



(21) (A1) **2,309,005**
(86) 1998/10/30
(87) 1999/05/14

(72) EL-SHERBEINI, MOHAMMED, US

(72) WONG, KENNY KIN CHUNG, US

(71) MERCK & CO., INC., US

(51) Int.Cl.⁶ C12N 1/00, C12Q 1/68, C12N 15/63, C12N 9/00, C12N 15/52,
C12N 5/10

(30) 1997/11/04 (60/064,205) US

(54) **PROTEINE MURD ET GENE DE STREPTOCOCCUS
PYOGENES**

(54) **MURD PROTEIN AND GENE OF STREPTOCOCCUS
PYOGENES**

(57) L'invention concerne le gène murD de Streptococcus pyogenes, ainsi que des protéines murD recombinantes isolées et purifiées. On décrit également des séquences d'acide nucléique codant des protéines murD fonctionnellement actives. On décrit enfin des dosages permettant d'identifier des modulateurs de l'expression de murD et des inhibiteurs de l'activité de murD.

(57) This invention provides the murD gene of Streptococcus pyogenes. Purified and isolated MurD recombinant proteins are also provided. Nucleic acid sequences which encode functionally active MurD proteins are described. Assays for the identification of modulators of the expression of murD and inhibitors of the activity of MurD, are also provided.



**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 1/00, 5/10, 9/00, 15/52, 15/63, C12Q 1/68	A1	(11) International Publication Number: WO 99/23201 (43) International Publication Date: 14 May 1999 (14.05.99)
(21) International Application Number: PCT/US98/23054 (22) International Filing Date: 30 October 1998 (30.10.98) (30) Priority Data: 60/064,205 4 November 1997 (04.11.97) US (71) Applicant (for all designated States except US): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): EL-SHERBEINI, Mohammed [EG/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). WONG, Kenny, Kin, Chung [CN/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). (74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).		(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>
(54) Title: MURD PROTEIN AND GENE OF <i>STREPTOCOCCUS PYOGENES</i> (57) Abstract This invention provides the <i>murD</i> gene of <i>Streptococcus pyogenes</i>. Purified and isolated MurD recombinant proteins are also provided. Nucleic acid sequences which encode functionally active MurD proteins are described. Assays for the identification of modulators of the expression of <i>murD</i> and inhibitors of the activity of MurD, are also provided.		

TITLE OF THE INVENTION

MURD PROTEIN AND GENE OF *STREPTOCOCCUS PYOGENES*

CROSS-REFERENCE TO RELATED APPLICATIONS

- 5 This application claims the benefit of U.S. Provisional Application No. 60/064,205, filed November 4, 1997, the contents of which are incorporated herein by reference in their entirety.

STATEMENT REGARDING FEDERALLY-SPONSORED R&D

- 10 Not applicable.

REFERENCE TO MICROFICHE APPENDIX

Not applicable.

15 FIELD OF THE INVENTION

 This invention relates to the genes and enzymes involved in cell wall synthesis in bacteria, and particularly to the inhibition of such enzymes.

20 BACKGROUND OF THE INVENTION

- The molecular target of many naturally-occurring antibiotics, including fosfomycin, cycloserine and b-lactams, is the synthesis of the bacterial cell wall. The frequency with which these types of antibiotics arose in evolution indicates that the pathway of cell wall biosynthesis is a particularly effective point of attack against bacteria. Genetic studies confirm the soundness of this process as a target, as temperature-sensitive alleles of the intracellular pathway genes are lytic, and therefore lethal. Since the building blocks of the cell wall are highly conserved structures in both Gram-positive and Gram-negative bacteria, but are unique to the eubacteria, novel inhibitors of cell wall formation are expected to be both broad spectrum and safe antibiotics.

- The bacterial cell wall is a polymer -- a single molecule composed of peptidoglycan -- that defines the boundary and shape of the cell. Assembled by crosslinking glycan chains with short peptide bridges (Rogers, H. J., H. R. Perkins, and J. B. Ward, 1980, Biosynthesis of peptidoglycan. p. 239-297. In Microbial cell walls and

membranes. Chapman & Hall Ltd. London), the completed structure is strong enough to maintain cell integrity against an osmotic pressure differential of over four atmospheres, but also flexible enough to allow the cell to move, grow and divide.

5 The construction of the peptidoglycan begins in the cytoplasm with an activated sugar molecule, UDP-N-acetylglucosamine. After two reactions (catalyzed by MurA and MurB) that result in the placement of a lactyl group on the 3-OH of the glucosamine moiety, a series of ATP-dependent amino acid ligases (MurC, -D, -E, and -F)
10 catalyze the stepwise synthesis of the pentapeptide sidechain using the newly synthesized lactyl carboxylate as the first acceptor site. After attachment of the sugar pentapeptide to a lipid carrier in the plasma membrane, another glucosamine unit is added to the 4-OH of the muramic acid moiety. The completed monomeric building block is
15 moved across the membrane into the periplasm where the penicillin-binding proteins enzymatically add it into the growing cell wall.

 Because the pentapeptide sidechain is not synthesized ribosomally it contains more diverse chemical functionality than a typical peptide, both structurally and stereochemically. Two of the
20 enzymes catalyze the addition of D-amino acids (MurD and MurF) and MurE mediates the formation of a peptide bond between the γ -carboxylate of D-glutamate and the amino group of L-lysine. Presumably these structures render the exposed peptidoglycan resistant to the action of proteases, but they also imply that the active sites of the
25 enzymes must have unusual structures in order to handle the somewhat uncommon substrates. These unusual active sites are targets to bind novel inhibitors that can have antimicrobial activity.

 Among these potential enzyme targets is MurD. The first partial purification and characterization of a D-glutamate-adding
30 enzyme was from *Staphylococcus aureus* (Ito, E. and J. L. Strominger, 1962, Enzymatic synthesis of the peptide in bacterial uridine nucleotides: Enzymatic addition of L-alanine, D-glutamic acid, and L-lysine. J. Biol. Chem. 237: 2689-2695; Nathenson, S. G., J. L. Strominger, and E. Ito, 1964, Enzymatic synthesis of the peptide in bacterial uridine nucleotides:
35 purification and properties of D-Glutamic acid-adding enzyme, J. Biol. Chem. 239: 1773-1777), followed by studies in more detail on the isolated *E. coli* enzyme (Blanot, D., A. Kretsovali, M. Abo-Ghalia, D. Mengin-

- Lecreulx, and J. van Heijenoort, 1983. Synthesis of analogues of precursors of bacterial peptidoglycan. p. 311-314. *In* Peptides. Blaha, K. and P. Malon, eds. pp. 311-314, Walter de Gryter and Co. Berlin, New York.; Jin, H., Emanuele, J. J., Jr., Fairman, R., Robertson, J. G.,
- 5 Hail, M. E., Ho, H.-T., Falk, P. and Villafranca, J. J., 1996. Structural studies of *Escherichia coli* UDP-N-acetylmuramate: L-alanine ligase. *Biochemistry* 35: 14423-14431; Ito E. and J. L. Strominger, 1973. Enzymatic synthesis of the peptide in bacterial uridine nucleotides: Comparative biochemistry. *J. Biol. Chem.* 248: 3131-3136; Michaud, C.
- 10 D. Blanot, B. Flouret, and J. van Heijenoort, 1987. Partial purification and specificity studies of the D-glutamate-adding and D-alanyl-D-alanine-adding enzymes from *Escherichia coli* K12. *Eur. J. Biochem.* 166: 631-637). Recently, a purified recombinant *E. coli* MurD was reported (Pratviel-Sosa F, D. Mengin-Lecreulx and J. van Heijenoort,
- 15 1991. Over-production, purification and properties of the uridine diphosphate N-acetylmuramoyl-L-alanine:D-glutamate ligase from *Escherichia coli*. *Eur. J. Biochem.* 202 (3):1169-1176) and genes encoding MurD have been cloned from several species of bacteria including *E. coli* (Ikeda, M., M. Wachi, F. Ishino, and M. Matsushashi, 1990, Nucleotide
- 20 sequence involving *murD* and an open reading frame ORF-Y spacing *murF* and *ftsW* in *Escherichia coli*. *Nucleic Acids Res.* 18:1058; Mengin-Lecreulx, D., C Parquet, L. Desviat, J. Pla, B. Flouret, J. Ayala and J. van Heijenoort, 1989, Organization of the *murE-murG* region of *Escherichia coli*: Identification of the *murD* gene encoding the D-
- 25 glutamic-acid-adding enzyme. *J. Bacteriol.* 171: 6126-6134) and *B. subtilis* (Daniel, R. A., and J. Errington, 1993, DNA sequence of the *murE-murD* region of *Bacillus subtilis* 168. *J. Gen. Microbiol.* 139:361-370; Henriques, A. O. de Lencaster, H. and P. J. Piggot, 1992, A *Bacillus subtilis* morphogene cluster that includes *spoVE* is homologous to the
- 30 *mra* region of *Escherichia coli*. *Biochimie.* 74: 735-748). Compounds have been designed and synthesized that have inhibitory activity against the *E. coli* enzyme (Tanner, M.E., S. Vaganay, van Heijenoort, J., and D. Blanot, 1996, Phosphinate Inhibitors of the D-Glutamic Acid-Adding Enzyme of Peptidoglycan Biosynthesis. *J. Org. Chem.* 61: 1756-1760),
- 35 although they do not have antibacterial activity.

SUMMARY OF THE INVENTION

Polynucleotides and polypeptides of *Streptococcus pyogenes* MurD, an enzyme involved in bacterial cell wall biosynthesis are provided. The recombinant MurD enzyme is catalytically active in ATP-dependent D-glutamate addition reactions. The enzyme is used in *in vitro* assays to screen for antibacterial compounds that target cell wall biosynthesis. The invention includes the purified polynucleotides, purified enzymes encoded by the polynucleotides, and host cells expressing the recombinant enzyme and their use in assays.

10

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A & 1B. Nucleotide sequence and the predicted amino acid sequence of *S. pyogenes murD*. The amino acid sequence (SEQ ID NO:2) is presented in three-letter code below the nucleotide sequence (nucleotides 18 to 1367 of SEQ ID NO: 1).

FIGS. 2A & 2B. Alignment of the deduced amino acid sequence of *S. pyogenes* MurD with MurD homologs of *E. coli* (Eco) (SEQ ID NO:3), *H. influenzae* (Hin) (SEQ ID NO:4), *B. subtilis* (Bsu) (SEQ ID NO:5) and *S. aureus* (Sau) (SEQ ID NO:6). Single letters amino acids are used. Identical residues are indicated by boxes. Gaps are indicated by dashes.

FIG. 3. The murein consensus sequence of the MurD of *S. pyogenes* (Spy). The murein consensus sequence (shown in bold) was generated by multiple alignments of murein genes murC, -D, -E, and -F from various bacterial genera (Eveland, S. S., D. L. Pompliano, and M. S. Anderson, 1997, Conditionally lethal *Escherichia coli* murein mutants contain point defects that map to regions conserved among murein and folyl poly-g-glutamate ligases: Identification of a ligase superfamily. Biochemistry. 36: 6223-6229).

30

DETAILED DESCRIPTION OF THE INVENTION

This invention provides polynucleotides and polypeptides of a cell wall biosynthesis gene from *Streptococcus pyogenes*, referred to herein as MurD. The polynucleotides and polypeptides are used to further provide expression vectors, host cells comprising the vectors, probes and primers, antibodies against the MurD protein and

35

polypeptides thereof, assays for the presence or expression of MurD and assays for the identification of modulators and inhibitors of MurD.

Bacterial UDP-N-acetylmuramyl-L-alanine:D-glutamate ligase (MurD), a cytoplasmic peptidoglycan biosynthetic enzyme,
5 catalyzes the ATP-dependent addition of D-glutamate to an alanyl residue of the UDP-N-acetylmuramyl-L-alanine precursor, generating the dipeptide.

The *murD* gene was cloned from *Streptococcus pyogenes*. Sequence analysis of the *S. pyogenes murD* gene revealed an open
10 reading frame of 452 amino acids. The deduced amino acid sequence of *S. pyogenes* MurD is highly homologous to MurD from *Escherichia coli*, *Haemophilus influenzae*, *Bacillus subtilis* and *S. aureus* (FIGS. 2A & 2B) Recombinant MurD protein from *S. pyogenes* was over-produced as His-tagged fusion protein in *Escherichia coli* host cells. The *S. pyogenes*
15 MurD enzyme was purified to apparent homogeneity. The recombinant enzyme catalyzed the ATP-dependent addition of D-glutamate to the precursor sugar peptide.

As used herein an "inhibitor" is a compound or molecule that interacts with and inhibits or prevents a polypeptide of MurD from
20 catalyzing the ATP-dependent addition of D-glutamate to an alanyl residue of the UDP-N-acetylmuramyl-L-alanine precursor.

As used herein a "modulator" is a compound or molecule that interacts with an aspect of cellular biochemistry to effect an increase or decrease in the amount of a polypeptide of MurD present at
25 the surface of a cell, or in the surrounding serum or media. The change in amount of the MurD polypeptide can be mediated by the effect of a modulator on the expression of the protein, e.g., the transcription, translation, post-translational processing, translocation or folding of the protein, or by affecting a component(s) of cellular biochemistry that
30 directly or indirectly participates in the expression of the protein. Alternatively, a modulator can act by accelerating or decelerating the turnover of the protein either by direct interaction with the receptor or by interacting with another component(s) of cellular biochemistry which directly or indirectly effects the change.

35 Nucleic acids encoding *murD* from *Streptococcus pyogenes* are useful in the expression and production of the *S. pyogenes* MurD

protein. The nucleic acids are also useful in providing probes for detecting the presence of *S. pyogenes*.

Polynucleotides

5 A preferred aspect of the present invention is disclosed in FIGS. 1A & 1B and SEQ ID NO:1, an isolated *murD* polynucleotide encoding a MurD protein from *Streptococcus pyogenes*, disclosed as follows:

GGATAGTGTG AGCTTAG
ATGAAAGTGA TAAGTAATTT TCAAAACAAA AAAATATTAA TATTGGGGTT
AGCCAAATCG GGCGAAGCAG CAGCAAAATT ATTGACCAAA CTTGGTGCTT
TAGTGACTGT TAATGATAGT AAACCATTTG ACCAAAATCC AGCGGCACAA
5 GCCTTGTTGG AAGAGGGGAT TAAGGTCATT TGTGGTAGCC ACCCAGTAGA
ATTATTAGAT GAGGACTTTG AGTACATGGT TAAAAACCCT GGGATTCCCTT
ATGATAATCC TATGGTTAAA CGCGCCCTTG CAAAGGAAAT TCCCATCTTG
ACTGAAGTAG AATTGGCTTA TTTCGTATCT GAAGCGCCTA TTATCGGGAT
TACAGGATCA AACGGGAAGA CAACCACAAC GACAATGATT GCCGATGTTT
10 TGAATGCTGG CGGGCAATCT GCACTCTTAT CTGGAAACAT TGGTTATCCT
GCTTCAAAAG TTGTTCAAAA AGCAATTGCT GGTGATACTT TGGTGATGGA
ATTGTCCTCT TTTCAATTAG TGGGAGTGAA TGCTTTTCGC CCTCATATTG
CTGTCATCAC TAATTTAATG CCGACTCATC TGGACTATCA TGGCAGTTTT
GAGGATTATG TTGCTGCTAA ATGGATGATT CAAGCTCAGA TGACAGAATC
15 AGACTACCTT ATTTTAAATG CTAATCAAGA GATTTCAGCA ACTCTAGCTA
AGACCACCCA AGCAACAGTG ATTCCTTTTT CAACTCAAAA AGTGGTTGAT
GGAGCTTATC TGAAGGATGG AATACTCTAT TTAAAGAAC AGGCGATTAT
AGCTGCAACT GACTTAGGTG TCCCAGGTAG CCACAACATT GAAAATGCCC
TAGCAACTAT TGCAGTTGCC AAGTTATCTG GTATTGCTGA TGATATTATT
20 GCCCAGTGCC TTTCACATTT TGGAGGCGTT AAACATCGTT TGCAACGGGT
TGGTCAAATC AAAGATATTA CCTTCTACAA TGACAGTAAG TCAACCAATA
TTTTAGCCAC TCAAAAAGCT TTATCAGGTT TTGATAACAG TCGCTTGATT
TTGATTGCTG GCGGTCTAGA TCGTGGCAAT GAATTTGACG ATTTGGTGCC
AGACCTTTTA GGACTTAAGC AGATGATTAT TTTGGGAGAA TCCGCAGAGC
25 GTATGAAGCG AGCTGCTAAC AAAGCAGAGG TCTCTTATCT TGAAGCTAGA
AATGTGGCAG AAGCAACAGA GCTTGCTTTT AAGCTGGCCC AAACAGGCGA
TACTATCTTG CTTAGCCCAG CTAATGCTAG CTGGGATATG TATCCTAATT
TTGAGGTTCTG TGGGGATGAA TTTTGGCAA CCTTTGATTG TTTAAGAGGA
GATGCCTAAT GCCTAAGAAG ATTTTATTTA CAGGTGGTGG
30

The isolated nucleic acid molecule of the present invention can include a deoxyribonucleic acid molecule (DNA), which can be single (coding or noncoding strand) or double stranded, as well as synthetic DNA, such as a synthesized, single stranded polynucleotide.

35 The isolated nucleic acid molecule of the present invention can also include a ribonucleic acid molecule (RNA).

The present invention also relates to recombinant vectors and recombinant hosts, both prokaryotic and eukaryotic, which contain the substantially purified nucleic acid molecules disclosed throughout this specification.

5 As used herein a "polynucleotide" is a nucleic acid of more than one nucleotide. A polynucleotide can be made up of multiple polynucleotide units that are referred to by description of the unit. For example, a polynucleotide can comprise within its bounds a polynucleotide(s) having a coding sequence(s), a polynucleotide(s) that is
10 a regulatory region(s) and/or other polynucleotide units commonly used in the art.

 An "expression vector" is a polynucleotide having regulatory regions operably linked to a coding region such that, when in a host cell, the vector can direct the expression of the coding sequence.
15 The use of expression vectors is well known in the art. Expression vectors can be used in a variety of host cells and, therefore, the regulatory regions are preferably chosen as appropriate for the particular host cell.

 A "regulatory region" is a polynucleotide that can promote
20 or enhance the initiation or termination of transcription or translation of a coding sequence. A regulatory region includes a sequence that is recognized by the RNA polymerase, ribosome, or associated transcription or translation initiation or termination factors of a host cell. Regulatory regions that direct the initiation of transcription or
25 translation can direct constitutive or inducible expression of a coding sequence.

 Polynucleotides of this invention contain full length or partial length sequences of the MurD gene sequences disclosed herein. Polynucleotides of this invention can be single or double stranded. If
30 single stranded, the polynucleotides can be a coding, "sense," strand or a complementary, "antisense," strand. Antisense strands can be useful as modulators of the gene by interacting with RNA encoding the MurD protein. Antisense strands are preferably less than full length strands having sequences unique or highly specific for RNA encoding the
35 protein.

 The polynucleotides can include deoxyribonucleotides, ribonucleotides or mixtures of both. The polynucleotides can be

produced by cells, in cell-free biochemical reactions or through chemical synthesis. Non-natural or modified nucleotides, including inosine, methyl-cytosine, deaza-guanosine, etc., can be present. Natural phosphodiester internucleotide linkages can be appropriate. However, 5 polynucleotides can have non-natural linkages between the nucleotides. Non-natural linkages are well known in the art and include, without limitation, methylphosphonates, phosphorothioates, phosphorodithionates, phosphoroamidites and phosphate ester linkages. Dephospho-linkages are also known, as bridges between 10 nucleotides. Examples of these include siloxane, carbonate, carboxymethyl ester, acetamidate, carbamate, and thioether bridges. "Plastic DNA," having, for example, N-vinyl, methacryloxyethyl, methacrylamide or ethyleneimine internucleotide linkages, can be used. "Peptide Nucleic Acid" (PNA) is also useful and resists degradation by 15 nucleases. These linkages can be mixed in a polynucleotide.

As used herein, "purified" and "isolated" are utilized interchangeably to stand for the proposition that the polynucleotides, proteins and polypeptides, or respective fragments thereof in question has been removed from its *in vivo* environment so that it can be 20 manipulated by the skilled artisan, such as but not limited to sequencing, restriction digestion, site-directed mutagenesis, and subcloning into expression vectors for a nucleic acid fragment as well as obtaining the protein or protein fragment in pure quantities so as to afford the opportunity to generate polyclonal antibodies, monoclonal 25 antibodies, amino acid sequencing, and peptide digestion. Therefore, the nucleic acids claimed herein can be present in whole cells or in cell lysates or in a partially purified or substantially purified form. A polynucleotide is considered purified when it is purified away from environmental contaminants. Thus, a polynucleotide isolated from cells 30 is considered to be substantially purified when purified from cellular components by standard methods while a chemically synthesized nucleic acid sequence is considered to be substantially purified when purified from its chemical precursors.

Polypeptides

The present invention also relates to a substantially purified form of the MurD protein from *Streptococcus pyogenes*, which is shown in FIGS. 1A & 1B and as set forth in SEQ ID NO:2, disclosed as follows:

5

MetLysValIleSerAsnPheGlnAsnLysLysIleLeuIleLeuGlyLeuAlaLysSer
 GlyGluAlaAlaAlaLysLeuLeuThrLysLeuGlyAlaLeuValThrValAsnAspSer
 LysProPheAspGlnAsnProAlaAlaGlnAlaLeuLeuGluGluGlyIleLysValIle
 CysGlySerHisProValGluLeuLeuAspGluAspPheGluTyrMetValLysAsnPro
 10 GlyIleProTyrAspAsnProMetValLysArgAlaLeuAlaLysGluIleProIleLeu
 ThrGluValGluLeuAlaTyrPheValSerGluAlaProIleIleGlyIleThrGlySer
 AsnGlyLysThrThrThrThrThrMetIleAlaAspValLeuAsnAlaGlyGlyGlnSer
 AlaLeuLeuSerGlyAsnIleGlyTyrProAlaSerLysValValGlnLysAlaIleAla
 GlyAspThrLeuValMetGluLeuSerSerPheGlnLeuValGlyValAsnAlaPheArg
 15 ProHisIleAlaValIleThrAsnLeuMetProThrHisLeuAspTyrHisGlySerPhe
 GluAspTyrValAlaAlaLysTrpMetIleGlnAlaGlnMetThrGluSerAspTyrLeu
 IleLeuAsnAlaAsnGlnGluIleSerAlaThrLeuAlaLysThrThrGlnAlaThrVal
 IleProPheSerThrGlnLysValValAspGlyAlaTyrLeuLysAspGlyIleLeuTyr
 PheLysGluGlnAlaIleIleAlaAlaThrAspLeuGlyValProGlySerHisAsnIle
 20 GluAsnAlaLeuAlaThrIleAlaValAlaLysLeuSerGlyIleAlaAspAspIleIle
 AlaGlnCysLeuSerHisPheGlyGlyValLysHisArgLeuGlnArgValGlyGlnIle
 LysAspIleThrPheTyrAsnAspSerLysSerThrAsnIleLeuAlaThrGlnLysAla
 LeuSerGlyPheAspAsnSerArgLeuIleLeuIleAlaGlyGlyLeuAspArgGlyAsn
 GluPheAspAspLeuValProAspLeuLeuGlyLeuLysGlnMetIleIleLeuGlyGlu
 25 SerAlaGluArgMetLysArgAlaAlaAsnLysAlaGluValSerTyrLeuGluAlaArg
 AsnValAlaGluAlaThrGluLeuAlaPheLysLeuAlaGlnThrGlyAspThrIleLeu
 LeuSerProAlaAsnAlaSerTrpAspMetTyrProAsnPheGluValArgGlyAspGlu
 PheLeuAlaThrPheAspCysLeuArgGlyAspAla

30

The present invention also relates to biologically active fragments and mutant or polymorphic forms of MurD polypeptide sequence as set forth as SEQ ID NO: 2, including but not necessarily limited to amino acid substitutions, deletions, additions, amino terminal truncations and carboxy-terminal truncations such that these
 35 mutations provide for proteins or protein fragments of diagnostic, therapeutic or prophylactic use and would be useful for screening for modulators, and/or inhibitors of MurD function.

Using the disclosure of polynucleotide and polypeptide sequences provided herein to isolate polynucleotides encoding naturally occurring forms of MurD, one of skill in the art can determine whether such naturally occurring forms are mutant or polymorphic forms of MurD by sequence comparison. One can further determine whether the encoded protein, or fragments of any MurD protein, is biologically active by routine testing of the protein or fragment in a *in vitro* or *in vivo* assay for the biological activity of the MurD protein. For example, one can express N-terminal or C-terminal truncations, or internal additions or deletions, in host cells and test for their ability to catalyzing the ATP-dependent addition of D-glutamate to an alanyl residue of the UDP-N-acetylmuramyl-L-alanine precursor.

It is known that there is a substantial amount of redundancy in the various codons which code for specific amino acids. Therefore, this invention is also directed to those DNA sequences encode RNA comprising alternative codons which code for the eventual translation of the identical amino acid, as shown below:

A=Ala=Alanine: codons GCA, GCC, GCG, GCU
 20 C=Cys=Cysteine: codons UGC, UGU
 D=Asp=Aspartic acid: codons GAC, GAU
 E=Glu=Glutamic acid: codons GAA, GAG
 F=Phe=Phenylalanine: codons UUC, UUU
 G=Gly=Glycine: codons GGA, GGC, GGG, GGU
 25 H=His=Histidine: codons CAC, CAU
 I=Ile=Isoleucine: codons AUA, AUC, AUU
 K=Lys=Lysine: codons AAA, AAG
 L=Leu=Leucine: codons UUA, UUG, CUA, CUC, CUG, CUU
 M=Met=Methionine: codon AUG
 30 N=Asp=Asparagine: codons AAC, AAU
 P=Pro=Proline: codons CCA, CCC, CCG, CCU
 Q=Gln=Glutamine: codons CAA, CAG
 R=Arg=Arginine: codons AGA, AGG, CGA, CGC, CGG, CGU
 S=Ser=Serine: codons AGC, AGU, UCA, UCC, UCG, UCU
 35 T=Thr=Threonine: codons ACA, ACC, ACG, ACU
 V=Val=Valine: codons GUA, GUC, GUG, GUU
 W=Trp=Tryptophan: codon UGG

Y=Tyr=Tyrosine: codons UAC, UAU

Therefore, the present invention discloses codon redundancy which can result in differing DNA molecules expressing an identical protein. For purposes of this specification, a sequence bearing one or more replaced
5 codons will be defined as a degenerate variation. Also included within the scope of this invention are mutations either in the DNA sequence or the translated protein which do not substantially alter the ultimate physical properties of the expressed protein. For example, substitution
10 of valine for leucine, arginine for lysine, or asparagine for glutamine may not cause a change in functionality of the polypeptide. However, any given change can be examined for any effect on biological function by simply assaying for the ability to catalyze the ATP-dependent addition of D-glutamate to an alanyl residue of the UDP-N-acetylmuramyl-L-
15 alanine precursor as compared to an unaltered MurD protein.

It is known that DNA sequences coding for a peptide can be altered so as to code for a peptide having properties that are different than those of the naturally occurring peptide. Methods of altering the DNA sequences include but are not limited to site directed mutagenesis.
20 Examples of altered properties include but are not limited to changes in the affinity of an enzyme for a substrate or a receptor for a ligand.

As used herein, a "biologically active equivalent" or "functional derivative" of a wild-type MurD possesses a biological activity that is substantially similar to the biological activity of a wild type MurD.
25 The term "functional derivative" is intended to include the "fragments," "mutants," "variants," "degenerate variants," "analogs" and "homologues" or to "chemical derivatives" of a wild type MurD protein. The term "fragment" is meant to refer to any polypeptide subset of wild-type MurD. The term "mutant" is meant to refer to a molecule that may
30 be substantially similar to the wild-type form but possesses distinguishing biological characteristics. Such altered characteristics include but are in no way limited to altered substrate binding, altered substrate affinity and altered sensitivity to chemical compounds affecting biological activity of the MurD or MurD functional derivative.
35 The term "variant" is meant to refer to a molecule substantially similar in structure and function to either the entire wild-type protein or to a fragment thereof. A molecule is "substantially similar" to a wild-type

MurD-like protein if both molecules have substantially similar structures or if both molecules possess similar biological activity. Therefore, if the two molecules possess substantially similar activity, they are considered to be variants even if the structure of one of the molecules is not found in the other or even if the two amino acid sequences are not identical. The term "analog" refers to a molecule substantially similar in function to either the full-length MurD protein or to a biologically active fragment thereof.

As used herein in reference to a MurD gene or encoded protein, a "polymorphic" MurD is a MurD that is naturally found in the population of *Streptococci* at large. A polymorphic form of MurD can have a different nucleotide sequence from the particular MurD genes and proteins disclosed herein. However, because of silent mutations, a polymorphic MurD gene can encode the same or different amino acid sequence as that disclosed herein. Further, some polymorphic forms MurD will exhibit biological characteristics that distinguish the form from wild-type MurD activity, in which case the polymorphic form is also a mutant.

A protein or fragment thereof is considered purified or isolated when it is obtained at least partially free from its natural environment at a concentration at least about five-fold to ten-fold higher than that found in nature. A protein or fragment thereof is considered substantially pure if it is obtained at a concentration of at least about 100-fold higher than that found in nature. A protein or fragment thereof is considered essentially pure if it is obtained at a concentration of at least about 1000-fold higher than that found in nature. We most prefer proteins that have been purified to homogeneity, that is, at least 10,000 - 100,000 fold.

30 Probes and Primers

Polynucleotide probes comprising full length or partial sequences of SEQ ID NO: 1 can be used to determine whether a cell or sample contains *S. pyogenes* MurD DNA or RNA. The effect of modulators that effect the transcription of the *murD* gene can be studied via the use of these probes. A preferred probe is a single stranded antisense probe having at least the full length of the coding sequence of MurD. It is also preferred to use probes that have less than the full

length sequence, and contain sequences highly specific for *S. pyogenes murD* DNA or RNA. The identification of a sequence(s) for use as a specific probe is well known in the art and involves choosing a sequence(s) that is unique to the target sequence, or is highly specific thereto. It is preferred that polynucleotides that are probes have at least about 25 nucleotides, more preferably about 30 to 35 nucleotides. The longer probes are believed to be more specific for *S. pyogenes murD* gene(s) and RNAs and can be used under more stringent hybridization conditions. Longer probes can be used but can be more difficult to prepare synthetically, or can result in lower yields from a synthesis. Examples of sequences that are useful as probes or primers for *S. pyogenes murD* gene(s) are Primer A (sense) 5'-GGA TAG TGT GAG CTT AGA TGA AAG TGA TAA GT-3' nucleotides 1 to 32 of SEQ ID NO:1 and Primer B (antisense) 5'-CCA CCA CCT GTA AAT AAA ATC TTC TTA GGC ATT AGG CAT CTC C-3' nucleotides 1354 to 1397 of SEQ ID NO:1. These primers are the first 32 (A) and the complement of the last 43 (B) nucleotides, respectively, of SEQ ID NO:1. However, one skilled in the art will recognize that these are only a few of the useful probe or primer sequences that can be derived from SEQ ID NO:1.

Polynucleotides having sequences that are unique or highly specific for *S. pyogenes murD* can be used as primers in amplification reaction assays. These assays can be used in tissue typing as described herein. Additionally, amplification reactions employing primers derived from *S. pyogenes murD* sequences can be used to obtain amplified *S. pyogenes murD* DNA using the *murD* DNA of the cells as an initial template. The *murD* DNA so obtained can be a mutant or polymorphic form of *S. pyogenes murD* that differs from SEQ ID NO:1 by one or more nucleotides of the MurD open reading frame or sequences flanking the ORF. The differences can be associated with a non-defective naturally occurring form or with a defective form of MurD. Thus, polynucleotides of this invention can be used in identification of various *S. pyogenes murD* genes or the detection of an organism having a *S. pyogenes murD* gene. Many types of amplification reactions are known in the art and include, without limitation, Polymerase Chain Reaction, Reverse Transcriptase Polymerase Chain Reaction, Strand Displacement Amplification and Self-Sustained Sequence Reaction.

Any of these or like reactions can be used with primers derived from SEQ ID NO:1.

Expression of MurD

5 A variety of expression vectors can be used to express recombinant MurD in host cells. Expression vectors are defined herein as DNA sequences that are required for the transcription of cloned DNA and the translation of their mRNAs in an appropriate host. Such vectors can be used to express bacteria DNA in a variety of hosts such as
10 bacteria, bluegreen algae, plant cells, insect cells and animal cells. Specifically designed vectors allow the shuttling of DNA between hosts such as bacteria-yeast or bacteria-animal cells. An appropriately constructed expression vector should contain: an origin of replication for autonomous replication in host cells, selectable markers, a limited
15 number of useful restriction enzyme sites, a potential for high copy number, and active promoters. A promoter is defined as a DNA sequence that directs RNA polymerase to bind to DNA and initiate RNA synthesis. A strong promoter is one which causes mRNAs to be initiated at high frequency. Expression vectors can include, but are not
20 limited to, cloning vectors, modified cloning vectors, specifically designed plasmids or viruses.

In particular, a variety of bacterial expression vectors can be used to express recombinant MurD in bacterial cells. Commercially available bacterial expression vectors which are suitable for
25 recombinant MurD expression include, but are not limited to pQE (Qiagen), pET11a or pET15b (Novagen), lambda gt11 (Invitrogen), and pKK223-3 (Pharmacia).

Alternatively, one can express *murD* DNA in cell-free transcription-translation systems, or *murD* RNA in cell-free translation
30 systems. Cell-free synthesis of MurD can be in batch or continuous formats known in the art.

One can also synthesize MurD chemically, although this method is not preferred.

A variety of host cells can be employed with expression
35 vectors to synthesize MurD protein. These can include *E. coli*, *Bacillus*, and *Salmonella*. Insect and yeast cells can also be appropriate.

Following expression of MurD in a host cell, MurD polypeptides can be recovered. Several protein purification procedures are available and suitable for use. MurD protein and polypeptides can be purified from cell lysates and extracts, or from culture medium, by various combinations of, or individual application of methods including ultrafiltration, acid extraction, alcohol precipitation, salt fractionation, 5 ionic exchange chromatography, phosphocellulose chromatography, lecithin chromatography, affinity (*e.g.*, antibody or His-Ni) chromatography, size exclusion chromatography, hydroxylapatite adsorption chromatography and chromatography based on hydrophobic or hydrophilic interactions. In some instances, protein denaturation and refolding steps can be employed. High performance liquid chromatography (HPLC) and reversed phase HPLC can also be useful. Dialysis can be used to adjust the final buffer composition.

15 The MurD protein itself is useful in assays to identify compounds that modulate the activity of the proteins -- including compounds that inhibit the activity of the protein. The MurD protein is also useful for the generation of antibodies against the protein, structural studies of the protein, and structure/function relationships of 20 the protein.

Modulators and Inhibitors of MurD

The present invention is also directed to methods for screening for compounds or molecules which modulate or inhibit the expression of DNA or RNA encoding a MurD protein. Compounds or 25 molecules which modulate or inhibit MurD can be DNA, RNA, peptides, proteins, or non-proteinaceous organic compounds or molecules. Compounds that modulate the expression of DNA or RNA encoding MurD or are inhibitors of the biological function thereof can be detected by a variety of assays. The assay can be a simple "yes/no" assay to 30 determine whether there is a change in expression or function. The assay can be made quantitative by comparing the expression or function of a test sample with the levels of expression or function in a standard sample, that is, a control. Kits containing MurD, antibodies to MurD, or modified MurD can be prepared by known methods for such uses. 35

The DNA molecules, RNA molecules, recombinant protein and antibodies of the present invention can be used to screen and

measure levels of MurD. The recombinant proteins, DNA molecules, RNA molecules and antibodies lend themselves to the formulation of kits suitable for the detection and analysis of MurD. Such a kit would comprise a compartmentalized carrier suitable to hold in close
5 confinement at least one container. The carrier would further comprise reagents such as recombinant MurD or anti- MurD antibodies suitable for detecting MurD. The carrier can also contain a means for detection such as labeled antigen or enzyme substrates or the like.

10 Pharmaceutical Compositions

Pharmaceutically useful compositions comprising inhibitors of MurD can be formulated according to known methods such as by the admixture of a pharmaceutically acceptable carrier. Examples of such carriers and methods of formulation can be found in
15 Remington's Pharmaceutical Sciences. To form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of the inhibitor.

Therapeutic or diagnostic compositions of the invention are administered to an individual in amounts sufficient to treat or diagnose
20 disorders. The effective amount can vary according to a variety of factors such as the individual's condition, weight, sex and age. Other factors include the mode of administration. The appropriate amount can be determined by a skilled physician

The pharmaceutical compositions can be provided to the
25 individual by a variety of routes such as subcutaneous, topical, oral and intramuscular.

The term "chemical derivative" describes a molecule that contains additional chemical moieties which are not normally a part of the base molecule. Such moieties can improve the solubility, half-life,
30 absorption, etc. of the base molecule. Alternatively the moieties can attenuate undesirable side effects of the base molecule or decrease the toxicity of the base molecule. Examples of such moieties are described in a variety of texts, such as Remington's Pharmaceutical Sciences.

Compounds identified according to the methods disclosed
35 herein can be used alone at appropriate dosages. Alternatively, co-administration or sequential administration of other agents can be desirable.

The present invention also provides a means to obtain suitable topical, oral, systemic and parenteral pharmaceutical formulations for use in the methods of treatment of the present invention. The compositions containing compounds or molecules identified according to this invention as the active ingredient can be administered in a wide variety of therapeutic dosage forms in conventional vehicles for administration. For example, the compounds can be administered in such oral dosage forms as tablets, capsules (each including timed release and sustained release formulations), pills, powders, granules, elixirs, tinctures, solutions, suspensions, syrups and emulsions, or by injection. Likewise, they can also be administered in intravenous (both bolus and infusion), intraperitoneal, subcutaneous, topical with or without occlusion, or intramuscular form, all using forms well known to those of ordinary skill in the pharmaceutical arts.

Advantageously, compounds of the present invention can be administered in a single daily dose, or the total daily dosage can be administered in divided doses of two, three or four times daily. Furthermore, compounds for the present invention can be administered in intranasal form via topical use of suitable intranasal vehicles, or via transdermal routes, using those forms of transdermal skin patches well known to those of ordinary skill in that art. To be administered in the form of a transdermal delivery system, the dosage administration will, of course, be continuous rather than intermittent throughout the dosage regimen.

For combination treatment with more than one active agent, where the active agents are in separate dosage formulations, the active agents can be administered concurrently, or they each can be administered at separately staggered times.

The dosage regimen utilizing the compounds of the present invention is selected in accordance with a variety of factors including type, species, age, weight, sex and medical condition of the patient; the severity of the condition to be treated; the route of administration; the renal, hepatic and cardiovascular function of the patient; and the particular compound thereof employed. A physician or veterinarian of ordinary skill can readily determine and prescribe the effective amount of the drug required to prevent, counter or arrest the progress of the condition. Optimal precision in achieving concentrations of drug within

the range that yields efficacy without toxicity requires a regimen based on the kinetics of the drug's availability to target sites. This involves a consideration of the distribution, equilibrium, and elimination of a drug.

The following examples are presented by the way of illustration and, because various other embodiments will be apparent to those in the art, the following is not to be construed as a limitation on the scope of the invention:

EXAMPLE 1

10 General Materials and Methods

All reagents were purchased from Sigma Chemical Co, St. Louis, MO, unless otherwise indicated. UDP-N-acetylmuramyl-L-alanine was synthesized and purified by a method known in the art (Jin, H., Emanuele, J. J., Jr., Fairman, R., Robertson, J. G., Hail, M. E., Ho, H.-T., Falk, P. and Villafranca, J. J, 1996, Structural studies of *Escherichia coli* UDP-N-acetylmuramate: L-alanine ligase, Biochemistry 35: 14423-14431).

DNA manipulations reagents and techniques. Restriction endonucleases and T4 ligase were obtained from Gibco-BRL. Agarose gel electrophoresis and plasmid DNA preparations were performed according to published procedures (Sambrook, J., E. F. Fritsch, and T. Maniatis, 1989, Molecular cloning: a Laboratory Manual, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory). Recombinant plasmids containing *S. pyogenes murD* were propagated in *E. coli* DH5a (Gibco-BRL, Rockville, MD) prior to protein expression in *E. coli* BL21(DE3)/plysS (Novagen, Madison, WI). SDS-PAGE was performed with precast gels (Novagen). DNA sequences were determined using an automated ABI Prism DNA sequencer (Perkin-Elmer ABI, Foster City, CA).

EXAMPLE 2

Cloning of *Streptococcus pyogenes murD*

Genomic DNA from *S. pyogenes* (strain MB4439) was prepared from 100 ml late stationary phase culture in Brain Heart Infusion broth (Difco, Detroit, MI). Cells were washed with 0.2 M sodium acetate, suspended in 10 ml of TEG (100 mM Tris, pH 7, containing 10 mM EDTA and 25% glucose) and lysed by incubation with 200 µg of N-acetylmuramidase (Sigma) for 1h at 37°C. Chromosomal DNA was purified from the cell lysate using a Qiagen (Santa Clarita, CA) genomic DNA preparation kit and following the manufacturers protocol. Briefly, the cell lysate was treated with protease K at 50°C for 45 min, loaded onto an equilibrated Qiagen genomic tip, entered into the resin by centrifugation at 3000 rpm for 2 min. Following washing the genomic tip, the genomic DNA was eluted in distilled water and kept at 4°C. Approximately 50 ng genomic DNA was used as a template in PCR reactions to clone *murD*.

A *murD* homologue was localized in contig. sequence #306 of *S. pyogenes* genomic DNA sequence (Roe, B. A., S. Clifton, M. McShan and J. Ferretti. The Streptococcal genome sequencing project, University of Oklahoma, Oklahoma, USA) by searching the *S. pyogenes* genome for homology to *S. aureus murD*. The cloning of *S. aureus murD* is described in copending, commonly owned U.S. Application Ser. No. _____, filed _____, incorporated herein by reference in its entirety. The sequence of *S. aureus murD* is provided therein as SEQ ID NO:1.

Two oligonucleotide primers (Midland Certified Reagent, Midland, TX) complementary to sequences at the 5' and the 3' ends of *S. pyogenes murD* were used to clone this gene using KlenTach advantage polymerase (Clontech, Palo Alto, CA). The primer nucleotide sequences were as follows: 5'-GG ATA GTG TGA GCT CAT ATG AAA GTG ATA AGT-3' (the first 34 nucleotides of SEQ ID NO: 1) and 5'-TTC AAG GTG ACA TGA CCT ACG GAT CCA CC-3' (the complement of the last 35 nucleotides of SEQ ID NO: 1). A PCR product representing *S. pyogenes murD* was verified by nucleotide sequence, digested with NdeI and BamHI, and cloned between the NdeI and BamHI sites of pET-15b,

creating plasmid pSpyMurD. This plasmid was used for expression of the *murD* gene in *E. coli*.

The plasmid pSpyMurD has been deposited with the American Type Culture Collection on October 31, 1997, under the terms of the Budapest Treaty for the Deposit of Microorganisms and has been designated as ATCC 98575. The deposited material is provided as a convenience and is not an admission that the deposited material is required to describe or practice the invention. The sequence of the polynucleotide of the deposit, and the encoded amino acid sequence, are incorporated herein by reference and are controlling in the event of a conflict with any description of the sequences provided in this specification or the associated drawings. A license may be required to make, use, sell or offer to sell the polynucleotide of the deposit or a protein of the amino acid sequence encoded by the polynucleotide. No such license is granted herein.

EXAMPLE 3

Sequence analysis of *Streptococcus pyogenes murD*

The nucleotide sequence of *murD*, determined in both orientations, and the deduced amino acid sequence of the MurD protein is depicted in FIGS. 1A & 1B. Sequence comparison using the BLAST (1) algorithm against the GenBank database showed that, to varying degrees, the cloned region is homologous (% similar, % identical) to *murD* genes from *B. subtilis* (67%, 48%) and *S. aureus murD* (66%, 47%).

In addition to the several regions of homology that are conserved among MurD proteins from both Gram-positive and Gram-negative bacteria (FIGS. 2A & 2B), there are at least two regions that are particularly conserved among MurD proteins of Gram-negative bacteria. The first region is amino acid residues ValLysAsnProGlyIleProTyr (amino acids 77-84 of *S. pyogenes murD*; SEQ ID NO: 2) and the second region is amino acid residues GlyGlyLeuAspArgGlyAsnGluPheAsp (codons 354-363 of *S. pyogenes murD*; SEQ ID NO:2).

Multiple sequence alignments of MurC (Ikeda, M., M. Wachi, H. K. Jung, F. Ishino, and M. Matsuhashi, 1990, Nucleotide sequence involving murG and murC in the mra gene cluster of *Escherichia coli*. Nucleic Acids Res. 18:4014), MurD, MurE (Tao, J.S, and E. E., Ishiguro, 1989, Nucleotide sequence of the murE gene of *Escherichia coli*. Can. J. Microbiol. 35:1051-1054), and MurF (Parquet, C., D. Mengin-Lecreulx, B. Flouret, D. Mengin-Lecreulx, and J. van Heijenoort, 1989, Nucleotide sequence of the murF gene encoding the UDP-MurNAc-pentapeptide synthetase of *Escherichia coli*, Nucleic Acids Res. 17:5379) proteins from several bacterial genera revealed four regions of homology with certain residues conserved amongst Mur ligases of both Gram-positive and Gram-negative bacteria (Eveland, S. S., D. L. Pompliano, and M. S. Anderson, 1997, Conditionally lethal *Escherichia coli* murein mutants contain point defects that map to regions conserved among murein and folyl poly-g-glutamate ligases: Identification of a ligase superfamily. Biochemistry, 36: 6223-6229, Ikeda, M., M. Wachi, H. K. Jung, F. Ishino, and M. Matsuhashi, 1990, Homology among MurC, MurD, MurE and MurF proteins in *Escherichia coli* and that between *E. coli* murG and a possible murG protein in *Bacillus subtilis*. J. Gen. Appl. Microbiol. 36: 179-187). The homologous regions may correlate with the catalytic functions of these enzymes (Eveland, *et al.*, 1997). Most notable is the putative ATP binding region I that was found in MurF (Parquet, C., D. Mengin-Lecreulx, B. Flouret, D. Mengin-Lecreulx, and J. van Heijenoort, 1989, Nucleotide sequence of the murF gene encoding the UDP-MurNAc-pentapeptide synthetase of *Escherichia coli*, Nucleic Acids Res. 17:5379) and is also conserved in *S. pyogenes* MurD protein GlySerAspGlyLysThrThr (codons 119 to 125, SEQ ID NO:2). While region I is an ATP-binding domain (Ikeda, *et al.*, 1990), the function of the other homologous regions is unknown. Of these three other regions, only two of them appear to be present in *S. pyogenes murD* (FIG. 3). Region III, which contains two contiguous Asp residues, is not conserved in *S. pyogenes* MurD, although it is present in MurD of *B. subtilis*. Within regions I, II, and IV, the specific residues conserved in the known murein ligases are also present in *S. pyogenes* MurD (FIG. 3).

EXAMPLE 4

Overexpression, purification and enzymatic activity of *Streptococcus pyogenes* MurD

murD was cloned into the expression vector pET-15b (Novagen) as described above to create plasmid pSpyMurD. The pET-15b vector incorporates the 6xHistidine-tag into the protein construct to allow rapid purification of MurD by affinity chromatography. The pET (Plasmids for Expression by T7 RNA polymerase) plasmids are derived from pBR322 and designed for protein over-production in *E. coli*. The vector pET-15b contains the ampicillin resistance gene, ColE1 origin of replication in addition to T7 phage promoter and terminator. The T7 promoter is recognized by the phage T7 RNA polymerase but not by the *E. coli* RNA polymerase. A host *E. coli* strain such as BL21(DE3)pLysS is engineered to contain integrated copies of T7 RNA polymerase under the control of lacUV5 that is inducible by IPTG. Production of a recombinant protein in the *E. coli* strain BL21(DE3)pLysS occurs after expression of T7RNA polymerase is induced.

The pSpyMurD plasmid was introduced into the host strain BL21 DE3/pLysS (Novagen) for expression of His-tagged MurD. Colonies were grown at 37°C in 100 ml of LB broth containing 100 mg/ml ampicillin and 32 µg/ml chloramphenicol. When cultures reached a cell density of A₆₀₀=0.5, cells were pelleted and then resuspended in M9ZB medium (Novagen) containing 1 mM IPTG. Cells were induced for 3 h at 30°C, pelleted at 3000g, and frozen at -80°C.

Cultures containing either the recombinant plasmid pSpyMurD or the control plasmid vector, pET-15b were grown at 30°C and induced with IPTG. Cells transformed with pSpyMurD contained an inducible protein of approximately 52 kDa, corresponding to the expected size of *S. pyogenes* MurD protein as shown by SDS-PAGE. There were no detectable protein bands after induction of cells transformed with the control plasmid vector, pET-15b.

Purification of recombinant MurD enzyme.

The cell pellet from 100 ml of induced culture prepared as described above was resuspended in 10 ml BT buffer (50 mM bis-tris-

propane, pH 8.0, containing 100 mM potassium chloride and 1% glycerol) at 4°C. Cells were lysed either by freeze-thaw or by French Press. After centrifugation, the supernatant was mixed with 15 ml of freshly prepared Talon resin and incubated for 30 min at room temp.

- 5 The resin was washed twice by centrifugation with 25 ml of BT buffer at room temperature. Finally, the resin loaded into a column and washed with 20 ml of BT, pH 7.0, containing 5 mM imidazole. Protein was eluted with 20 ml of BT buffer pH 8.0, containing 100 mM imidazole. Fractions (0.5 ml) were collected and analyzed by SDS-Gel
- 10 electrophoresis. Fractions containing MurD were pooled, cleaved with thrombin (20 units/mg of His-tagged MurD) to remove the histidine tag. Finally the MurD protein was purified on Superdex 75 size exclusion column and eluted in 10 mM tris, pH 7.4, 150 mM NaCl and dialyzed in 100 mM MOPS, pH 7.5, using a Slide-A-Lyzer (Pierce, Rockford, IL).
- 15 This resulted in a partially purified preparation of *S. pyogenes* MurD protein that could be used in activity assays. The protein may be purified further, if desired, using methods known in the art.

Assay for activity of MurD enzyme.

- The ATP-dependent MurD activity was assayed by
- 20 monitoring the formation of product ADP using the pyruvate kinase and lactate dehydrogenase coupled enzyme assay. The reaction was monitored spectrophotometrically.

- Typically, the assay contained 100 mM BIS-TRIS-propane, pH 8.0, 200 μ M NADH, 1 mM ATP, 20 mM PEP, 5 mM MgCl_2 , 1 mM
- 25 DTT, 350 μ M UDP-N-acetyl-muramyl-L-alanine, 1 mM D-glutamate, 33 units/ml of pyruvate kinase and 1660 units/ml of lactate dehydrogenase in a final volume of 200 or 400 μ L. The mixture was incubated at 25°C for 5 min and the reaction initiated by the addition of 1-10 μ g of MurD. ADP formation was monitored by the decrease in absorbance at 340 nm as a
- 30 function of time using a Molecular Devices SpectramaxPlus microtiterplate spectrophotometer (for 200 μ l assays) or a Hewlett-Packard HP8452A spectrophotometer equipped with a circulating water bath (for 400 μ l assays). Rates were calculated from the linear portions of the progress curves using the extinction coefficient for NADH, $\epsilon = 6220 \text{ cm}^{-1}$
- 35 M^{-1} . One unit of MurD activity is equal to 1 μ mol of ADP formed per min at 25°C. MurD activity co-eluted with a ~52 kDa protein.

Table 1
Specific activities of recombinant MurD
from *E. coli* and *S. pyogenes*.

Species	protein used (mg)	units (μ mole/min) ¹	Specific activity (μ mol/min mg)
<i>E. coli</i> ²	0.00277	0.012250	4.40
<i>S. pyogenes</i> ³	0.00137	0.00017	0.13

5 ¹ Concentration of UDP-N-acetylmuramyl-L-ala, D-glutamate, ATP were 350 μ M, 1mM, and 1 mM, respectively. Volume of the reactions were 200 μ l at 25°C.

² *E. coli* Mur D was prepared described in Pratviel-Sosa, *et al.* (1991).

³ *S. pyogenes* MurD was partially purified as described above.

10

 Assays have been conducted using 120 and 350 μ M UDP-N-acetyl-muramyl-L-alanine. However, it has been observed that at the higher level of 350 μ M UDP-N-acetyl-muramyl-L-alanine, substrate inhibition of the *E. coli* MurD occurs. At the lower level, the specific
15 activity of the *E. coli* enzyme can be in the area of 8 units/mg. It is unclear whether the *S. pyogenes* enzyme is similarly inhibited.

EXAMPLE 5

Screening for inhibitors of MurD

20

 One assay for the measurement of the activity of MurD is provided in Example 4. That assay, and other assays for MurD activity can be adapted for screening assays to detect inhibitors of MurD. For inhibition studies, inhibitors in DMSO are added at the desired concentration to the assay mixture. In a separate, control reaction, only
25 DMSO is added to the assay mixture. The reactions are initiated by the addition of enzyme (MurD). Rates were calculated as described above. Relative activities were calculated from the equation 1:

relative activity = rate with inhibitor/rate without inhibitor. (1)

Inhibition constant (IC₅₀) values were determined from a range of inhibitor concentrations and calculated from the equation 2 using Sigma Plot (Jandel Scientific).

$$\text{relative activity} = 1/(1 + [I]/IC_{50}) \quad (2)$$

We prefer inhibitors of MurD that result in relative activities of the MurD enzyme of at least less than 75%, more preferably, 25-50% or 10-25%. We most prefer inhibitors resulting in relative activities of less than 20%, particularly less than 10% of the activity or Mur D in the absence of the inhibitor.

We also prefer inhibitors that effectively lower the relative activity of MurD when present at very low levels.

EXAMPLE 8

Pharmaceutical compositions

An appropriate amount of an inhibitor of MurD is formulated in a pharmaceutically acceptable carrier, in an appropriate dosage form.

EXAMPLE 9

Therapy using inhibitors of MurD

A patient presenting with an indication of infection with a microorganism susceptible to inhibitors of MurD, *e.g.*, gram positive organisms, including *S. pyogenes*, can be treated by administration of inhibitors of MurD. Physicians skilled in the art are familiar with administering therapeutically effective amounts of inhibitors of microbial enzymes. Such skilled persons can readily determine an appropriate dosing scheme to achieve a desired therapeutic effect.

REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. Myers, and D. J. Lipman, 1990, Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- 5 Blanot, D., A. Kretsovali, M. Abo-Ghalia, D. Mengin-Lecreulx, and J. van Heijenoort, 1983, Synthesis of analogues of precursors of bacterial peptidoglycan. p. 311-314. In *peptides*. Blaha, K. and P. Malon, eds. PP311-314. Walter de Gruyter and Co. Berlin, New York.
- 10 Daniel, R. A., and J. Errington, 1993, DNA sequence of the *murE-murD* region of *Bacillus subtilis* 168, *J. Gen. Microbiol.* 139:361-370.
- Eveland, S. S., D. L. Pompliano, and M. S. Anderson, 1997, Conditionally lethal *Escherichia coli* murein mutants contain point
- 15 defects that map to regions conserved among murein and folyl poly-glutamate ligases: Identification of a ligase superfamily. *Biochemistry*, 36: 6223-6229.
- Flouret, B., D. Mengin-Lecreulx, and J. van Heijenoort, 1981, Reverse-phase high-pressure liquid chromatography of Uridine
- 20 Diphosphate N-Acetylmuramyl peptide precursors of bacterial cell wall peptidoglycan, *Anal. Biochem.* 114:59-63.
- Henriques, A. O. de Lencaster, H. and P. J. Piggot, 1992, A *Bacillus subtilis* morphogene cluster that includes *spoVE* is homologous to the *mra* region of *Escherichia coli*. *Biochimie.* 74: 735-748.
- 25 Ikeda, M., M. Wachi, F. Ishino, and M. Matsuhashi, 1990, Nucleotide sequence involving *murD* and an open reading frame ORF-Y spacing *murF* and *ftsW* in *Escherichia coli*. *Nucleic Acids Res.* 18:1058.
- Ikeda, M., M. Wachi, H. K. Jung, F. Ishino, and M. Matsuhashi, 1990, Nucleotide sequence involving *murG* and *murC* in
- 30 the *mra* gene cluster of *Escherichia coli*. *Nucleic Acids Res.* 18:4014.
- Ikeda, M., M. Wachi, H. K. Jung, F. Ishino, and M. Matsuhashi, 1990, Homology among *MurC*, *MurD*, *MurE* and *MurF* proteins in *Escherichia coli* and that between *E. coli murG* and a possible *murG* protein in *Bacillus subtilis*. *J. Gen. Appl. Microbiol.* 36:
- 35 179-187.

- Ito, E. and J. L. Strominger, 1962, Enzymatic synthesis of the peptide in bacterial uridine nucleotides: Enzymatic addition of L-alanine, D-glutamic acid, and L-lysine. *J. Biol. Chem.* 237: 2689-2695.
- Ito E. and J. L. Strominger, 1973, Enzymatic synthesis of
5 the peptide in bacterial uridine nucleotides: Comparative biochemistry, *J. Biol. Chem.* 248: 3131-3136.
- Jin, H., Emanuele, J. J., Jr., Fairman, R., Robertson, J. G., Hail, M. E., Ho, H.-T., Falk, P. and Villafranca, J. J., 1996, Structural studies of *Escherichia coli* UDP-N-acetylmuramate: L-
10 alanine ligase, *Biochemistry* 35: 14423-14431.
- Lugtenberg, E. J. J., 1972, Studies on *Escherichia coli* enzymes involved in the synthesis of Uridine Diphosphate-N-Acetyl-Muramyl-pentapeptide. *J. Bacteriol.* 110:26-34.
- Mengin-Lecreulx, D., B. Flouret, and J. van Heijenoort,
15 1982, Cytoplasmic steps of peptidoglycan synthesis in *Escherichia coli*. *J. Bacteriol.* 151: 1109-1117.
- Mengin-Lecreulx, D., C Michaud, C. Richaud, D. Blanot and J. van Heijenoort, 1988, Incorporation of LL-diaminopimelic acid into peptidoglycan of *Escherichia coli* mutants lacking diaminopimelate
20 epimerase encoded by *dapF*, *J. Bacteriol.* 170:2031-2039.
- Mengin-Lecreulx, D., C Parquet, L. Desviat, J. Pla, B. Flouret, J. Ayala and J. van Heijenoort, 1989, Organization of the *murE-murG* region of *Escherichia coli*: Identification of the *murD* gene encoding the D-glutamic-acid-adding enzyme, *J. Bacteriol.* 171: 6126-
25 6134.
- Mengin-Lecreulx, D. and J. van Heijenoort, 1990, Nucleotide sequence of the *murD* gene encoding the UDP-MurNAc-L-Ala-D-Glu synthetase of *Escherichia coli*. *Nucleic Acids Research* 18:183.
- 30 Michaud, C. D. Blanot, B. Flouret, and J. van Heijenoort, 1987, Partial purification and specificity studies of the D-glutamate-adding and D-alanyl-D-alanine-adding enzymes from *Escherichia coli* K12. *Eur. J. Biochem.* 166: 631-637.
- 35 Michaud, C. , C., Parquet, B. Flouret, D. Blanot, and J. van Heijenoort, 1990, Revised interpretation of the sequence containing the *murE* gene encoding the UDP-N-acetyl-muramyl-tripeptide synthetase of *Escherichia coli*. *Biochem. J.* 269:277-280.

Nathenson, S. G., J. L. Strominger, and E. Ito, 1964, Enzymatic synthesis of the peptide in bacterial uridine nucleotides: purification and properties of D-Glutamic acid-adding enzyme, *J. Biol. Chem.* 239: 1773-1776.

- 5 Parquet, C., D. Mengin-Lecreulx, B. Flouret, D. Mengin-Lecreulx, and J. van Heijenoort, 1989, Nucleotide sequence of the murF gene encoding the UDP-MurNAc-pentapeptide synthetase of *Escherichia coli.*, *Nucleic Acids Res.* 17:5379.

- Pratviel-Sosa F, D. Mengin-Lecreulx and J. van Heijenoort,
10 1991, Over-production, purification and properties of the uridine diphosphate N-acetylmuramoyl-L-alanine:D-glutamate ligase from *Escherichia coli.* *Eur. J. Biochem.* 202 (3):1169-1176.

- Roe, B. A., S. Clifton, M. McShan and J. Ferretti. The Streptococcal genome sequencing project, University of Oklahoma,
15 Oklahoma, USA.

 Rogers, H. J., H. R. Perkins, and J. B. Ward, 1980, Biosynthesis of peptidoglycan. p. 239-297. In *Microbial cell walls and membranes.* Chapman & Hall Ltd. London.

- Sambrook, J., E. F. Fritsch, and T. Maniatis, 1989,
20 Molecular cloning: a Laboratory Manual, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

 Schleifer, K. H. and O. Kandler, 1972, Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Review.* 36:407-477.

- 25 Tanner, M.E., S. Vaganay, van Heijenoort, J., and D. Blanot, 1996, Phosphinate Inhibitors of the D-Glutamic Acid-Adding Enzyme of Peptidoglycan Biosynthesis. *J. Org. Chem.* 61:1756-1760.

 Tao, J.S, and E. E., Ishiguro, 1989, Nucleotide sequence of the murE gene of *Escherichia coli.* *Can. J. Microbiol.* 35:1051-1054.

WO 99/23201

PCT/US98/23054

SEQUENCE LISTING

<110> Merck & Co., Inc.

<120> MURD PROTEIN AND GENE OF STREPTOCOCCUS
PYOGENES

<130> 20070 PCT

<150> 60/064,205

<151> 1997-11-04

<160> 6

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 1407

<212> DNA

<213> Streptococcus pyogenes murD

<400> 1

ggatagtgtg	agcttagatg	aaagtgataa	gtaattttca	aaacaaaaaa	atattaatat	60
tgggggttagc	caaatcgggc	gaagcagcag	caaaattatt	gaccaaactt	ggtgcttttag	120
tgactgttaa	tgatagtaaa	ccatttgacc	aaaatccagc	ggcacaagcc	ttgttggaag	180
aggggattaa	ggtcatttgt	ggtagccacc	cagtagaatt	attagatgag	gactttgagt	240
acatgggttaa	aaaccctggg	attccttatg	ataatcctat	ggttaaacgc	gcccttgcaa	300
aggaaattcc	catcttgact	gaagtagaat	tggcttattt	cgtatctgaa	gcgcctatta	360
tcgggattac	aggatcaaac	gggaagacaa	ccacaacgac	aatgattgcc	gatgttttga	420
atgctggcgg	gcaatctgca	ctcttatctg	gaaacattgg	ttatcctgct	tcaaaagtgtg	480
ttcaaaaagc	aattgctggg	gatactttgg	tgatggaatt	gtcctctttt	caattagtgg	540
gagtgaatgc	ttttcgccct	catattgctg	tcatcactaa	tttaatgccg	actcatctgg	600
actatcatgg	cagttttgag	gattatgttg	ctgctaaatg	gatgattcaa	gctcagatga	660
cagaatcaga	ctaccttatt	ttaaatgcta	atcaagagat	ttcagcaact	ctagctaaga	720
ccacccaagc	aacagtgatt	cctttttcaa	ctcaaaaagt	ggttgatgga	gcttatctga	780
aggatggaat	actctatttt	aaagaacagg	cgattatagc	tgcaactgac	ttaggtgtcc	840
caggtagcca	caacattgaa	aatgccctag	caactattgc	agttgccaag	ttatctggta	900
ttgctgatga	tattattgcc	cagtgccttt	cacatttttg	aggcggttaa	catcgtttgc	960
aacgggttgg	tcaaatcaaa	gatattacct	tctacaatga	cagtaagtca	accaatattt	1020
tagccactca	aaaagcttta	tcaggttttg	ataacagtcg	cttgattttg	attgctggcg	1080
gtctagatcg	tggcaatgaa	tttgacgatt	tggtgccaga	ccttttagga	cttaagcaga	1140
tgattatttt	gggagaatcc	gcagagcgta	tgaagcgagc	tgctaacaaa	gcagaggtct	1200
cttatcttga	agctagaaat	gtggcagaag	caacagagct	tgcttttaag	ctggcccaaa	1260
caggcgatac	tatcttgctt	agcccagcta	atgctagctg	ggatatgtat	cctaattttg	1320
aggttcgtgg	ggatgaattt	ttggcaacct	ttgattgttt	aagaggagat	gcctaattgcc	1380
taagaagatt	ttatttacag	gtggtgg				1407

<210> 2

<211> 452

<212> PRT

<213> Streptococcus pyogenes MurD

<400> 2

Met	Lys	Val	Ile	Ser	Asn	Phe	Gln	Asn	Lys	Lys	Ile	Leu	Ile	Leu	Gly
1				5				10				15			

WO 99/23201

PCT/US98/23054

Leu Ala Lys Ser Gly Glu Ala Ala Ala Lys Leu Leu Thr Lys Leu Gly
 20 25 30
 Ala Leu Val Thr Val Asn Asp Ser Lys Pro Phe Asp Gln Asn Pro Ala
 35 40 45
 Ala Gln Ala Leu Leu Glu Glu Gly Ile Lys Val Ile Cys Gly Ser His
 50 55 60
 Pro Val Glu Leu Leu Asp Glu Asp Phe Glu Tyr Met Val Lys Asn Pro
 65 70 75 80
 Gly Ile Pro Tyr Asp Asn Pro Met Val Lys Arg Ala Leu Ala Lys Glu
 85 90 95
 Ile Pro Ile Leu Thr Glu Val Glu Leu Ala Tyr Phe Val Ser Glu Ala
 100 105 110
 Pro Ile Ile Gly Ile Thr Gly Ser Asn Gly Lys Thr Thr Thr Thr
 115 120 125
 Met Ile Ala Asp Val Leu Asn Ala Gly Gly Gln Ser Ala Leu Leu Ser
 130 135 140
 Gly Asn Ile Gly Tyr Pro Ala Ser Lys Val Val Gln Lys Ala Ile Ala
 145 150 155 160
 Gly Asp Thr Leu Val Met Glu Leu Ser Ser Phe Gln Leu Val Gly Val
 165 170 175
 Asn Ala Phe Arg Pro His Ile Ala Val Ile Thr Asn Leu Met Pro Thr
 180 185 190
 His Leu Asp Tyr His Gly Ser Phe Glu Asp Tyr Val Ala Ala Lys Trp
 195 200 205
 Met Ile Gln Ala Gln Met Thr Glu Ser Asp Tyr Leu Ile Leu Asn Ala
 210 215 220
 Asn Gln Glu Ile Ser Ala Thr Leu Ala Lys Thr Thr Gln Ala Thr Val
 225 230 235 240
 Ile Pro Phe Ser Thr Gln Lys Val Val Asp Gly Ala Tyr Leu Lys Asp
 245 250 255
 Gly Ile Leu Tyr Phe Lys Glu Gln Ala Ile Ile Ala Ala Thr Asp Leu
 260 265 270
 Gly Val Pro Gly Ser His Asn Ile Glu Asn Ala Leu Ala Thr Ile Ala
 275 280 285
 Val Ala Lys Leu Ser Gly Ile Ala Asp Asp Ile Ile Ala Gln Cys Leu
 290 295 300
 Ser His Phe Gly Gly Val Lys His Arg Leu Gln Arg Val Gly Gln Ile
 305 310 315 320
 Lys Asp Ile Thr Phe Tyr Asn Asp Ser Lys Ser Thr Asn Ile Leu Ala
 325 330 335
 Thr Gln Lys Ala Leu Ser Gly Phe Asp Asn Ser Arg Leu Ile Leu Ile
 340 345 350
 Ala Gly Gly Leu Asp Arg Gly Asn Glu Phe Asp Asp Leu Val Pro Asp
 355 360 365
 Leu Leu Gly Leu Lys Gln Met Ile Ile Leu Gly Glu Ser Ala Glu Arg
 370 375 380
 Met Lys Arg Ala Ala Asn Lys Ala Glu Val Ser Tyr Leu Glu Ala Arg
 385 390 395 400
 Asn Val Ala Glu Ala Thr Glu Leu Ala Phe Lys Leu Ala Gln Thr Gly
 405 410 415
 Asp Thr Ile Leu Leu Ser Pro Ala Asn Ala Ser Trp Asp Met Tyr Pro
 420 425 430
 Asn Phe Glu Val Arg Gly Asp Glu Phe Leu Ala Thr Phe Asp Cys Leu
 435 440 445
 Arg Gly Asp Ala
 450

WO 99/23201

PCT/US98/23054

<210> 3
 <211> 438
 <212> PRT
 <213> Escherichia coli MurD

<400> 3
 Met Ala Asp Tyr Gln Gly Lys Asn Val Val Ile Ile Gly Leu Gly Leu
 1 5 10 15
 Thr Gly Leu Ser Cys Val Asp Phe Phe Leu Ala Arg Gly Val Thr Pro
 20 25 30
 Arg Val Met Asp Thr Arg Met Thr Pro Pro Gly Leu Asp Lys Leu Pro
 35 40 45
 Glu Ala Val Glu Arg His Thr Gly Ser Leu Asn Asp Glu Trp Leu Met
 50 55 60
 Ala Ala Asp Leu Ile Val Ala Ser Pro Gly Ile Ala Leu Ala His Pro
 65 70 75 80
 Ser Leu Ser Ala Ala Ala Asp Ala Gly Ile Glu Ile Val Gly Asp Ile
 85 90 95
 Glu Leu Phe Cys Arg Glu Ala Gln Ala Pro Ile Val Ala Ile Thr Gly
 100 105 110
 Ser Asn Gly Lys Ser Thr Val Thr Thr Leu Val Gly Glu Met Ala Lys
 115 120 125
 Ala Ala Gly Val Asn Val Gly Val Gly Gly Asn Ile Gly Leu Pro Ala
 130 135 140
 Leu Met Leu Leu Asp Asp Glu Cys Glu Leu Tyr Val Leu Glu Leu Ser
 145 150 155 160
 Ser Phe Gln Leu Glu Thr Thr Ser Ser Leu Gln Ala Val Ala Ala Thr
 165 170 175
 Ile Leu Asn Val Thr Glu Asp His Met Asp Arg Tyr Pro Phe Gly Leu
 180 185 190
 Gln Gln Tyr Arg Ala Ala Lys Leu Arg Ile Tyr Glu Asn Ala Lys Val
 195 200 205
 Cys Val Val Asn Ala Asp Asp Ala Leu Thr Met Pro Ile Arg Gly Ala
 210 215 220
 Asp Glu Arg Cys Val Ser Phe Gly Val Asn Met Gly Asp Tyr His Leu
 225 230 235 240
 Asn His Gln Gln Gly Glu Thr Trp Leu Arg Val Lys Gly Glu Lys Val
 245 250 255
 Leu Asn Val Lys Glu Met Lys Leu Ser Gly Gln His Asn Tyr Thr Asn
 260 265 270
 Ala Leu Ala Ala Leu Ala Leu Ala Asp Ala Ala Gly Leu Pro Arg Ala
 275 280 285
 Ser Ser Leu Lys Ala Leu Thr Thr Phe Thr Gly Leu Pro His Arg Phe
 290 295 300
 Glu Val Val Leu Glu His Asn Gly Val Arg Trp Ile Asn Asp Ser Lys
 305 310 315 320
 Ala Thr Asn Val Gly Ser Thr Glu Ala Ala Leu Asn Gly Leu His Val
 325 330 335
 Asp Gly Thr Leu His Leu Leu Leu Gly Gly Asp Gly Lys Ser Ala Asp
 340 345 350
 Phe Ser Pro Leu Ala Arg Tyr Leu Asn Gly Asp Asn Val Arg Leu Tyr
 355 360 365
 Cys Phe Gly Arg Asp Gly Ala Gln Leu Ala Ala Leu Arg Pro Glu Val
 370 375 380

WO 99/23201

PCT/US98/23054

Ala Glu Gln Thr Glu Thr Met Glu Gln Ala Met Arg Leu Leu Ala Pro
 385 390 395 400
 Arg Val Gln Pro Gly Asp Met Val Leu Leu Ser Pro Ala Cys Ala Ser
 405 410 415
 Leu Asp Gln Phe Lys Asn Phe Glu Gln Arg Gly Asn Glu Phe Ala Arg
 420 425 430
 Leu Ala Lys Glu Leu Gly
 435

<210> 4
 <211> 437
 <212> PRT
 <213> Haemophilus influenzae MurD

<400> 4
 Met Asn Ala Tyr Gln Asn Lys Asn Ile Thr Ile Ile Gly Leu Gly Lys
 1 5 10 15
 Thr Gly Leu Ser Cys Val Asp Tyr Leu Leu Ser Gln Gln Ala Asn Ile
 20 25 30
 Arg Val Ile Asp Thr Arg Lys Asn Pro Thr Gly Ile Asp Lys Leu Pro
 35 40 45
 Gln Asn Ile Pro Leu His Thr Gly Ser Leu Asn Gln Glu Trp Leu Leu
 50 55 60
 Glu Ser Asp Met Ile Val Ile Ser Pro Gly Leu Ala Val Lys Thr Pro
 65 70 75 80
 Glu Ile Gln Thr Ala Leu Lys Ala Gly Val Glu Val Ile Gly Asp Ile
 85 90 95
 Glu Leu Phe Cys Arg Ala Ala Thr Lys Pro Ile Val Gly Ile Thr Gly
 100 105 110
 Ser Asn Gly Lys Ser Thr Val Thr Thr Leu Val Tyr Glu Met Ala Lys
 115 120 125
 Ala Ala Gly Val Lys Val Gly Met Gly Gly Asn Ile Gly Ile Pro Ala
 130 135 140
 Leu Ser Leu Leu Asn Glu Asp Cys Glu Leu Tyr Val Leu Glu Leu Ser
 145 150 155 160
 Ser Phe Gln Leu Glu Thr Thr Tyr Ser Leu Lys Ala Ala Ala Thr
 165 170 175
 Val Leu Asn Val Thr Glu Asp His Met Asp Arg Tyr Met Asp Leu Glu
 180 185 190
 Asp Tyr Arg Gln Ala Lys Leu Arg Ile Tyr His Asn Ala Lys Val Gly
 195 200 205
 Val Leu Asn Asn Glu Asp Arg Leu Thr Phe Gly Glu Asn Glu Asn Gln
 210 215 220
 Ala Lys His Thr Val Ser Phe Ala Glu Asn Ser Ala Asp Tyr Trp Leu
 225 230 235 240
 Lys Thr Glu Asn Gly Lys Gln Tyr Leu Met Val Lys Asp Glu Val Ile
 245 250 255
 Leu Pro Cys Glu Glu Ala Thr Leu Val Gly Arg His Asn Tyr Met Asn
 260 265 270
 Ile Leu Ala Ala Thr Ala Leu Ala Gln Ala Ile Gly Ile Asn Leu Asp
 275 280 285
 Ser Ile Arg Thr Ala Leu Arg His Phe Lys Gly Leu Asp His Arg Phe
 290 295 300
 Gln Leu Val His Gln Ala Asn Gly Ile Arg Trp Ile Asn Asp Ser Lys
 305 310 315 320

[illegible]

```
<210> 5
<211> 451
<212> PRT
<213> Bacillus subtilis MurD
```

<400> 5															
Val 1	Glu	Asn	Asp	Gln	Phe	Leu	Gln	Lys	Gln	His	Phe	Leu	Ile	Leu	Gly
			5						10					15	
Leu	Ala	Lys	Ser	Gly	Tyr	Ala	Ala	Ala	Ser	Ile	Leu	His	Glu	Lys	Gly
			20					25					30		
Ile	Tyr	Val	Ala	Val	Asn	Asp	Gln	Lys	Pro	Phe	Glu	Glu	Asn	Glu	Pro
		35					40					45			
Ala	Gln	Lys	Leu	Ser	Glu	Lys	Gly	Ile	Glu	Val	Val	Cys	Gly	Glu	His
	50					55					60				
Pro	Val	Ser	Leu	Phe	Asp	Gln	His	Gln	Ile	Thr	Ile	Leu	Ile	Lys	Asn
65					70					75					80
Pro	Gly	Ile	Pro	Tyr	Glu	Asn	Ile	Met	Val	Gln	Glu	Ala	Glu	Lys	Arg
			85					90						95	
Gly	Ile	Pro	Val	Trp	Thr	Glu	Ile	Glu	Leu	Ala	Tyr	Tyr	Leu	Thr	Ser
			100					105					110		
Ala	Lys	Phe	Ile	Gly	Ile	Thr	Gly	Ser	Asn	Gly	Lys	Thr	Thr	Thr	Thr
		115					120					125			
Thr	Leu	Ile	Tyr	Glu	Met	Leu	Lys	Ala	Asp	Ser	Gln	Lys	Ala	Leu	Ile
	130					135					140				
Ala	Gly	Asn	Ile	Gly	Thr	Val	Ala	Ser	Glu	Val	Ala	Tyr	His	Ala	Asp
145					150					155					160
Gly	Asp	Glu	Trp	Ile	Val	Thr	Glu	Leu	Ser	Ser	Phe	Gln	Leu	Met	Gly
			165					170						175	
Thr	His	Ala	Phe	Arg	Pro	Glu	Ile	Ser	Leu	Ile	Leu	Asn	Val	Phe	Asp
			180					185					190		
Ala	His	Leu	Asp	Tyr	His	His	Thr	Arg	Glu	Asn	Tyr	Glu	Lys	Ala	Lys
		195					200					205			
Gln	Lys	Val	Tyr	Leu	His	Gln	Thr	Ala	Ser	Asp	Lys	Ala	Ile	Val	Asn
	210					215					220				
Gln	Asp	Asp	Glu	Thr	Val	Val	Arg	Leu	Ala	Glu	Ala	Gly	Lys	Ala	Glu
225					230					235					240
Ile	Val	Pro	Phe	Ser	Val	Ser	Lys	Thr	Leu	Glu	Gln	Gly	Ala	Tyr	Val
			245						250					255	

WO 99/23201

PCT/US98/23054

Lys Asp Ser Met Ile Met Phe Asn Gly Glu Ala Ile Leu Pro Leu Glu
 260 265 270
 Glu Val Val Leu Pro Gly Ala His Asn Leu Glu Asn Ile Leu Ala Ala
 275 280 285
 Ile Ala Val Val Lys Thr Ala Gly Ala Ser Asn Glu Ala Val Lys Lys
 290 295 300
 Val Leu Thr Ser Phe Thr Gly Val Lys His Arg Leu Gln Tyr Val Thr
 305 310 315 320
 Thr Val Asn Gly Arg Lys Phe Tyr Asn Asp Ser Lys Ala Thr Asn Ile
 325 330 335
 Leu Ala Thr Ser Lys Ala Leu Ser Ala Phe Asp Lys Pro Val Ile Leu
 340 345 350
 Leu Ala Gly Gly Leu Asp Arg Gly Asn Gly Phe Asp Asp Leu Lys Pro
 355 360 365
 Tyr Met Lys His Val Lys Ala Val Leu Thr Phe Gly Gln Thr Ala Pro
 370 375 380
 Lys Leu Glu Lys Leu Gly Asn Glu Leu Gly Ile Gln His Val Lys Arg
 385 390 395 400
 Val Asp Asn Val Glu Gln Ala Val Ser Ala Ala Phe Ala Leu Ser Asn
 405 410 415
 Glu Gly Asp Val Ile Leu Leu Ser Pro Ala Cys Ala Ser Trp Asp Gln
 420 425 430
 Phe Lys Thr Phe Glu Glu Arg Gly Asp Met Phe Ile Asp Ala Val His
 435 440 445
 Met Leu Lys
 450

<210> 6

<211> 449

<212> PRT

<213> Staphylococcus aureus MurD

<400> 6

Met Leu Asn Tyr Thr Gly Leu Glu Asn Lys Asn Val Leu Val Val Gly
 1 5 10 15
 Leu Ala Lys Ser Gly Tyr Glu Ala Ala Lys Leu Leu Ser Lys Leu Gly
 20 25 30
 Ala Asn Val Thr Val Asn Asp Gly Lys Asp Leu Ser Gln Asp Ala His
 35 40 45
 Ala Lys Asp Leu Glu Ser Met Gly Ile Ser Val Val Ser Gly Ser His
 50 55 60
 Pro Leu Thr Leu Leu Asp Asn Asn Pro Ile Ile Val Lys Asn Pro Gly
 65 70 75 80
 Ile Pro Tyr Thr Val Ser Ile Ile Asp Glu Ala Val Lys Arg Gly Leu
 85 90 95
 Lys Ile Leu Thr Glu Val Glu Leu Ser Tyr Leu Ile Ser Glu Ala Pro
 100 105 110
 Ile Ile Ala Val Thr Gly Thr Asn Gly Lys Thr Thr Val Thr Ser Leu
 115 120 125
 Ile Gly Asp Met Phe Lys Lys Ser Arg Leu Thr Gly Arg Leu Ser Gly
 130 135 140
 Asn Ile Gly Tyr Val Ala Ser Lys Val Ala Gln Glu Val Lys Pro Thr
 145 150 155 160
 Asp Tyr Leu Val Thr Glu Leu Ser Ser Phe Gln Leu Leu Gly Ile Glu
 165 170 175

WO 99/23201

PCT/US98/23054

Lys	Tyr	Lys	Pro	His	Ile	Ala	Ile	Ile	Thr	Asn	Ile	Tyr	Ser	Ala	His
			180					185					190		
Leu	Asp	Tyr	His	Glu	Asn	Leu	Glu	Asn	Tyr	Gln	Asn	Ala	Lys	Lys	Gln
		195					200					205			
Ile	Tyr	Lys	Asn	Gln	Thr	Glu	Glu	Asp	Tyr	Leu	Ile	Cys	Asn	Tyr	His
	210					215					220				
Gln	Arg	Gln	Val	Ile	Glu	Ser	Glu	Glu	Leu	Lys	Ala	Lys	Thr	Leu	Tyr
225					230					235					240
Phe	Ser	Thr	Gln	Gln	Glu	Val	Asp	Gly	Ile	Tyr	Ile	Lys	Asp	Gly	Phe
			245						250					255	
Ile	Val	Tyr	Lys	Gly	Val	Arg	Ile	Ile	Asn	Thr	Glu	Asp	Leu	Val	Leu
			260					265					270		
Pro	Gly	Glu	His	Asn	Leu	Glu	Asn	Ile	Leu	Ala	Ala	Val	Leu	Ala	Cys
		275					280					285			
Ile	Leu	Ala	Gly	Val	Pro	Ile	Lys	Ala	Ile	Ile	Asp	Ser	Leu	Thr	Thr
	290					295					300				
Phe	Ser	Gly	Ile	Glu	His	Arg	Leu	Gln	Tyr	Val	Gly	Thr	Asn	Arg	Thr
305					310					315					320
Asn	Lys	Tyr	Tyr	Asn	Asp	Ser	Lys	Ala	Thr	Asn	Thr	Leu	Ala	Thr	Gln
				325					330					335	
Phe	Ala	Leu	Asn	Ser	Phe	Asn	Gln	Pro	Ile	Ile	Trp	Leu	Cys	Gly	Gly
			340					345					350		
Leu	Asp	Arg	Gly	Asn	Glu	Phe	Asp	Glu	Leu	Ile	Pro	Tyr	Met	Glu	Asn
	355						360					365			
Val	Arg	Ala	Met	Val	Val	Phe	Gly	Gln	Thr	Lys	Ala	Lys	Phe	Ala	Lys
	370					375					380				
Leu	Gly	Asn	Ser	Gln	Gly	Lys	Ser	Val	Ile	Glu	Ala	Asn	Asn	Val	Glu
385					390					395					400
Asp	Ala	Val	Asp	Lys	Val	Gln	Asp	Ile	Ile	Glu	Pro	Asn	Asp	Val	Val
				405					410					415	
Leu	Leu	Ser	Pro	Ala	Cys	Ala	Ser	Trp	Asp	Gln	Tyr	Ser	Thr	Phe	Glu
			420					425					430		
Glu	Arg	Gly	Glu	Lys	Phe	Ile	Glu	Arg	Phe	Arg	Ala	His	Leu	Pro	Ser
		435					440					445			
Tyr															

WHAT IS CLAIMED:

1. An isolated polynucleotide selected from the group consisting of:
 - (a) a polynucleotide encoding a polypeptide having an amino acid sequence of SEQ ID NO: 2.
 - (b) a polynucleotide which is complementary to the polynucleotide of (a),
 - (c) a polynucleotide representing a naturally occurring mutant or polymorphic form of (a), and
 - (d) a polynucleotide comprising at least 25 nucleotides of the polynucleotide of (a), (b) or (c), said 25 nucleotides being highly specific for *murD* gene of *Streptococcus pyogenes*.
2. The polynucleotide of claim 1 wherein the polynucleotide comprises nucleotides selected from the group consisting of natural, non-natural and modified nucleotides.
3. The polynucleotide of claim 1 wherein the internucleotide linkages are selected from the group consisting of natural and non-natural linkages.
4. The polynucleotide of claim 1 comprising the nucleotide sequence of SEQ ID NO:1.
5. A polynucleotide that is an expression vector comprising a polynucleotide of claim 1.
6. A host cell comprising the expression vector of claim 5.
7. A process for expressing a MurD protein of *Streptococcus pyogenes* in a recombinant host cell, comprising:
 - (a) transfecting a suitable host cell with an expression vector of claim 5; and,
 - (b) culturing the host cell of step (a) in conditions under which allow expression of said the MurD protein from said expression vector.

8. A substantially purified polypeptide having an amino acid sequence selected from the group consisting of
- 5 ID NO:2,
- (a) a polypeptide having an amino acid sequence of SEQ
- (b) a polypeptide representing a naturally occurring mutant or polymorphic form of (a).
9. A method of determining whether candidate
- 10 compounds or molecules are inhibitors of a *Streptococcus pyogenes* MurD polypeptide of claim 8 comprising:
- (a) providing a cell according to claim 6,
- (b) contacting said cell with the candidate under conditions sufficient to permit the interaction of the candidate with the
- 15 protein of the candidate, and
- (c) determining whether the candidate is an inhibitor by ascertaining the relative activity of the murd polypeptide in the presence of the candidate.

1/5

10 30 50
ATGAAAGTGATAAGTAATTTTCAAAACAAAAAATATTAATATTGGGGTTAGCCAAATCG
MetLysValIleSerAsnPheGlnAsnLysLysIleLeuIleLeuGlyLeuAlaLysSer

70 90 110
GGCGAAGCAGCAGCAAAATTATTGACTAACTTGGTGCTTTAGTGACTGTTAATGATAGT
GlyGluAlaAlaAlaLysLeuLeuThrLysLeuGlyAlaLeuValThrValAsnAspSer

130 150 170
AAACCATTTGACCAAAATCCAGCGGCACAAGCCTTGTTGGAAGAGGGGATTAAGGTCATT
LysProPheAspGlnAsnProAlaAlaGlnAlaLeuLeuGluGluGlyIleLysValIle

190 210 230
TGTGGTAGCCACCCAGTAGAATTATTAGATGAGGACTTTGAGTACATGGTTAAAAACCT
CysGlySerHisProValGluLeuLeuAspGluAspPheGluTyrMetValLysAsnPro

250 270 290
GGGATTCCTTATGATAATCCTATGGTTAAACGCGCCCTTGCAAAGGAAATTCCCATCTTG
GlyIleProTyrAspAsnProMetValLysArgAlaLeuAlaLysGluIleProIleLeu

310 330 350
ACTGAAGTAGAATTGGCTTATTTTGTATCTGAAGCGCCTATTATCGGGATTACAGGATCA
ThrGluValGluLeuAlaTyrPheValSerGluAlaProIleIleGlyIleThrGlySer

370 390 410
AACGGGAAGACAACCACAACGACAATGATTGCCGATGTTTTGAATGCTGGCGGGCAATCT
AsnGlyLysThrThrThrThrMetIleAlaAspValLeuAsnAlaGlyGlyGlnSer

430 450 470
GCACTCTTATCTGGAAACATTGGTTATCCTGCTTCAAAAGTTGTTCAAAAAGCAATTGCT
AlaLeuLeuSerGlyAsnIleGlyTyrProAlaSerLysValValGlnLysAlaIleAla

490 510 530
GGTGATACTTTGGTGATGGAATTGTCCTCTTTTCAATTAGTGGGAGTGAATGCTTTTCGC
GlyAspThrLeuValMetGluLeuSerSerPheGlnLeuValGlyValAsnAlaPheArg

550 570 590
CCTCATATTGCTGTCATCACTAATTTAATGCCGACTCATCTGGACTATCATGGCAGTTTT
ProHisIleAlaValIleThrAsnLeuMetProThrHisLeuAspTyrHisGlySerPhe

610 630 650
GAGGATTATGTTGCTGCTAAATGGATGATTCAAGCTCAGATGACAGAATCAGACTACCTT
GluAspTyrValAlaAlaLysTrpMetIleGlnAlaGlnMetThrGluSerAspTyrLeu

670 690 710
ATTTTAAATGCTAATCAAGAGATTTTCAAGCACTCTAGCTAAGACCAACCAAGCAACAGTG
IleLeuAsnAlaAsnGlnGluIleSerAlaThrLeuAlaLysThrThrGlnAlaThrVal

730 750 770
ATTCCTTTTTCAACTCAAAAAGTGGTTGATGGAGCTTATCTGAAGGATGGAATACTCTAT
IleProPheSerThrGlnLysValValAspGlyAlaTyrLeuLysAspGlyIleLeuTyr

FIG. 1A

2/5

790 810 830
TTTAAAGAACAAGCGATTATAGCTGCAACTGACTTAGGTGTCCCAGGTAGCCACAACATT
PheLysGluGlnAlaIleIleAlaAlaThrAspLeuGlyValProGlySerHisAsnIle

850 870 890
GAAAATGCCCTAGCAACTATTGCAGTTGCCAAGTTATCTGGTATTGCTGATGATATTATT
GluAsnAlaLeuAlaThrIleAlaValAlaLysLeuSerGlyIleAlaAspAspIleIle

910 930 950
GCCCAGTGCCTTTCACATTTTGGAGGCGTTAAACATCGTTTGCAACGGGTGGTCAAATC
AlaGlnCysLeuSerHisPheGlyGlyValLysHisArgLeuGlnArgValGlyGlnIle

970 990 1010
AAAGATATTACCTTCTACAATGACAGTAAGTCAACCAATATTTTAGCCACTCAAAAAGCT
LysAspIleThrPheTyrAsnAspSerLysSerThrAsnIleLeuAlaThrGlnLysAla

1030 1050 1070
TTATCAGGTTTTGATAACAGTCGCTTGATTTTGATTGCTGGCGGTCTAGATCGTGGCAAT
LeuSerGlyPheAspAsnSerArgLeuIleLeuIleAlaGlyGlyLeuAspArgGlyAsn

1090 1110 1130
GAATTTGACGATTTGGTGCCAGACCTTTTAGGACTTAAGCAGATGATTATTTTAGGAGAA
GluPheAspAspLeuValProAspLeuLeuGlyLeuLysGlnMetIleIleLeuGlyGlu

1150 1170 1190
TCCGCAGAGCGTATGAAGCGAGCTGCTAACAAAGCAGAGGTCTTATCTTGAAGCTAGA
SerAlaGluArgMetLysArgAlaAlaAsnLysAlaGluValSerTyrLeuGluAlaArg

1210 1230 1250
AATGTGGCAGAAGCAACAGAGCTTGCTTTTAAGCTGGCCCAAACAGGCGATACTATCTTG
AsnValAlaGluAlaThrGluLeuAlaPheLysLeuAlaGlnThrGlyAspThrIleLeu

1270 1290 1310
CTTAGCCCAGCTAATGCTAGCTGGGATATGTATCCTAATTTTGAGGTTCGTGGGGATGAA
LeuSerProAlaAsnAlaSerTrpAspMetTyrProAsnPheGluValArgGlyAspGlu

1330 1350
TTTTTGGCAACCTTTGATTGTTTAAGAGGAGATGCC
PheLeuAlaThrPheAspCysLeuArgGlyAspAla

FIG. 1B

FIG. 2A

Eco}	--	MADY	QGN	V	I	IGL	GL	T	G	L	SC	VD	F	F	L	A	R	G	V	T	P	R	V	M	D	T	R	--	M	T	P	P	C	L	44																	
Hin}	--	MNAY	QNK	N	I	I	GL	G	K	T	L	SC	VD	Y	L	S	Q	Q	A	N	I	R	V	I	D	T	R	--	K	N	P	T	G	I	44																	
Bsu}	V	E	N	D	Q	F	L	Q	K	Q	H	F	L	I	L	GL	A	K	S	G	Y	A	A	S	I	L	L	H	E	K	G	I	Y	V	A	50																
Spy}	M	K	V	I	S	N	F	Q	N	K	I	L	I	L	GL	A	K	S	G	E	A	A	K	L	L	L	T	K	L	G	A	L	V	T	V	N	D	50														
Sau}	M	L	N	Y	T	G	L	E	N	K	N	V	L	V	V	GL	A	K	S	G	Y	E	A	A	K	L	L	S	K	L	G	A	N	V	T	V	N	D	50													
Eco}	D	K	L	P	E	A	V	E	R	H	T	G	S	L	N	D	E	W	L	M	A	A	--	D	L	I	V	A	S	P	G	I	A	L	A	H	P	S	L	S	A	A	D	A	G	I	E	I	92			
Hin}	D	K	L	P	Q	N	I	P	L	H	T	G	S	L	N	Q	E	W	L	L	E	S	--	D	M	I	V	I	S	P	G	L	A	V	K	T	P	E	I	Q	T	A	L	K	A	G	V	E	V	92		
Bsu}	K	L	S	E	K	G	I	E	V	V	C	G	E	H	P	V	S	L	F	D	Q	H	Q	I	T	I	L	I	K	N	P	G	I	P	Y	E	N	I	M	V	Q	E	A	L	E	K	R	G	I	P	V	100
Spy}	A	L	L	E	E	G	I	K	V	I	C	G	S	H	P	V	E	L	L	D	E	N	F	--	E	Y	M	V	K	N	P	G	I	P	Y	D	N	P	M	V	K	R	A	L	A	K	E	I	P	I	99	
Sau}	D	L	E	S	M	G	I	S	V	V	S	C	S	H	P	L	T	L	L	D	N	N	--	P	I	I	V	K	N	P	G	I	P	Y	T	V	S	I	I	D	E	A	V	K	R	G	L	K	I	98		
Eco}	V	G	D	I	E	L	F	C	R	E	A	Q	A	P	I	V	A	I	T	G	S	N	G	K	S	T	V	T	T	L	V	G	E	M	A	K	A	A	G	V	N	V	G	V	G	N	I	G	L	142		
Hin}	I	G	D	I	E	L	F	C	R	A	A	T	K	P	I	V	G	I	T	G	S	N	G	K	S	T	V	T	T	L	V	Y	E	M	A	K	A	A	G	V	K	V	G	M	G	N	I	G	I	142		
Bsu}	W	T	E	I	E	L	A	Y	Y	L	T	S	A	K	F	I	G	I	T	G	S	N	G	K	T	T	T	T	T	L	I	Y	E	M	L	K	A	D	S	Q	K	A	L	I	A	G	N	I	G	T	150	
Spy}	L	T	E	V	E	L	A	Y	F	V	S	E	A	P	I	I	G	I	T	G	S	N	G	K	T	T	T	T	T	M	I	A	D	V	L	N	A	G	Q	S	A	L	L	S	G	N	I	G	Y	149		
Sau}	L	T	E	V	E	L	S	Y	L	I	S	E	A	P	I	I	A	V	T	G	T	N	G	K	T	T	V	T	S	L	I	G	D	M	F	K	K	S	R	L	T	G	R	L	S	G	N	I	G	Y	148	
Eco}	P	A	L	M	L	D	D	E	--	C	E	L	Y	V	L	E	L	S	S	F	Q	L	E	T	T	S	S	L	Q	A	V	A	A	T	I	L	N	V	T	E	D	H	M	D	R	Y	P	F	190			
Hin}	P	A	L	S	L	N	E	D	--	C	E	L	Y	V	L	E	L	S	S	F	Q	L	E	T	T	Y	S	L	K	A	A	A	T	V	L	N	V	T	E	D	H	M	D	R	Y	--	M	189				
Bsu}	V	A	S	E	V	A	Y	H	A	D	G	D	E	W	I	V	T	E	L	S	S	F	Q	L	M	G	T	H	A	F	R	P	E	I	S	L	I	L	N	V	F	D	A	H	L	D	Y	H	--	H	199	
Spy}	P	A	S	K	V	V	Q	K	A	I	A	G	D	T	L	V	M	E	L	S	S	F	Q	L	V	G	V	N	A	F	R	P	H	I	A	V	I	I	N	L	M	P	T	H	L	D	Y	H	--	G	198	
Sau}	V	A	S	K	V	A	Q	E	V	K	P	T	D	Y	L	V	T	E	L	S	S	F	Q	L	L	G	I	E	K	Y	K	P	H	I	A	I	I	T	N	I	Y	S	A	H	L	D	Y	H	--	E	197	
Eco}	G	L	Q	Q	Y	R	A	K	L	R	I	Y	E	N	--	A	K	V	C	V	V	N	A	D	D	A	L	T	M	P	I	R	--	G	A	D	E	R	C	V	S	F	G	V	N	M	G	236				
Hin}	D	L	E	D	Y	R	Q	A	K	L	R	I	Y	H	N	--	A	K	V	G	V	L	N	E	D	R	L	T	F	G	E	N	E	N	Q	A	K	H	T	V	S	F	A	E	N	S	A	236				
Bsu}	T	R	E	N	Y	E	K	A	K	Q	K	V	Y	L	H	Q	T	A	S	D	K	A	I	V	N	Q	D	D	E	T	V	V	R	L	A	E	A	G	K	A	E	I	V	P	F	S	V	S	K	T	249	
Spy}	S	F	E	D	Y	V	A	A	K	W	M	I	Q	A	Q	M	T	E	S	D	Y	L	I	L	N	A	N	Q	E	I	S	A	T	L	A	K	T	T	K	A	T	V	I	P	F	S	T	Q	K	V	248	
Sau}	N	L	E	N	Y	Q	N	A	K	K	I	Y	K	N	Q	T	E	E	D	Y	L	I	C	N	Y	H	Q	R	Q	V	I	E	--	S	E	E	L	K	A	K	T	L	Y	F	S	T	Q	Q	E	246		

Eco	D	Y	H	L	N	H	Q	Q	G	E	T	W	L	R	V	K	G	E	K	V	L	N	V	K	E	M	K	L	S	G	Q	H	N	Y	T	N	A	L	A	L	A	A	L	A	L	A	A	L	A	D	A	A	G	L	P	286
Hin	D	Y	W	L	K	T	E	N	G	K	Q	Y	L	M	V	K	D	E	V	I	L	P	C	E	E	A	T	L	V	G	R	H	N	Y	M	N	I	L	A	A	I	L	A	A	I	L	A	Q	A	I	G	I	N	286		
Bsu	L	E	Q	G	-	A	Y	V	K	D	S	M	I	M	F	N	G	E	A	I	L	P	L	E	E	V	V	P	G	A	H	N	L	E	N	I	L	A	A	I	L	A	A	V	V	K	T	A	G	A	S	298				
Spy	V	D	G	-	-	A	Y	L	K	D	G	I	L	Y	F	K	E	Q	A	I	I	A	T	D	L	G	V	P	G	S	H	N	I	E	N	A	L	A	T	I	A	V	A	K	L	S	G	I	A	296						
Sau	V	D	G	-	-	I	Y	I	K	D	G	F	I	V	Y	K	G	V	R	I	I	N	T	E	D	L	V	L	P	G	E	H	N	L	E	N	I	L	A	A	V	L	A	C	I	L	A	G	V	P	294					

Eco	R	A	S	S	L	K	A	L	T	T	F	T	G	L	P	H	R	F	F	V	V	L	E	H	N	G	V	R	W	I	N	D	S	K	A	T	N	V	G	S	T	E	A	A	L	N	G	L	H	V	
Hin	L	D	S	I	R	T	A	L	R	H	F	Q	L	V	D	H	R	F	Q	L	V	H	Q	A	N	G	I	R	W	I	N	D	S	K	A	T	N	V	G	S	T	V	A	A	L	N	G	L	Y	I	
Bsu	N	E	A	V	K	K	V	L	T	S	F	T	G	V	K	H	R	L	Q	R	V	T	I	K	D	I	R	F	F	Y	N	D	S	K	A	T	N	I	L	A	T	S	K	A	L	S	A	F	D	-	
Spy	D	D	I	I	A	Q	C	L	S	H	F	G	V	K	H	R	L	Q	R	V	V	G	I	N	R	T	I	K	F	F	Y	N	D	S	K	A	T	N	I	L	A	T	S	K	A	L	S	G	F	D	N
Sau	I	K	A	I	I	D	S	L	T	T	F	G	I	E	H	R	L	Q	R	V	V	G	T	N	R	T	I	K	F	F	Y	N	D	S	K	A	T	N	I	L	A	T	S	K	A	L	N	S	F	N	-

Eco	D	G	T	L	H	L	L	L	G	G	D	G	K	S	A	D	F	S	P	L	A	R	Y	L	N	G	D	N	V	R	L	Y	C	F	G	R	D	G	A	Q	L	A	L	R	P	E	V	-	-	384	
Hin	E	G	K	L	H	L	L	L	L	G	G	D	G	K	G	A	D	F	S	E	L	A	E	L	I	N	Q	P	H	I	C	Y	C	F	G	R	D	G	A	L	L	A	K	F	S	Q	-	-	383		
Bsu	-	K	P	V	I	L	L	L	A	G	G	L	D	R	G	N	G	F	D	L	K	P	Y	M	K	H	-	V	K	A	V	L	T	F	G	Q	T	A	P	K	L	E	K	L	G	N	E	L	G	I	395
Spy	-	S	R	L	I	L	I	A	G	G	L	D	R	G	N	E	F	D	L	V	P	D	L	L	G	-	L	K	Q	M	I	I	L	G	E	S	A	E	R	M	K	R	A	N	K	A	E	V	394		
Sau	-	Q	P	I	I	W	L	C	G	G	L	D	R	G	N	E	F	D	E	I	P	Y	M	E	N	-	V	R	A	M	V	V	F	G	Q	T	K	A	K	F	A	K	L	G	N	S	Q	G	K	391	

Eco}	-	-	A	E	Q	T	E	T	M	E	Q	A	M	R	L	L	A	P	R	V	Q	P	G	D	M	V	L	L	S	P	A	C	A	S	L	D	Q	F	K	N	F	E	F	A	R	432
Hin}	-	-	S	Y	L	F	D	T	M	E	Q	A	I	E	F	L	R	P	T	L	Q	S	G	D	M	V	L	L	S	P	A	C	A	S	L	D	Q	F	A	S	F	E	F	T	H	431
Bsu}	Q	H	V	K	R	V	D	N	V	E	Q	A	V	S	A	A	F	A	L	S	N	E	G	D	V	I	L	L	S	P	A	C	A	S	W	D	Q	F	K	T	F	I	D	445		
Spy}	S	-	Y	L	E	A	R	N	V	A	E	A	T	E	L	A	F	K	L	A	Q	T	G	D	T	I	L	L	S	P	A	N	A	S	W	D	M	Y	P	N	F	L	A	443		
Sou}	S	-	V	I	E	A	N	N	V	E	D	A	V	D	K	V	Q	D	I	I	E	P	N	D	V	V	L	L	S	P	A	C	A	S	W	D	Q	Y	S	T	I	E	440			

Eco}	L	A	K	E	L	G	-	-	-	438
Hin}	L	A	Q	C	L	T	-	-	-	437
Bsu}	A	V	H	M	L	K	-	-	-	451
Spy}	T	F	D	C	L	R	G	D	A	452
Sau}	R	F	R	A	H	L	P	S	Y	449

FIG. 2B

	Region I	Region II	Region IV
Spy	GSNGKTT	VMELSSFQL...N...HLDYH...AK	G.HN...N...AT.A...G...HRLQ

FIG.3