVANGDGTVSLVIKNIGNIQS 95
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6557> which encodes the amino acid sequence <SEQ ID 6558>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3336(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 246/325 (75%), Positives = 286/325 (87%)

Query: 6   MMDFYIKQIIIHQFSPNDTELVLSDTPLTLTPRIDDYFRKKLSKVFSDEAKRGYFGEDNV 65
            M+D YIK+I+IHQFSPNDTEL+LSD  +++TPRID+YFRKKL+KVFSDEAKRG F  +N
Sbjct: 1   MLDSYIKRIVIHQFSPNDTELLLSDRLVSITPRIDEYFRKKLAKVFSDEAKRGQFEANNT 60

Query: 66  FMSHLQDDLYVSSCQIAQLWKEEFVISEDQKTNDLVFIQFDKDGMEHFAFLRISLKEQFA 125
            F + + DDL  +S  IAQLWKE FVISEDQKTNDLVF+QFDKDG   FAFLRI+LKEQFA
Sbjct: 61  FPTTIGDDLLETSVTIAQLWKEAFVISEDQKTNDLVFVQFDKDGEPFFAFLRIALKEQFA 120

Query: 126 HVSENQEQPITITQNNLPSAAQTPDEALVVNKSSKQYYLIEKRIKHNGSFANYFSENLLQ 185
            H+S+N E P T+TQNNLPS  QTPDEALV+N  S QYYLIEKR+KHNGSFANYFSE+LL+
Sbjct: 121 HLSDNYEHPFTVTQNNLPSPTQTPDEALVINLKSGQYYLIEKRVKHNGSFANYFSEHLLK 180

Query: 186 VQPEQSVKKSIKMVEQTAQKIAENFNKDDFSFQSKMKSAIYKNLEEEQELSPEKLADQLF 245
            V PEQSVKKSIKM+EQTAQKIAE+FN+DDF+FQSKMKS ++K LE +  LSPEKLADQLF
Sbjct: 181 VTPEQSVKKSIKMIEQTAQKIAEHFNQDDFTFQSKMKSTLFKQLEADDVLSPEKLADQLF 240

Query: 246 DDNLTARLTFVDELKDAIPGPVQVSDIDHSRQIKKLENQKLSLSNGIELIVPNNVYQDAE 305
            DDNLTARLTFVD++KD IP P+++SDI+HSRQIKKLENQKLSLSNGIEL VPN +YQDAE
Sbjct: 241 DDNLTARLTFVDQVKDVIPEPIKISDIEHSRQIKKLENQKLSLSNGIELTVPNAIYQDAE 300

Query: 306 SVEFIQNPDGTYSILIKNIQDIQNK 330
            +VEF+ N DGTYSILIKNI+DI+ K
Sbjct: 301 AVEFLLNDDGTYSILIKNIEDIKTK 325
```

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

Example 2125

A DNA sequence (GBSx2241) was identified in *S.agalactiae* <SEQ ID 6559> which encodes the amino acid sequence <SEQ ID 6560>. This protein is predicted to be Serine hydroxymethyltransferase (glyA-1). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3876(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:AAD35802 GB:AE001743 serine hydroxymethyltransferase [Thermotoga maritima]
 Identities = 243/416 (58%), Positives = 307/416 (73%), Gaps = 7/416 (1%)

Query: 9   KEFDQELWQAIHDEEIRQQNNIELIASENVVSKAVMAAQGSVLTNKYAEGYPSHRYYGGT 68
            K+ D E+++ + +E  RQ+  +ELIASEN  S AV+   GS+LTNKYAEGYP  RYYGG
Sbjct: 6   KQVDPEIYEVLVNELKRQEYGLELIASENFASLAVIETMGSMLTNKYAEGYPKKRYYGGC 65

Query: 69  DCVDVVESLAIERAKTLFNAEFANVQPHSGSQANAAAYMALIEPGDTVLGMDLAAGGHLT 128
            + VD  E  AIERAK LF A+FANVQPHSGSQAN A Y+AL +PGDT++GM L+ GGHLT
Sbjct: 66  EWVDRAEERAIERAKRLFGAKFANVQPHSGSQANMAVYLALAQPGDTIMGMSLSHGGHLT 125

Query: 129 HGASVSFSGKTYHFVSYSVDPKTEMLDYDNILKIAQETQPKLIVAGASAYSRIIDFEKFR 188
            HGA V+FSGK +  V Y V+ +TE +DYD + ++A E +PK+IVAG SAY+RIIDF++FR
Sbjct: 126 HGAPVNFSGKIFKVVPYGVNLETETIDYDEVRRLALEHKPKIIVAGGSAYARIIDFKRFR 185

Query: 189 QIADAVDAYLMVDMAHIAGLVASGHHPSPIPYAHVTTTTTHKTLRGPRGGLILTNDEAIA 248
            +IAD V AYLMVDMAH AGLVA+G HP+P+ YAHV T+TTHKTLRGPRGGLILTND  IA
Sbjct: 186 EIADEVGAYLMVDMAHFAGLVAAGIHPNPLEYAHVVTSTTHKTLRGPRGGLILTNDPEIA 245

Query: 249 KKINSAVFPGLQGGPLEHVIAAKAVALKEALDPSFKIYGEDIIKNAQAMAKVFKEDDDFH 308
```

```
          K ++  +FPG+QGGPL HVIAAKAV  KEA+   FK Y + ++KNA+ MA+ F++   +
Sbjct: 246 KAVDKTIFPGIQGGPLMHVIAAKAVCFKEAMTEEFKEYQKQVVKNAKKMAEEFQK-RGYR 304

Query: 309 LISDGTDNHLFLVDVTKVIENGKKAQNVLEEVNITLNKNSIPFERLSPFKTSGIRIGTPA 368
           ++S GTD HLFLVD+T     GK A+  LE   IT+NKN+IP E+ SPF  SGIRIGTPA
Sbjct: 305 IVSGGTDTHLFLVDLTPKDITGKAAEKALESCGITVNKNTIPNEKRSPFVASGIRIGTPA 364

Query: 369 ITSRGMGVEESRRIAELMIKALKN--HENQDVLTEVRQE----IKSLTDAFPLYEN 418
           +T+RGM  EE   IAE++   L N   EN  V  EVR+E     ++ L + FPLY +
Sbjct: 365 VTTRGMKEEEMEEIAEMIDLVLSNVIDENGTVKPEVREEVSKKVRELCERFPLYRD 420
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6561> which encodes the amino acid sequence <SEQ ID 6562>. Analysis of this protein sequence reveals the following:

```
      Possible site: 47

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.00   Transmembrane  196 - 212 ( 196 - 212)

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

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

```
>GP:CAB15707 GB:Z99122 serine hydroxymethyltransferase [Bacillus subtilis]
 Identities = 250/407 (61%), Positives = 311/407 (75%), Gaps = 2/407 (0%)

Query: 14  DKELWDAIHAEEERQEHHIELIASENMVSKAVMAAQGSVLTNKYAEGYPGNRYYGGTECV 73
           D+++++AI  E ERQ+  IELIASEN VS+AVM AQGSVLTNKYAEGYPG RYYGG E V
Sbjct: 8   DEQVFNAIKNERERQQTKIELIASENFVSEAVMEAQGSVLTNKYAEGYPGKRYYGGCEHV 67

Query: 74  DIVETLAIERAKKLFGAAFANVQAHSGSQANAAAYMALIEAGDTVLGMDLAAGGHLTHGS 133
           D+VE +A +RAK++FGA   NVQ HSG+QAN A Y  ++E GDTVLGM+L+ GGHLTHGS
Sbjct: 68  DVVEDIARDRAKEIFGAEHVNVQPHSGAQANMAVYFTILEQGDTVLGMNLSHGGHLTHGS 127

Query: 134 PVNFSGKTYHFVGYSVDTDTEMLNYEAILEQAKAVQPKLIVAGASAYSRSIDFEKFRAIA 193
           PVNFSG  Y+FV Y VD +T+ ++Y+ + E+A A +PKLIVAGASAY R+IDF+KFR IA
Sbjct: 128 PVNFSGVQYNFVEYGVDKETQYIDYDDVREKALAHKPKLIVAGASAYPRTIDFKKFREIA 187

Query: 194 DHVGAYLMVDMAHIAGLVAAGVHPSPVPYAHIVTSTTHKTLRGPRGGLILTNDEALAKKI 253
           D VGAY MVDMAHIAGLVAAG+HP+PVPYA  VT+TTHKTLRGPRGG+IL  +E   KKI
Sbjct: 188 DEVGAYFMVDMAHIAGLVAAGLHPNPVPYADFVTTTTHKTLRGPRGGMILCREE-FGKKI 246

Query: 254 NSAVFPGLQGGPLEHVIAAKAVAFKEALDPAFKDYAQAIIDNTAAMAAVFAQDDRFRLIS 313
           + ++FPG+QGGPL HVIAAKAV+F E L   FK YAQ +I N   +A    ++   +L+S
Sbjct: 247 DKSIFPGIQGGPLMHVIAAKAVSFGEVLQDDFKTYAQNVISNAKRLAEALTKEG-IQLVS 305

Query: 314 GGTDNHVFLVDVTKVIANGKLAQNLLDEVNITLNKNAIPFETLSPFKTSGIRIGCAAITS 373
           GGTDNH+ LVD+  + GK+A+++LDE+ IT NKNAIP++  PF TSGIR+G AA+TS
Sbjct: 306 GGTDNHLILVDLRSLGLTGKVAEHVLDEIGITSNKNAIPYDPEKPFVTSGIRLGTAAVTS 365

Query: 374 RGMGVKESQTIARLIIKALVNHDQETILEEVRQEVRQLTDAFPLYKK 420
           RG      + +  +I  AL NH+ E  LEE RQ V  LTD FPLYK+
Sbjct: 366 RGFDGDALEEVGAIIALALKNHEDEGKLEEARQRVAALTDKFPLYKE 412
```

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

```
 Identities = 330/417 (79%), Positives = 368/417 (88%)

Query: 1   MIFDKDNFKEFDQELWQAIHDEEIRQQNNIELIASENVVSKAVMAAQGSVLTNKYAEGYP 60
           MIFDK N ++FD+ELW AIH EE RQ+++IELIASEN+VSKAVMAAQGSVLTNKYAEGYP
Sbjct: 3   MIFDKGNVEDFDKELWDAIHAEEERQEHHIELIASENMVSKAVMAAQGSVLTNKYAEGYP 62

Query: 61  SHRYYGGTDCVDVVESLAIERAKTLFNAEFANVQPHSGSQANAAAYMALIEPGDTVLGMD 120
           +RYYGGT+CVD+VE+LAIERAK LF A FANVQ HSGSQANAAAYMALIE GDTVLGMD
```

```
Sbjct: 63  GNRYYGGTECVDIVETLAIERAKKLFGAAFANVQAHSGSQANAAAYMALIEAGDTVLGMD 122

Query: 121 LAAGGHLTHGASVSFSGKTYHFVSYSVDPKTEMLDYDNILKIAQETQPKLIVAGASAYSR 180
           LAAGGHLTHG+ V+FSGKTYHFV YSVD  TEML+Y+ IL+ A+  QPKLIVAGASAYSR
Sbjct: 123 LAAGGHLTHGSPVNFSGKTYHFVGYSVDTDTEMLNYEAILEQAKAVQPKLIVAGASAYSR 182

Query: 181 IIDFEKFRQIADAVDAYLMVDMAHIAGLVASGHHPSPIPYAHVTTTTTHKTLRGPRGGLI 240
            IDFEKFR IAD V AYLMVDMAHIAGLVA+G HPSP+PYAH+ T+TTHKTLRGPRGGLI
Sbjct: 183 SIDFEKFRAIADHVGAYLMVDMAHIAGLVAAGVHPSPVPYAHIVTSTTHKTLRGPRGGLI 242

Query: 241 LTNDEAIAKKINSAVFPGLQGGPLEHVIAAKAVALKEALDPSFKIYGEDIIKNAQAMAKV 300
           LTNDEA+AKKINSAVFPGLQGGPLEHVIAAKAVA KEALDP+FK Y + II N  AMA V
Sbjct: 243 LTNDEALAKKINSAVFPGLQGGPLEHVIAAKAVAFKEALDPAFKDYAQAIIDNTAAMAAV 302

Query: 301 FKEDDDFHLISDGTDNHLFLVDVTKVIENGKKAQNVLEEVNITLNKNSIPFERLSPFKTS 360
           F +DD F LIS GTDNH+FLVDVTKVI NGK AQN+L+EVNITLNKN+IPFE LSPFKTS
Sbjct: 303 FAQDDRFRLISGGTDNHVFLVDVTKVIANGKLAQNLLDEVNITLNKNAIPFETLSPFKTS 362

Query: 361 GIRIGTPAITSRGMGVEESRRIAELMIKALKNHENQDVLTEVRQEIKSLTDAFPLYE 417
           GIRIG  AITSRGMGV+ES+ IA L+IKAL NH+ + +L EVRQE++ LTDAFPLY+
Sbjct: 363 GIRIGCAAITSRGMGVKESQTIARLIIKALVNHDQETILEEVRQEVRQLTDAFPLYK 419
```

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

Example 2126

A DNA sequence (GBSx2242) was identified in *S.agalactiae* <SEQ ID 6563> which encodes the amino acid sequence <SEQ ID 6564>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2289(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 9839> which encodes amino acid sequence <SEQ ID 9840> was also identified.

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

```
>GP:AAD35934 GB:AE001752 conserved hypothetical protein [Thermotoga maritima]
 Identities = 71/198 (35%), Positives = 114/198 (56%), Gaps = 4/198 (2%)

Query: 1   MNDLGQILEDHGAVIMPTETVYGIFAKALSEEAVNHVYELKKRPRDKAMNLNICDFETIL 60
           + +  ++L +   +I PTETVYGI A A +EEA   +++LK+RP D  + ++I  F+ +
Sbjct: 17  LKEAAAELLRNGEVIIFPTETVYGIGADAYNEEACKKIFKLKERPADNPLIVHIHSFKQLE 76

Query: 61  KYSKNQPTYLKQLYDAFLPGPLTIIL-EASQEVPHWINSGLLSVGFRMPKHPVTLDMIAN 119
           + ++    +L  L   F PGPLT+I  + S+++P  + + L +V  RMP HPV L +I
Sbjct: 77  EIAEGYEPHLDFL-KKFWPGPLTVIFRKKSEKIPPVVTADLPTVAVRMPAHPVALKLIEL 135

Query: 120 HG-PLIGPSANISGCDSGRVFSEIQKQFNHQV-LGIEDDKALTGVDSTIIDLSGDRVKIL 177
            G P+  PSANISG  S      + + F  +V L I+      G++STI+DL+ ++  +L
Sbjct: 136 FGHPIAAPSANISGRPSATNVKHVIEDFMGKVKLIIDAGDTPFGLESTIVDLTKEKPVLL 195

Query: 178 RQGAITQEVLTATIPELI 195
           R G +  E L    PEL+
Sbjct: 196 RPGPVEVERLKELFPELV 213
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6565> which encodes the amino acid sequence <SEQ ID 6566>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0282(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 = 127/196 (64%), Positives = 154/196 (77%)

Query: 1   MNDLGQILEDHGAVIMPTETVYGIFAKALSEEAVNHVYELKKRPRDKAMNLNICDFETIL 60
           M  L  I+E   A+++PTETVYG+FAKAL E+AVN VY+LK+RPRDKAMNLN+ DF +IL
Sbjct: 11  MEYLASIIESGDALVLPTETVYGLFAKALDEKAVNAVYDLKQRPRDKAMNLNVADFNSIL 70

Query: 61  KYSKNQPTYLKQLYDAFLPGPLTIILEASQEVPHWINSGLLSVGFRMPKHPVTLDMIANH 120
            +SK QP YLK+LY AFLPGPLTIIL+A+ +VP+WINSGL +VGFR+P HP+T  +I
Sbjct: 71  AFSKEQPRYLKKLYQAFLPGPLTIILKANDQVPYWINSGLSTVGFRLPSHPITAALIQKT 130

Query: 121 GPLIGPSANISGCDSGRVFSEIQKQFNHQVLGIEDDKALTGVDSTIIDLSGDRVKILRQG 180
           GPLIGPSAN+SG  SGRVF  I + F+ QV G  DD  LTG DSTI+DLSG+R  ILRQG
Sbjct: 131 GPLIGPSANLSGKASGRVFDHIMQDFDFQVFGYADDPFLTGKDSTILDLSGERAVILRQG 190

Query: 181 AITQEVLTATIPELIF 196
           AIT+E L A +PEL F
Sbjct: 191 AITKEELLANVPELRF 206
```

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

Example 2127

A DNA sequence (GBSx2243) was identified in *S.agalactiae* <SEQ ID 6567> which encodes the amino acid sequence <SEQ ID 6568>. This protein is predicted to be protoporphyrinogen oxidase (hemK). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3000(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:BAB07493 GB:AP001519 protoporphyrinogen oxidase [Bacillus halodurans]
 Identities = 94/236 (39%), Positives = 132/236 (55%), Gaps = 12/236 (5%)

Query: 49  DTDQQLMENIFQQLKKHRSP---QYITGKAYFRDLIFFVDERVLIPRPETEELVDLILSE 105
           + D +L + + + L  H S    Q++ G   F    F VD+ VLIPRPETEELV  +L E
Sbjct: 46  ELDGELFQRLEEDLAAHASGVPVQHLIGVESFYGRQFQVDQHVLIPRPETEELVLAVLKE 105

Query: 106 -----NKVEDCSVLDIGTGSGAIAISLKKERPSWDVLASDISVSALDLAKENANNCDAEV 160
                K E+ ++LDIGTGSGAIA++L  E    +V A DIS  AL +A +NA    A V
Sbjct: 106 IRRQFKKEEEITILDIGTGSGAIAVTLALEEERTNVTAVDISRDALQVAADNARRLGANV 165

Query: 161 TFIESDV---FSNISGKFDIIVSNPPYISYNDKDEVGKNVLASEPHSALFADEEGLAIYR 217
            I  D+    F      +FD+IVSNPPYI    +KD +  +V   EP  ALF    +GL +YR
Sbjct: 166 QLIHGDLGEPFLKTGERFDVIVSNPPYIPTVEKDTLAVHVRDHEPALALFGGVDGLDVYR 225

Query: 218 KIIENSREYL-QPRGKLYFEIGYKQGDDLRSLLKRYFPNNRCRVLKDIFGKDRMVV 272
           +++         + +G +  EIG  QG D+  L++  +P      VL D+ GKDR+V+
Sbjct: 226 RLMSQLPALTKEEKGMVALEIGAGQGMDVEKLMQTAYPKAAVDVLYDLNGKDRIVL 281
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6569> which encodes the amino acid sequence <SEQ ID 6570>. Analysis of this protein sequence reveals the following:

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

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

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

```
 Identities = 174/274 (63%), Positives = 207/274 (75%)

Query: 1   MNYAQLIKHYGQLLEACGEEVENFIYVLKDLKQWSTTDYLLNQNSSVSDTDQQLMENIFQ 60
            MNYA LI+ Y   LE   E+ EN  YV +++K+WS+ D L++QN +V+  D  L+E+IF
Sbjct: 1   MNYATLIRTYEDKLEQIDEDRENLAYVFREIKEWSSLDMLIHQNQAVTPEDAVLLEHIFC 60

Query: 61  QLKKHRSPQYITGKAYFRDLIFFVDERVLIPRPETEELVDLILSENKVEDCSVLDIGTGS 120
            L +H SPQYITG AYFRDL   VD+RVLIPRPETEELVD+IL+EN     +VLDIGTGS
Sbjct: 61  SLSQHLSPQYITGNAYFRDLKLAVDKRVLIPRPETEELVDMILAENLDAPLNVLDIGTGS 120

Query: 121 GAIAISLKKERPSWDVLASDISVSALDLAKENANNCDAEVTFIESDVFSNISGKFDIIVS 180
            GAIAISLKKERP+W V ASDIS +ALDLAK NA+    ++TFIESDVFS IS  FDIIVS
Sbjct: 121 GAIAISLKKERPNWQVTASDISRAALDLAKANADAYQLDITFIESDVFSLISETFDIIVS 180

Query: 181 NPPYISYNDKDEVGKNVLASEPHSALFADEEGLAIYRKIIENSREYLQPRGKLYFEIGYK 240
            NPPYISY DK+EV  NVL SEPH ALFA E G AIYRKIIE +  YL   GKLYFEIGYK
Sbjct: 181 NPPYISYEDKEEVSLNVLQSEPHLALFAKENGYAIYRKIIEQADNYLTKEGKLYFEIGYK 240

Query: 241 QGDDLRSLLKRYFPNNRCRVLKDIFGKDRMVVLD 274
            Q + ++ +L+ YFP    R + DIFGK+RMVV+D
Sbjct: 241 QAEGIKDMLQAYFPQRHIRAVTDIFGKERMVVVD 274
```

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

Example 2128

A DNA sequence (GBSx2244) was identified in *S.agalactiae* <SEQ ID 6571> which encodes the amino acid sequence <SEQ ID 6572>. This protein is predicted to be peptide chain release factor RF-1 (prfA). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3446(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:CAB15718 GB:Z99122 peptide chain release factor 1 [Bacillus subtilis]
 Identities = 211/351 (60%), Positives = 280/351 (79%), Gaps = 1/351 (0%)

Query: 5   DQLQAVEDRYEELGELLSDPDVVSDTKRFMELSREEASTRETVTAYREYKQVIQNISDAE 64
            D+L+++E+RYE+L ELLSDP+VV+D K+  E S+E++  +ETV  YR+Y+   + ++DA+
Sbjct: 3   DRLKSIEERYEKLNELLSDPEVVNDPKKLREYSKEQSDIQETVDVYRQYRDASEQLADAK 62

Query: 65  EMIKDASGDAELEEMAKEELKESKAAKEEYEERLKILLLPKDPNDDKNIILEIRGAAGGD 124
             M+++   DAE+ +M KEE+ E +   E   ERLK+LL+PKDPNDDKN+I+EIRGAAGG+
```

```
Sbjct: 63  AMLEEKL-DAEMRDMVKEEISELQKETETLSERLKVLLIPKDPNDDKNVIMEIRGAAGGE 121

Query: 125 EAALFAGDLLTMYQKYAETQGWRFEVMESSVNGVGGIKEVVAMVSGQSVYSKLKYESGAH 184
           EAALFAG+L  MY +YAE QGW+ EVME++V G GG KE++ M++G   YSKLKYE+GAH
Sbjct: 122 EAALFAGNLYRMYSRYAELQGWKTEVMEANVTGTGGYKEIIFMITGSGAYSKLKYENGAH 181

Query: 185 RVQRVPVTESQGRVHTSTATVLVMPEVEEVEYEIDQKDLRVDIYHASGAGGQNVNKVATA 244
           RVQRVP TES GR+HTSTATV  +PE EEVE +I +KD+RVD + +SG GGQ+VN   +A
Sbjct: 182 RVQRVPETESGGRIHTSTATVACLPEAEEVEVDIHEKDIRVDTFASSGPGGQSVNTTMSA 241

Query: 245 VRMVHIPTGIKVEMQEERTQQKNRDKAMKIIRARVADHFAQIAQDEQDAERKSTVGTGDR 304
           VR+ H+PTG+ V  Q+E++Q KN++KAMK++RAR+ D F Q AQ E D  RKS VG+GDR
Sbjct: 242 VRLTHLPTGVVVSCQDEKSQIKNKEKAMKVLRARIYDKFQQEAQAEYDQTRKSAVGSGDR 301

Query: 305 SERIRTYNFPQNRVTDHRIGLTLQKLDTILSGKMDEVIDALVMYDQTQKLE 355
           SERIRTYNFPQNRVTDHRIGLT+QKLD IL GK+DEV++AL++ DQ  KL+
Sbjct: 302 SERIRTYNFPQNRVTDHRIGLTIQKLDQILEGKLDEVVEALIVEDQASKLQ 352
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6573> which encodes the amino acid sequence <SEQ ID 6574>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3446(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 = 349/358 (97%), Positives = 354/358 (98%)

Query: 1   MNIYDQLQAVEDRYEELGELLSDPDVVSDTKRFMELSREEASTRETVTAYREYKQVIQNI 60
           MNIYDQLQAVEDRYEELGELLSDPDVVSDTKRFMELSREE +TRETVTAYREYKQVIQ I
Sbjct: 1   MNIYDQLQAVEDRYEELGELLSDPDVVSDTKRFMELSREETNTRETVTAYREYKQVIQTI 60

Query: 61  SDAEEMIKDASGDAELEEMAKEELKESKAAKEEYEERLKILLLPKDPNDDKNIILEIRGA 120
           SDAEEMIKDASGD ELEEMAKEELKESKAAKEEYEE+LKILLLPKDPNDDKNIILEIRGA
Sbjct: 61  SDAEEMIKDASGDPELEEMAKEELKESKAAKEEYEEKLKILLLPKDPNDDKNIILEIRGA 120

Query: 121 AGGDEAALFAGDLLTMYQKYAETQGWRFEVMESSVNGVGGIKEVVAMVSGQSVYSKLKYE 180
           AGGDEAALFAGDLLTMYQKYAETQGWRFEVMESSVNGVGGIKEVVAMVSGQSVYSKLKYE
Sbjct: 121 AGGDEAALFAGDLLTMYQKYAETQGWRFEVMESSVNGVGGIKEVVAMVSGQSVYSKLKYE 180

Query: 181 SGAHRVQRVPVTESQGRVHTSTATVLVMPEVEEVEYEIDQKDLRVDIYHASGAGGQNVNK 240
           SGAHRVQRVPVTESQGRVHTSTATVLVMPEVEEVEY+ID KDLRVDIYHASGAGGQNVNK
Sbjct: 181 SGAHRVQRVPVTESQGRVHTSTATVLVMPEVEEVEYDIDPKDLRVDIYHASGAGGQNVNK 240

Query: 241 VATAVRMVHIPTGIKVEMQEERTQQKNRDKAMKIIRARVADHFAQIAQDEQDAERKSTVG 300
           VATAVRMVHIPTGIKVEMQEERTQQKNRDKAMKIIRARVADHFAQIAQDEQDAERKSTVG
Sbjct: 241 VATAVRMVHIPTGIKVEMQEERTQQKNRDKAMKIIRARVADHFAQIAQDEQDAERKSTVG 300

Query: 301 TGDRSERIRTYNFPQNRVTDHRIGLTLQKLDTILSGKMDEVIDALVMYDQTQKLEALN 358
           TGDRSERIRTYNFPQNRVTDHRIGLTLQKLDTILSGKMDEVIDALVMYDQT+KLE+LN
Sbjct: 301 TGDRSERIRTYNFPQNRVTDHRIGLTLQKLDTILSGKMDEVIDALVMYDQTKKLESLN 358
```

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

Example 2129

A DNA sequence (GBSx2245) was identified in *S.agalactiae* <SEQ ID 6575> which encodes the amino acid sequence <SEQ ID 6576>. This protein is predicted to be thymidine kinase (tdk). Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2244(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 9841> which encodes amino acid sequence <SEQ ID 9842> was also identified.

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

```
>GP:AAB02289 GB:L40415 thymidine kinase [Streptococcus gordonii]
 Identities = 158/189 (83%), Positives = 175/189 (91%)

Query: 1   MAQLYYKYGTMNSGKTIEILKVAHNYEEQGKPVVIMTSALDTRDEFGVVSSRIGMRREAV 60
           MAQLYYKYGTMNSGKTIEILKVAHNYEEQGK VVIMTSA+DTRD  G VSSRIGM+R+A+
Sbjct: 1   MAQLYYKYGTMNSGKTIEILKVAHNYEEQGKGVVIMTSAVDTRDGVGYVSSRIGMKRQAM 60

Query: 61  PISDDMDIFSYIQNLPQKPYCVLIDECQFLSKKNVYDLARVVDDLDVPVMAFGLKNDFQN 120
            I DD DI  YI+NLP+KPYC+LIDE QFL + +VYDLARVVD+LDVPVMAFGLKNDF+N
Sbjct: 61  AIEDDTDIILGYIKNLPEKPYCILIDEAQFLKRHHVYDLARVVDELDVPVMAFGLKNDFRN 120

Query: 121 NLFEGSKHLLLLADKIDEIKTICQYCSKKATMVLRTENGKPVYEGDQIQIGGNETYIPVC 180
            LFEGSKHLLLLADKI+EIKTICQYCS+KATMVLRT++GKPVY+G+QIQIGGNETYIPVC
Sbjct: 121 ELFEGSKHLLLLADKIEEIKTICQYCSRKATMVLRTDHGKPVYDGEQIQIGGNETYIPVC 180

Query: 181 RKHYFNPDI 189
           RKHYF PDI
Sbjct: 181 RKHYFKPDI 189
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6577> which encodes the amino acid sequence <SEQ ID 6578>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2244(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

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

```
 Identities = 174/189 (92%), Positives = 184/189 (97%)

Query: 1   MAQLYYKYGTMNSGKTIEILKVAHNYEEQGKPVVIMTSALDTRDEFGVVSSRIGMRREAV 60
           +AQLYYKYGTMNSGKTIEILKVAHNYEEQGKPVVIMTSALDTRD FG+VSSRIGMRREA+
Sbjct: 1   LAQLYYKYGTMNSGKTIEILKVAHNYEEQGKPVVIMTSALDTRDGFGIVSSRIGMRREAI 60

Query: 61  PISDDMDIFSYIQNLPQKPYCVLIDECQFLSKKNVYDLARVVDDLDVPVMAFGLKNDFQN 120
           PIS+DMDIF++I  L +KPYCVLIDE QFLSK+NVYDLARVVD+L+VPVMAFGLKNDFQN
Sbjct: 61  PISNDMDIFTFIAQLEEKPYCVLIDESQFLSKQNVYDLARVVDELNVPVMAFGLKNDFQN 120

Query: 121 NLFEGSKHLLLLADKIDEIKTICQYCSKKATMVLRTENGKPVYEGDQIQIGGNETYIPVC 180
           NLFEGSKHLLLLADKIDEIKTICQYCSKKATMVLRTENGKPVYEGDQIQIGGNETYIPVC
Sbjct: 121 NLFEGSKHLLLLADKIDEIKTICQYCSKKATMVLRTENGKPVYEGDQIQIGGNETYIPVC 180
```

```
Query: 181 RKHYFNPDI 189
           RKHYFNPDI
Sbjct: 181 RKHYFNPDI 189
```

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

Example 2130

A DNA sequence (GBSx2246) was identified in *S.agalactiae* <SEQ ID 6579> which encodes the amino acid sequence <SEQ ID 6580>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3995(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:AAA26046 GB:M95650 4-oxalocrotonate tautomerase [Plasmid pWW0]
 Identities = 27/60 (45%), Positives = 36/60 (60%)

Query: 1  MPFVKIDLFEGRSQEQKNELAREVTEVVSRIAKAPKENIHVFINDMPEGTYYPQGELKKK 60
          MP  +I + EGRS EQK  L REV+E +SR   AP  ++ V I +M +G +   GEL  K
Sbjct: 1  MPIAQIHILEGRSDEQKETLIREVSEAISRSLDAPLTSVRVIITEMAKGHFGIGGELASK 60
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6581> which encodes the amino acid sequence <SEQ ID 6582>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4128(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 = 56/60 (93%), Positives = 59/60 (98%)

Query: 1  MPFVKIDLFEGRSQEQKNELAREVTEVVSRIAKAPKENIHVFINDMPEGTYYPQGELKKK 60
          MPFV IDLFEGRSQEQKN+LAREVTEVVSRIAKAPKENIHVFINDMPEGTYYPQGE+K+K
Sbjct: 1  MPFVTIDLFEGRSQEQKNQLAREVTEVVSRIAKAPKENIHVFINDMPEGTYYPQGEMKQK 60
```

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

Example 2131

A DNA sequence (GBSx2247) was identified in *S.agalactiae* <SEQ ID 6583> which encodes the amino acid sequence <SEQ ID 6584>. Analysis of this protein sequence reveals the following:

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

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2154(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```