



US 20050101504A1

(19) **United States**

(12) **Patent Application Publication**  
**Estell**

(10) **Pub. No.: US 2005/0101504 A1**

(43) **Pub. Date: May 12, 2005**

(54) **PROTEASES FROM GRAM POSITIVE ORGANISMS**

**Publication Classification**

(76) Inventor: **David A. Estell, San Mateo, CA (US)**

Correspondence Address:  
**Genencor International, Inc.**  
**925 Page Mill Road**  
**Palo Alto, CA 94034-1013 (US)**

(51) **Int. Cl.<sup>7</sup>** ..... **C11D 3/386**; C12N 9/50;  
C12N 9/54; C07H 21/04;  
C12N 1/21; C12N 15/74  
(52) **U.S. Cl.** ..... **510/320**; 435/221; 435/69.1;  
435/320.1; 435/252.31; 536/23.2

(21) Appl. No.: **11/013,992**

(57) **ABSTRACT**

(22) Filed: **Dec. 16, 2004**

**Related U.S. Application Data**

(62) Division of application No. 10/460,718, filed on Jun. 12, 2003, now Pat. No. 6,872,807, which is a division of application No. 09/554,998, filed on May 23, 2000, now Pat. No. 6,599,731, filed as 371 of international application No. PCT/US98/27020, filed on Dec. 17, 1998.

The present invention relates to the identification of a novel metalloproteases in gram positive microorganisms. The present invention provides the nucleic acid and amino acid sequences for the metalloprotease. The present invention also provides host cells having a mutation or deletion of part or all of the gene encoding the metalloprotease. The present invention provides host cells which further comprises a nucleic acid encoding desired heterologous proteins such as enzymes. The present invention also provides cleaning compositions, animal feeds and compositions used to treat a textile that include the metalloprotease of the present invention.

(30) **Foreign Application Priority Data**

Dec. 30, 1997 (GB) ..... 9727464.1





GTCTCCTTATGTGAGGAGGAAGGCAGCCGTTTCCCAATGACCTATTGGGGGTCGGGGAAT 540  
 CAGAGGAATACACTCCTCCTCCGTCGGCAAAAGGTTACTGGATAACCCCCAGCCCTTAA

Val Ser Leu Cys Glu Glu Glu Ser Arg Phe Pro Met Thr Tyr Trp Gly Ser Gly Asn  
 \_\_\_\_\_yurH\_\_\_\_\_

ATGACGGGTGTTTTTCAGAACAAAGATGCCAAAAGAGCCCAAGAGACGAGTCGGGAGTCICC 600  
 TACTGCCCCACAAAAAAGTCTTGTTCTACGTTTTCTCGGTTCTCTGCTCAGCCCTCAGAGG

Met Thr Gly Val Phe Ser Glu Gln Asp Ala Lys Glu Pro Arg Asp Glu Ser Gly Val Ser  
 \_\_\_\_\_yurH\_\_\_\_\_

CTCCAAACAGCGGATGCATGAGAGCGGTTTTGGCAAAGGTGTTTCAGTCAGCATAATAGG 660  
 GAGGTTGTGCTACGTACTCTCGCCAAAACCGTTTCCACACAAAAGTCAGTCGTATAITCC

Leu Gln Thr Ala Met His Glu Ser Gly Phe Gly Lys Gly Val Phe Gln Ser Ala Tyr Arg  
 \_\_\_\_\_yurH\_\_\_\_\_

ACAGATATCAGCGGTTTGTGAGCTGCATATTGAACAGGGGAAGACGTTGGAAATGTCA 720  
 TGTCTATAGTCGGCAAAACAACTCGACGTATAACTTGTCCTCCCTTCTGCAACCTTTACAGT

Thr Asp Ile Ser Ala Phe Val Glu Leu His Ile Glu Gln Gly Lys Thr Leu Glu Met Ser  
 \_\_\_\_\_yurH\_\_\_\_\_

FIG.- 1C

GGCCGGGATCTCGGCATCGTGACAAGTATTGGGGGCGAGAGACGATATCTCGTCACGCCTC  
 780  
 CCGGCCCTAGAGCCGTAGCACTGTTTCATAACGCCCCCGTCTCTGCTATAGAGCAGTGCAG

Gly Arg Asp Leu Gly Ile Val Thr Ser Ile Ala Gly Gln Arg Arg Tyr Leu Val Thr Leu  
 \_\_\_\_\_yurH\_\_\_\_\_

GAAGGAGAAATGCAATCACGCAGGAACCCACCTCCATGAAATGGCGCAAGGACCCCGCTCGCA  
 840  
 CTTCCCTTACGTTAGTGGTCCCTGGTGGAGGTACTTTACCCGGTTCCTGGGGGAGCGGT

Glu Gly Glu Cys Asn His Ala Gly Thr Thr Ser Met Lys Trp Arg Lys Asp Pro Leu Ala  
 \_\_\_\_\_yurH\_\_\_\_\_

GCCAGCAGCCGTATCATGAGCTGCTGCTGGGTCGGATGAGCTGCCGGATGAGCTC  
 900  
 CGGTGTCGGCATAGTAGTACTCGACGACGACGCCAGCCAGCTACTCGACGGCCACTCGAG

Ala Ser Ser Arg Ile Ile His Glu Leu Leu Leu Leu Arg Ser Asp Glu Leu Pro Asp Glu Leu  
 \_\_\_\_\_yurH\_\_\_\_\_

CGTCTGACATGGGAAAAAATAACGGCAGAGCCCAATGTAGCCAAATGTCATACCGGGCCCGC  
 960  
 GCAGACTGTACGCCCTTTTATTGCCGCTCGGGTTACATCGGTTACAGTATGGCCCGGGCG

Arg Leu Thr Cys Gly Lys Ile Thr Ala Glu Pro Asn Val Ala Asn Val Ile Pro Gly Arg  
 \_\_\_\_\_yurH\_\_\_\_\_

FIG.-1D

GTCCAGTTTTCAATCGATATTCGCCCATCAGCATCAGCATGTGCTGGAACAGTTTCATCAA  
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
 CAGGTCAAAAAGTTAGCTATAAGCGGTAGTCGIAGTCGTACACGACCTTGTCAAAAGTAGTT  
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

1020

Val Gln Phe Ser Ile Asp Ile Arg His Gln His Gln His Val Leu Glu Gln Phe His Gln  
 -----yurH-----

GACATGGTTGCTTTGATCAACCGGCATTTGCCCTACAAAAAGGAATTCGGCIGTGATTGAT  
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
 CTGTACCAACGAAACTAGTTGCCGTAACGGGATGTTTTCCCTTAAGCGCGACACTAACTA  
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

1080

Asp Met Val Ala Leu Ile Asn Gly Ile Cys Leu Gln Lys Gly Ile Arg Ala Val Ile Asp  
 -----yurH-----

GAATATATGCGGATAGAGCCITGTGCCGATGGACGAAAGGCTGAAGGCTGCGGCTTTTGAA  
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
 CTTATATACGCCCTATCTCGGACACGGCTACCTGCTTTCCGACTTCCGACGCCCGGAAAACTT  
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

1140

Glu Tyr Met Arg Ile Glu Pro Val Pro Met Asp Glu Arg Leu Lys Ala Ala Phe Glu  
 -----yurH-----

ACAGCATTAGAAAACGGCTTCAGCTGTGAGGAAATGGTGAGCGGAGCAGGGCATGACCGCG  
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
 TGTCGTAATCTTTTGCCGAAGTCGACACTCCCTTACCACCTCGCCTCGTCCCGTACTGCGCG  
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

1200

Thr Ala Leu Glu Asn Gly Phe Ser Cys Glu Glu Met Val Ser Gly Ala Gly His Asp Ala  
 -----yurH-----

FIG.--1E

CAAATGATCGGAAGGCGCTATCCTTGCTGTATGCTGTTTGTGCCGAGCCGAGGCGGCGGTC 1260

GTTTACTAGCCTTCCGGGATAGGACCAACATACGACAAACACGGCTCGGCTCCGCCCGCAG

Gln Met Ile Gly Arg Arg Tyr Pro Ala Cys Met Leu Phe Val Pro Ser Arg Gly Val  
yurH

AGCCACTCACCGAAGGAATATACGTCAGCCAGACAGCTTGAGATCGGCGTCCCGGCGCACTG 1320

TCGGTGAGTGGCTTCCTTATATGCAGTCGGTCTGTGCGAACTCTAGCCGCGAGGCGCGGTGAC

Ser His Ser Pro Lys Glu Tyr Thr Ser Ala Arg Gln Leu Glu Ile Gly Val Arg Ala Leu  
yurH

ACTGATTTATTATACAAACTGGCTTACTGATAAAGGAGGAATGGCTGTGTCAGGCAGAAG 1380

TGACTAAATAATATGTTGACCGAATGACTATTTCCCTCCTTACCAGACACAGTCCGTCCTC

Thr Asp Leu Leu Tyr Lys Leu Ala Tyr  
yurH

AGAATTATGCACCCCGTTAAGAACGATTATGACACCCGGGCGCTGTGAGGTTGATCCGCG 1440

TCCTAATACGTGGGCAATCTTGCTAATACTGTGGCCCCGGACAACCTCCAAC TAGGCGC

TGTATTAAGA  
ACATAATTCT 1450

FIG.-1F

ID HYUC\_PSESN STANDARD; PRT; 414 AA.DE HYDANTOIN UTILIZATION  
 GN PROTEIN C (ORF4).  
 OS HYUC.  
 OS PSEUDOMONAS SP. (STRAIN NS671).CC -I- FUNCTION: CONVERTS N-CARBAMYL-L-AMINO ACIDS TO L-AMINO ACIDS.  
 CC -I- PATHWAY: CONVERSION OF 5-SUBSTITUTED HYDANTOINS TO CORRESPONDING L-AMINO ACIDS.  
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M40, ALSO KNOWN AS THE AMA/HIPO/HYUC FAMILY OF HYDROLASES.  
 DR EMBL; D10494; G216833; --  
 DR EMBL; M72717; G151284; --  
 KW HYDROLASE; PLASMID; AMINO-ACID BIOSYNTHESIS.  
 SQ SEQUENCE 414 AA; 45683 MW; A5D0C350 CRC32;  
 MKVTISKER LRHIEQLGE IGTKDKGVQ RLALSKEDRE ATLLVSEWMR EAGLTVTHDH  
 FGNLIGRKEG ETPSLPSVMI GSHIDSVRNG GKFDGVIQVL AGIEIVHAIS EANVVHEHSI  
 EVVAFCEEEG SRFNDGLFGS RGMVGKVKPE DLQKVDNNV TRYEALKTFG FGIDPDFTHQ  
 SIREIGDIKH YFEMHIEQGP YLEKNNYPIG IVSGIAGPSW FKVRLVGEAG HAGTVPMSLR  
 KDPLVGAEEV IKEVETLCMN DPNAFTVGTV GRIAAFPGGS NIIPESVEFT LDIRDIELER  
 RNKIIEKIEE KIKLVSNTRG LEYQIEKNMA AVPVKCSENL INSLKQSCKE LEIDAPIIVS  
 GAGHDAMFLA EITEIGMVTV RCRNGISHSP KEWAEIDDL TGTKVLYESI IKHI

[NOTE: generally conserved residues are D94, E129, E130, H195, and E339 using hyuc\_psesn numbering. - DAE]  
 YurH: 412 AA - 3342 kb - Function unknown; similar to Bacillus stearothermophilus  
 N-carboxyl-L-aminoacid amidohydrolase

The score is:                   init1   initn   opt   z-sc E(3981)..  
 BSUPEP:YURH    Begin: 9   End: 409                   326   662   747   804.9   2e-39  
 | 412 AA - 3342 kb  
 hyuc\_psesn.pep  
 BSUPEP:YURH  
 YurH 412 AA - 3342 kb

SCORES    Init1: 326   Initn: 662   Opt: 747   z-score: 804.9   E(): 2e-39  
 Smith-Waterman score: 747;   34.6% identity in 410 aa overlap

FIG.-2A



```

10      20      30      40      50      59
hyuc_psesn.p  MKTVTISKERLRIHIEQLGEIGKTKDKGVQRLALSKEDEATLLVSEWMREAGLTVTHD
|: : : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |
10      20      30      40      50      60
YURH  MEKQKLSVKNSITDYIEWLAQYGASADGGVTRLLYTKREWDAQLAVKTEMSSFGLETRFD
10      20      30      40      50      60
60      70      80      90      100     110     119
hyuc_psesn.p  HFCNLIGRKEGETPSLPSVMIGSHIDSVRNGGKFDGVIQVLAGIEIVHAISEANVVHEHS
| | : | | | | : : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
70      80      90      100     110     120
YURH  DVGNVFGRLSGTQSPDEVIVTGSHTVINGGKYDCAYGVLAAAMLAKQLKETYGAPKKT
120     130     140     150     160     170
hyuc_psesn.p  IEVVAFCSEEGSRFNDGLFGSRGMVGVKVPEDLQKV-DDNNVTRYEALKTFGFGIDPDDFT
: | : | : | | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
130     140     150     160     170
YURH  LEAVSLCEEGSRFPMTYWGSGNMTGVFSEQDAKEPRDESGVSLQTAMHESGFGKG---V
180     190     200     210     220     230
hyuc_psesn.p  HQSIREIGDIKHVFEMHIEQGPYLEKNNYPIGIVSGIAGPSWFKVRLVGEAGHAGTVPMS
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
180     190     200     210     220     230
YURH  FQSAVRT-DISAFVELHIEQKTELEMSGRDLGIVTSGIAGRRYLVTLGECECNHAGTTSMK
240     250     260     270     280     290
hyuc_psesn.p  LRKDPLVGAAEVIKEVETLCMNDPNAPTGVTVGRIAAFPGGSNIIPESVEFTLDIRDIEL
| | | | : : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
240     250     260     270     280     290
YURH  WRKDPAAASSRIIHELLLRSDLELRL-TCGKITAEPNVANVIPGRVQFSIDIR----
300     310     320     330     340     350
hyuc_psesn.p  ERRNKIIEKI-EEKIKLVSNNT--RGLEYQIEKNMAAVPVKSENLINSKQSCKELEID
: : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
300     310     320     330     340     350
YURH  HQHQHVLEQFHQDMVALINGICLQKGIKIRAVIDEYMRIEPVPMDERLKAFAFETALENGFSS
360     370     380     390     400     410
hyuc_psesn.p  APIIVSGAGHDAMFLAEITEIGMVFVRCRNGIASHSPKWEAIEDDILTGTKVLYESIIKHI
: | | | | | | | | : : : : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
360     370     380     390     400     410
YURH  CEEMVSGAGHDAQMIGRRYPACMLFVPSRGGVSHSPKEYTSARQLEIGVRALDILLYKLA
360     370     380     390     400     410
YURH  Y

```

FIG.-2B

## PROTEASES FROM GRAM POSITIVE ORGANISMS

### FIELD OF THE INVENTION

[0001] The present invention relates to metalloproteases derived from gram positive microorganisms. The present invention provides nucleic acid and amino acid sequences of a metalloproteases identified in *Bacillus*. The present invention also provides methods for the production of the metalloprotease in host cells as well as the production of heterologous proteins in a host cell having a mutation or deletion of part or all of the metalloprotease of the present invention.

### BACKGROUND OF THE INVENTION

[0002] Gram positive microorganisms, such as members of the group *Bacillus*, have been used for large-scale industrial fermentation due, in part, to their ability to secrete their fermentation products into the culture media. In gram positive bacteria, secreted proteins are exported across a cell membrane and a cell wall, and then are subsequently released into the external media usually maintaining their native conformation.

[0003] Various gram positive microorganisms are known to secrete extracellular and/or intracellular proteases at some stage in their life cycles. Some of these proteases are produced in large quantities for industrial purposes. However, a negative aspect of the presence of proteases in gram positive organisms is their contribution to the overall degradation of secreted heterologous or foreign proteins.

[0004] The classification of proteases found in microorganisms is based on their catalytic mechanism which results in four groups: serine proteases, metalloproteases, cysteine proteases, and aspartic proteases. These categories can be distinguished by their sensitivity to various inhibitors. For example, serine proteases are inhibited by phenylmethylsulfonylfluoride (PMSF) and diisopropylfluorophosphate (DIFP); metalloproteases by chelating agents; cysteine proteases by iodoacetamide and heavy metals and aspartic proteases by pepstatin. Further, in general, serine proteases have alkaline pH optima, metalloproteases are optimally active around neutrality, and cysteine and aspartic proteases have acidic pH optima (*Biotechnology Handbooks. Bacillus*. Vol. 2, edited by Harwood, 1989, Plenum Press, New York).

[0005] Metalloproteases form the most diverse of the catalytic types of proteases. Family m40 includes bacterial enzymes such as the hippurate hydrolase from *Campylobacter jejuni* (HipO) and the hydantoin utilization protein C (HyuC) from *Pseudomonas sp.*

### SUMMARY OF THE INVENTION

[0006] The present invention relates to the discovery of a heretofore unknown metalloprotease (MP) found in gram positive microorganisms, uses of the MP in industrial applications, and advantageous strain improvements based on genetically engineering such microorganisms to delete, underexpress or overexpress that MP. The present invention is based upon the discovery that MP has overall amino acid relatedness to *Escherichia coli* ptilysin.

[0007] The present invention is based upon Applicant's discovery of this new metalloprotease, MP (YurH), which in addition to providing a new and useful protease and methods

of detecting DNA encoding other such proteases in a gram positive microorganism, provides several advantages which may facilitate optimization and/or modification of strains of gram positive microorganisms, such as *Bacillus*, for expression of desired, e.g. heterologous, proteins. Such optimizations, as described below in detail, allow the construction of strains that can have decreased proteolytic degradation of desired expression products.

[0008] Due to the relatedness of MP to hippurate hydrolase and hydantoin utilization protein C, it can be concluded that MP is also an endopeptidase and would be expected to behave similarly to hippurate hydrolase and hydantoin utilization protein C.

[0009] The present invention encompasses the naturally occurring MP encoded by nucleic acid found in a *Bacillus* species as well as the nucleic acid and amino acid molecules having the sequences disclosed in SEQ ID NOS: 1 and 2. In one embodiment, the gram positive microorganism is a *Bacillus*. In a further embodiment, the *Bacillus* is preferably selected from the group consisting of *Bacillus subtilis*, *Bacillus stearothermophilus*, *Bacillus licheniformis* and *Bacillus amyloliquefaciens*. The invention further provides for a metalloprotease that has at least 80%, preferably at least 90%, most preferably 95% homology with the amino acid sequence of SEQ ID NO: 2. The invention also provides for a nucleic acid which encodes a metalloprotease that has at least 80%, preferably at least 90%, most preferably 95% homology with the nucleotide sequence shown in SEQ ID NO: 1.

[0010] In a preferred embodiment, the present invention encompasses the naturally occurring MP nucleic acid molecule having the sequence found in *Bacillus subtilis* I-168 strain (*Bacillus* Genetic Stock Center, accession number 1A1, Columbus, Ohio) in the region of about 3342 kb from the point of origin. In another preferred embodiment, the *Bacillus subtilis* MP nucleic acid and amino acid molecules have the sequences as shown in FIGS. 1A-1F (SEQ ID NOS:1 and 2).

[0011] The present invention provides isolated polynucleotide and amino acid sequences for *Bacillus subtilis* MP in FIGS. 1A-1F (SEQ ID NOS:1 and 2). Due to the degeneracy of the genetic code, the present invention encompasses any nucleic acid sequence that encodes the *Bacillus subtilis* MP amino acid sequence. The present invention provides expression vectors and host cells comprising a nucleic acid encoding a gram positive MP. The present invention also provides methods of making the gram positive MP.

[0012] The present invention encompasses novel amino acid variations of gram positive MP amino acid sequences disclosed herein that have proteolytic activity. Naturally occurring gram positive MP as well as proteolytically active amino acid variations or derivatives thereof, have application in the textile industry, in cleaning compositions and in animal feed.

[0013] The present invention also encompasses amino acid variations or derivatives of gram positive MP that do not have the characteristic proteolytic activity as long as the nucleic acid sequences encoding such variations or derivatives would have sufficient 5' and 3' coding regions to be capable of being integrated into a gram positive organism genome. Such variants would have applications in gram

positive expression systems where it is desirable to delete, mutate, alter or otherwise incapacitate the naturally occurring metalloprotease in order to diminish or delete its proteolytic activity. Such an expression system would have the advantage of allowing for greater yields of recombinant heterologous proteins or polypeptides.

[0014] The present invention provides methods for detecting gram positive microorganism homologues of *B. subtilis* MP that comprises hybridizing part or all of the nucleic acid encoding *B. subtilis* MP with nucleic acid derived from gram positive organisms, either of genomic or cDNA origin. Accordingly, the present invention provides a method for detecting a gram positive microorganism MP, comprising the steps of hybridizing gram positive microorganism nucleic acid under low stringency conditions to a probe, wherein the probe comprises part or all of the nucleic acid sequence shown in FIGS. 1A-1F (SEQ ID NO:1); and isolating the gram positive nucleic acid which hybridizes to said probe.

[0015] The production of desired heterologous proteins or polypeptides in gram positive microorganisms may be hindered by the presence of one or more proteases which degrade the produced heterologous protein or polypeptide. One advantage of the present invention is that it provides methods and expression systems which can be used to prevent that degradation, thereby enhancing yields of the desired heterologous protein or polypeptide.

[0016] Accordingly, the present invention provides a gram positive microorganism that can be used as a host cell having a mutation or deletion of part or all of the gene encoding MP, which results in the inactivation of the MP proteolytic activity, either alone or in combination with mutations in other proteases, such as apr, npr, epr, mpr, bpf or isp, or other proteases known to those of skill in the art. In one embodiment of the present invention, the gram positive microorganism is a member of the genus *Bacillus*. In a preferred embodiment, the *Bacillus* is selected from the group consisting of *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. coagulans*, *B. circulans*, *B. lautus* and *B. thuringiensis*. In a further preferred embodiment, the *Bacillus* is *Bacillus subtilis*.

[0017] In another aspect, the gram positive host cell having one or more metalloprotease deletions or mutations is further genetically engineered to produce a desired protein. In one embodiment of the present invention, the desired protein is heterologous to the gram positive host cell. In another embodiment, the desired protein is homologous to the host cell.

[0018] In another embodiment, a host cell is engineered to produce MP. The gram positive microorganism may be normally sporulating or non-sporulating. In a preferred embodiment, the gram positive host cell is a *Bacillus*. In another embodiment, the *Bacillus* is selected from the group consisting of *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. coagulans*, *B. circulans*, *B. lautus* and *B. thuringiensis*. In a further preferred embodiment, the *Bacillus* host cell is *Bacillus subtilis*.

[0019] In a further aspect of the present invention, gram positive MP is produced on an industrial fermentation scale

in a microbial host expression system. In another aspect, isolated and purified recombinant MP is used in compositions intended for use in the textile industry, in cleaning compositions, such as detergents, and in animal feeds. Accordingly, the present invention provides a cleaning composition, animal feed and a composition for the treatment of a textile comprising MP. The metalloprotease, MP, may be used alone or in combination with other enzymes and/or mediators or enhancers.

[0020] As noted, the present invention provides a cleaning composition comprising a metalloprotease, MP, comprising the amino acid sequence shown in SEQ ID NO:2. Also provided are cleaning compositions comprising a metalloprotease having at least 80%, preferably 90%, more preferably 95% homology with the amino acid sequence shown in SEQ ID NO:2 or comprising a metalloprotease encoded by a gene that hybridizes with the nucleic acid shown in SEQ ID NO:1.

[0021] Further there is provided an animal feed comprising a metalloprotease, MP, comprising the amino acid sequence shown in SEQ ID NO:2. Also provided are animal feeds comprising a metalloprotease having at least 80%, preferably 90%, more preferably 95% homology with the amino acid sequence shown in SEQ ID NO:2 or comprising a metalloprotease encoded by a gene that hybridizes with the nucleic acid shown in SEQ ID NO:1.

[0022] Also provided is a composition for the treatment of a textile comprising a metalloprotease, MP, comprising the amino acid sequence shown in SEQ ID NO:2. Also provided are compositions for the treatment of a textile comprising a metalloprotease having at least 80%, preferably 90%, more preferably 95% homology with the amino acid sequence shown in SEQ ID NO:2 or comprising a metalloprotease encoded by a gene that hybridizes with the nucleic acid shown in SEQ ID NO:1.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0023] FIGS. 1A-1F show the DNA and amino acid sequence for *Bacillus subtilis* MP (YurH).

[0024] FIGS. 2A-2B show an amino acid alignment of *Escherichia coli* succinyl-diaminopimelate desuccinylase and *Bacillus subtilis* MP (YurH).

#### DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0025] Definitions

[0026] As used herein, the genus *Bacillus* includes all members known to those of skill in the art, including but not limited to *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. coagulans*, *B. circulans*, *B. lautus* and *B. thuringiensis*.

[0027] The present invention relates to a newly characterized metalloprotease (MP) from gram positive organisms. In a preferred embodiment, the gram positive organisms is a *Bacillus*. In another preferred embodiment, the *Bacillus* is selected from the group consisting of *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. coagulans*, *B. circulans*, *B. lautus* and *B. thuringiensis*.

[0028] In another preferred embodiment, the gram positive organism is *Bacillus subtilis* and MP has the amino acid sequence encoded by the nucleic acid molecule having the sequence that occurs around 3342 kilobases from the point of origin of *Bacillus subtilis* I-168.

[0029] In another preferred embodiment, MP has the nucleic acid and amino acid sequence as shown in FIGS. 1A-1F (SEQ ID NOS: 1 and 2). The present invention encompasses the use of amino acid variations of the amino acid sequences disclosed in FIGS. 1A-1F (SEQ ID NO: 2) that have proteolytic activity. Such proteolytic amino acid variants can be used in the textile industry, animal feed and cleaning compositions. The present invention also encompasses the use of *B. subtilis* amino acid variations or derivatives that are not proteolytically active. DNA encoding such variants can be used in methods designed to delete or mutate the naturally occurring host cell MP.

[0030] As used herein, "nucleic acid" refers to a nucleotide or polynucleotide sequence, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be double-stranded or single-stranded, whether representing the sense or antisense strand. As used herein "amino acid" refers to peptide or protein sequences or portions thereof. A "polynucleotide homologue" as used herein refers to a gram positive microorganism polynucleotide that has at least 80%, preferably at least 90% and more preferably at least 95% identity to *B. subtilis* MP, or which is capable of hybridizing to *B. subtilis* MP under conditions of high stringency and which encodes an amino acid sequence having metalloprotease activity.

[0031] The terms "isolated" or "purified" as used herein refer to a nucleic acid or amino acid that is removed from at least one component with which it is naturally associated.

[0032] As used herein, the term "heterologous protein" refers to a protein or polypeptide that does not naturally occur in the chosen gram positive host cell. Examples of heterologous proteins include enzymes such as hydrolases including proteases, cellulases, carbohydrases such as amylases, and lipases; isomerases such as racemases, epimerases, tautomerases, or mutases; oxidases, reductases, transferases, kinases and phosphatases. The heterologous gene may encode therapeutically significant proteins or peptides, such as growth factors, cytokines, ligands, receptors and inhibitors, as well as vaccines and antibodies. The gene may encode commercially important industrial proteins or peptides, such as proteases, carbohydrases such as amylases and glucoamylases, cellulases, oxidases and lipases. The gene of interest may be a naturally occurring gene, a mutated gene or a synthetic gene.

[0033] The term "homologous protein" refers to a protein or polypeptide native or naturally occurring in the chosen gram positive host cell. The invention includes host cells producing the homologous protein via recombinant DNA technology. The present invention encompasses a gram positive host cell having a deletion or interruption of the nucleic acid encoding the naturally occurring homologous protein, such as a protease, and having nucleic acid encoding the homologous protein re-introduced in a recombinant form. In another embodiment, the host cell produces the homologous protein.

[0034] As used herein, the term "overexpressing" when referring to the production of a protein in a host cell means

that the protein is produced in greater amounts than in its naturally occurring environment.

[0035] As used herein, the phrase "proteolytic activity" refers to a protein that is able to hydrolyze a peptide bond. Enzymes having proteolytic activity are described in Enzyme Nomenclature, 1992, edited Webb Academic Press, Inc.

[0036] The unexpected discovery of the metalloprotease MP found in translated uncharacterised *B. subtilis* genomic sequences provides a basis for producing host cells, expression methods and systems which can be used to prevent the degradation of recombinantly produced heterologous proteins.

[0037] Accordingly, in a preferred embodiment, the host cell is a gram positive host cell that has a deletion or mutation in the naturally occurring nucleic acid encoding MP said mutation resulting in deletion or inactivation of the production by the host cell of the MP proteolytic gene product. The host cell may additionally be genetically engineered to produce a desired protein or polypeptide.

[0038] It may also be desired to genetically engineer host cells of any type to produce a gram positive metalloprotease. Such host cells are used in large scale fermentation to produce large quantities of the metalloprotease which may be isolated or purified and used in cleaning products, such as detergents.

#### [0039] I. Metalloprotease Sequences

[0040] The present invention encompasses the use of MP polynucleotide homologues encoding gram positive microorganism metalloproteases MP which have at least 80%, preferably at least 90%, more preferably at least 95% identity to *B. subtilis* MP as long as the homologue encodes a protein that has proteolytic activity. A preferred MP polynucleotide homologue has 96% homology to *B. subtilis* MP.

[0041] Gram positive polynucleotide homologues of *B. subtilis* MP may be obtained by standard procedures known in the art from, for example, cloned DNA (e.g., a DNA "library"), genomic DNA libraries, by chemical synthesis once identified, by cDNA cloning, or by the cloning of genomic DNA, or fragments thereof, purified from a desired cell. (See, for example, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover, D. M. (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K. Vol. I, II.) A preferred source is from genomic DNA.

[0042] As will be understood by those of skill in the art, the polynucleotide sequence and amino acid sequence disclosed in FIGS. 1A-1F may reflect inadvertent errors inherent to nucleic acid sequencing technology. The present invention encompasses the naturally occurring nucleic acid molecule having the nucleic acid sequence obtained from the genomic sequence of *Bacillus* species.

[0043] Nucleic acid encoding *Bacillus subtilis* MP starts around 3342 kilobases counting from the point of origin in the *Bacillus subtilis* strain 1-168 (Anagnostopala, 1961, *J. Bacteriol.*, 81:741-746 or Bacillus Genomic Stock Center, accession 1A1, Columbus, Ohio). The *Bacillus subtilis* point of origin has been described in Ogasawara, N. (1995,

Microbiology 141:Pt.2 257-59). *Bacillus subtilis* MP has a length of 396 amino acids. Based upon the location of the DNA encoding *Bacillus subtilis* MP, naturally occurring *B. subtilis* MP can be obtained by methods known to those of skill in the art including PCR technology.

[0044] Oligonucleotide sequences or primers of about 10-30 nucleotides in length can be designed from the polynucleotide sequence disclosed in FIGS. 1A-1F and used in PCR technology to isolate the naturally occurring sequence from *B. subtilis* genomic sequences.

[0045] Another general strategy for the "cloning" of *B. subtilis* genomic DNA pieces for sequencing uses inverse PCR. A known region is scanned for a set of appropriate restriction enzyme cleavage sites and inverse PCR is performed with a set of DNA primers determined from the outermost DNA sequence. The DNA fragments from the inverse PCR are directly used as template in the sequencing reaction. The newly derived sequences can be used to design new oligonucleotides. These new oligonucleotides are used to amplify DNA fragments with genomic DNA as template. The sequence determination on both strands of a DNA region is finished by applying a primer walking strategy on the genomic PCR fragments. The benefit of multiple starting points in the primer walking results from the series of inverse PCR fragments with different sizes of new "cloned" DNA pieces. From the most external DNA sequence, a new round of inverse PCR is started. The whole inverse PCR strategy is based on the sequential use of conventional taq polymerase and the use of long range inverse PCR in those cases in which the taq polymerase failed to amplify DNA fragments. Nucleic acid sequencing is performed using standard technology. One method for nucleic acid sequencing involves the use of a Perkin-Elmer Applied Biosystems 373 DNA sequencer (Perkin-Elmer, Foster City, Calif.) according to manufacturer's instructions.

[0046] Nucleic acid sequences derived from genomic DNA may contain regulatory regions in addition to coding regions. Whatever the source, the isolated MP gene should be molecularly cloned into a suitable vector for propagation of the gene.

[0047] In molecular cloning of the gene from genomic DNA, DNA fragments are generated, some of which will encode the desired gene. The DNA may be cleaved at specific sites using various restriction enzymes. Alternatively, one may use DNase in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The linear DNA fragments can then be separated according to size by standard techniques, including but not limited to, agarose and polyacrylamide gel electrophoresis and column chromatography.

[0048] Once the DNA fragments are generated, identification of the specific DNA fragment containing the MP may be accomplished in a number of ways. For example, a *B. subtilis* MP gene of the present invention or its specific RNA, or a fragment thereof, such as a probe or primer, may be isolated and labeled and then used in hybridization assays to detect a gram positive MP gene. (Benton, W. and Davis, R., 1977, *Science* 196:180; Grunstein, M. and Hogness, D., 1975, *Proc. Natl. Acad. Sci. USA* 72:3961). Those DNA fragments sharing substantial sequence similarity to the probe will hybridize under stringent conditions.

[0049] Accordingly, the present invention provides a method for the detection of gram positive MP polynucleo-

otide homologues which comprises hybridizing part or all of a nucleic acid sequence of *B. subtilis* MP with gram positive microorganism nucleic acid of either genomic or cDNA origin.

[0050] Also included within the scope of the present invention is the use of gram positive microorganism polynucleotide sequences that are capable of hybridizing to the nucleotide sequence of *B. subtilis* MP under conditions of intermediate to maximal stringency. Hybridization conditions are based on the melting temperature ( $T_m$ ) of the nucleic acid binding complex, as taught in Berger and Kimmel (1987, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, Vol. 152, Academic Press, San Diego, Calif.) incorporated herein by reference, and confer a defined "stringency" as explained below.

[0051] "Maximum stringency" typically occurs at about  $T_m - 5^\circ \text{C}$ . ( $5^\circ \text{C}$ . below the  $T_m$  of the probe); "high stringency" at about  $5^\circ \text{C}$ . to  $10^\circ \text{C}$ . below  $T_m$ ; "intermediate stringency" at about  $10^\circ \text{C}$ . to  $20^\circ \text{C}$ . below  $T_m$ ; and "low stringency" at about  $20^\circ \text{C}$ . to  $25^\circ \text{C}$ . below  $T_m$ . As will be understood by those of skill in the art, a maximum stringency hybridization can be used to identify or detect identical polynucleotide sequences while an intermediate or low stringency hybridization can be used to identify or detect polynucleotide sequence homologues.

[0052] The term "hybridization" as used herein shall include "the process by which a strand of nucleic acid joins with a complementary strand through base pairing" (Coombs, J., (1994), *Dictionary of Biotechnology*, Stockton Press, New York, N.Y.).

[0053] The process of amplification as carried out in polymerase chain reaction (PCR) technologies is described in Dieffenbach, C W and G S Dveksler, (*PCR Primer, a Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y., 1995). A nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides from *B. subtilis* MP, preferably about 12 to 30 nucleotides, and more preferably about 20-25 nucleotides can be used as a probe or PCR primer.

[0054] The *B. subtilis* MP amino acid sequences (shown in FIGS. 1A-1F) were identified via a BLAST search (Altschul, Stephen, Basic local alignment search tool, *J. Mol. Biol.*, 215:403-410) of *Bacillus subtilis* genomic nucleic acid sequences. *B. subtilis* MP (YurH) was identified by its overall nucleic acid identity to the metalloprotease, succinyl-diaminopimelate desuccinylase from *Escherichia coli*.

## [0055] II. Expression Systems

[0056] The present invention provides host cells, expression methods and systems for the enhanced production and secretion of desired heterologous or homologous proteins in gram positive microorganisms. In one embodiment, a host cell is genetically engineered to have a deletion or mutation in the gene encoding a gram positive MP such that the respective activity is deleted. In another embodiment of the present invention, a gram positive microorganism is genetically engineered to produce and/or overproduce a metalloprotease of the present invention.

[0058] Inactivation of a Gram Positive Metalloprotease in a Host Cell

[0059] Producing an expression host cell incapable of producing the naturally occurring metalloprotease necessitates the replacement and/or inactivation of the naturally occurring gene in the genome of the host cell. In a preferred embodiment, the mutation is a non-reverting mutation.

[0060] One method for mutating a nucleic acid encoding a gram positive metalloprotease is to clone the nucleic acid or part thereof, modify the nucleic acid by site directed mutagenesis and reintroduce the mutated nucleic acid into the cell on a plasmid. By homologous recombination, the mutated gene can be introduced into the chromosome. In the parent host cell, the result is that the naturally occurring nucleic acid and the mutated nucleic acid are located in tandem on the chromosome. After a second recombination, the modified sequence is left in the chromosome having thereby effectively introduced the mutation into the chromosomal gene for progeny of the parent host cell.

[0061] Another method for inactivating the metalloprotease proteolytic activity is through deleting the chromosomal gene copy. In a preferred embodiment, the entire gene is deleted, the deletion occurring in such a way as to make reversion impossible. In another preferred embodiment, a partial deletion is produced, provided that the nucleic acid sequence left in the chromosome is too short for homologous recombination with a plasmid encoding the metalloprotease gene. In another preferred embodiment, nucleic acid encoding the catalytic amino acid residues are deleted.

[0062] Deletion of the naturally occurring gram positive microorganism metalloprotease can be carried out as follows. A metalloprotease gene including its 5' and 3' regions is isolated and inserted into a cloning vector. The coding region of the metalloprotease gene is deleted from the vector in vitro, leaving behind a sufficient amount of the 5' and 3' flanking sequences to provide for homologous recombination with the naturally occurring gene in the parent host cell. The vector is then transformed into the gram positive host cell. The vector integrates into the chromosome via homologous recombination in the flanking regions. This method leads to a gram positive strain in which the protease gene has been deleted.

[0063] The vector used in an integration method is preferably a plasmid. A selectable marker may be included to allow for ease of identification of desired recombinant microorganisms. Additionally, as will be appreciated by one of skill in the art, the vector is preferably one which can be selectively integrated into the chromosome. This can be achieved by introducing an inducible origin of replication, for example, a temperature sensitive origin into the plasmid. By growing the transformants at a temperature to which the origin of replication is sensitive, the replication function of the plasmid is inactivated, thereby providing a means for selection of chromosomal integrants. Integrants may be selected for growth at high temperatures in the presence of the selectable marker, such as an antibiotic. Integration mechanisms are described in WO 88/06623.

[0064] Integration by the Campbell-type mechanism can take place in the 5' flanking region of the protease gene, resulting in a protease positive strain carrying the entire plasmid vector in the chromosome in the metalloprotease

locus. Since illegitimate recombination will give different results, it will be necessary to determine whether the complete gene has been deleted, such as through nucleic acid sequencing or restriction maps.

[0065] Another method of inactivating the naturally occurring metalloprotease gene is to mutagenize the chromosomal gene copy by transforming a gram positive microorganism with oligonucleotides which are mutagenic. Alternatively, the chromosomal metalloprotease gene can be replaced with a mutant gene by homologous recombination.

[0066] The present invention encompasses host cells having additional protease deletions or mutations, such as deletion of or mutation(s) in the genes encoding apr, npr, epr, mpr and others known to those of skill in the art.

[0067] One assay for the detection of mutants involves growing the *Bacillus* host cell on medium containing a protease substrate and measuring the appearance or lack thereof, of a zone of clearing or halo around the colonies. Host cells which have an inactive protease will exhibit little or no halo around the colonies.

### [0068] III. Production of Metalloprotease

[0069] For production of metalloprotease in a host cell, an expression vector comprising at least one copy of nucleic acid encoding a gram positive microorganism MP, and preferably comprising multiple copies, is transformed into the host cell under conditions suitable for expression of the metalloprotease. In accordance with the present invention, polynucleotides which encode a gram positive microorganism MP, or fragments thereof, or fusion proteins or polynucleotide homologue sequences that encode amino acid variants of *B. subtilis* MP, may be used to generate recombinant DNA molecules that direct their expression in host cells. In a preferred embodiment, the gram positive host cell belongs to the genus *Bacillus*. In a further preferred embodiment, the gram positive host cell is *B. subtilis*.

[0070] As will be understood by those of skill in the art, it may be advantageous to produce polynucleotide sequences possessing non-naturally occurring codons. Codons preferred by a particular gram positive host cell (Murray, E. et al., (1989), *Nuc. Acids Res.*, 17:477-508) can be selected, for example, to increase the rate of expression or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, than transcripts produced from a naturally occurring sequence.

[0071] Altered MP polynucleotide sequences which may be used in accordance with the invention include deletions, insertions or substitutions of different nucleotide residues resulting in a polynucleotide that encodes the same or a functionally equivalent MP homologue, respectively. As used herein a "deletion" is defined as a change in the nucleotide sequence of the MP resulting in the absence of one or more amino acid residues.

[0072] As used herein, an "insertion" or "addition" is that change in the nucleotide sequence which results in the addition of one or more amino acid residues as compared to the naturally occurring MP.

[0073] As used herein, "substitution" results from the replacement of one or more nucleotides or amino acids by different nucleotides or amino acids, respectively. The

change(s) in the nucleotides(s) can either result in a change in the amino acid sequence or not.

[0074] The encoded protein may also show deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent MP variant. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the variant retains its proteolytic ability. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine; glycine, alanine; asparagine, glutamine; serine, threonine, phenylalanine, and tyrosine.

[0075] The MP polynucleotides of the present invention may be engineered in order to modify the cloning, processing and/or expression of the gene product. For example, mutations may be introduced using techniques which are well known in the art, i.e., site-directed mutagenesis to insert new restriction sites, to alter glycosylation patterns or to change codon preference, for example.

[0076] In one embodiment of the present invention, a gram positive microorganism MP polynucleotide may be ligated to a heterologous sequence to encode a fusion protein. A fusion protein may also be engineered to contain a cleavage site located between the metalloprotease nucleotide sequence and the heterologous protein sequence, so that the metalloprotease may be cleaved and purified away from the heterologous moiety.

#### [0077] IV. Vector Sequences

[0078] Expression vectors used in expressing the metalloproteases of the present invention in gram positive microorganisms comprise at least one promoter associated with MP, which promoter is functional in the host cell. In one embodiment of the present invention, the promoter is the wild-type promoter for the selected metalloprotease and in another embodiment of the present invention, the promoter is heterologous to the metalloprotease, but still functional in the host cell. In one preferred embodiment of the present invention, nucleic acid encoding the metalloprotease is stably integrated into the microorganism genome.

[0079] In a preferred embodiment, the expression vector contains a multiple cloning site cassette which preferably comprises at least one restriction endonuclease site unique to the vector, to facilitate ease of nucleic acid manipulation. In a preferred embodiment, the vector also comprises one or more selectable markers. As used herein, the term "selectable marker" refers to a gene capable of expression in the gram positive host which allows for ease of selection of those hosts containing the vector. Examples of such selectable markers include but are not limited to antibiotics, such as, erythromycin, actinomycin, chloramphenicol and tetracycline.

#### [0080] V. Transformation

[0081] A variety of host cells can be used for the production *Bacillus subtilis* MP or MP homologues including bacterial, fungal, mammalian and insects cells. General transformation procedures are taught in *Current Protocols In*

*Molecular Biology*, (Vol. 1, edited by Ausubel et al., John Wiley & Sons, Inc., 1987, Chapter 9) and include calcium phosphate methods, transformation using DEAE-Dextran and electroporation. Plant transformation methods are taught in Rodriguez (WO 35 95/14099, published May 26, 1995).

[0082] In a preferred embodiment, the host cell is a gram positive microorganism and in another preferred embodiment, the host cell is *Bacillus*. In one embodiment of the present invention, nucleic acid encoding one or more metalloprotease(s) of the present invention is introduced into a host cell via an expression vector capable of replicating within the *Bacillus* host cell. Suitable replicating plasmids for *Bacillus* are described in *Molecular Biological Methods for Bacillus*, Ed. Harwood and Cutting, John Wiley & Sons, 1990, hereby expressly incorporated by reference; see chapter 3 on plasmids. Suitable replicating plasmids for *B. subtilis* are listed on page 92.

[0083] In another embodiment, nucleic acid encoding a metalloprotease(s) of the present invention is stably integrated into the microorganism genome. Preferred host cells are gram positive host cells. Another preferred host is *Bacillus*. Another preferred host is *Bacillus subtilis*. Several strategies have been described in the literature for the direct cloning of DNA in *Bacillus*. Plasmid marker rescue transformation involves the uptake of a donor plasmid by competent cells carrying a partially homologous resident plasmid (Contente et al., *Plasmid*, 2:555-571 (1979); Haima et al., *Mol. Gen. Genet.*, 223:185-191 (1990); Weinrauch et al., *J. Bacteriol.*, 154(3):1077-1087 (1983); and Weinrauch et al., *J. Bacteriol.*, 169(3):1205-1211 (1987)). The incoming donor plasmid recombines with the homologous region of the resident "helper" plasmid in a process that mimics chromosomal transformation.

[0084] Transformation by protoplast transformation is described for *B. subtilis* in Chang and Cohen, (1979), *Mol. Gen. Genet.*, 168:111-115; for *B. megaterium* in Vorobjeva et al., (1980), *FEMS Microbiol. Letters*, 7:261-263; for *B. amyloliquefaciens* in Smith et al., (1986), *Appl. and Env. Microbiol.*, 51:634; for *B. thuringiensis* in Fisher et al., (1981), *Arch. Microbiol.*, 139:213-217; for *B. sphaericus* in McDonald, (1984), *J. Gen. Microbiol.*, 130:203; and *B. larvae* in Bakhiet et al., (1985, *Appl. Environ. Microbiol.* 49:577). Mann et al., (1986, *Current Microbiol.*, 13:131-135) report on transformation of *Bacillus* protoplasts and Holubova, (1985), *Folia Microbiol.*, 30:97) disclose methods for introducing DNA into protoplasts using DNA containing liposomes.

#### [0085] VI. Identification of Transformants

[0086] Whether a host cell has been transformed with a mutated or a naturally occurring gene encoding a gram positive MP, detection of the presence/absence of marker gene expression can suggest whether the gene of interest is present. However, its expression should be confirmed. For example, if the nucleic acid encoding a metalloprotease is inserted within a marker gene sequence, recombinant cells containing the insert can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with nucleic acid encoding the metalloprotease under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the metalloprotease as well.

[0087] Alternatively, host cells which contain the coding sequence for a metalloprotease and express the protein may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridization and protein bioassay or immunoassay techniques which include membrane-based, solution-based, or chip-based technologies for the detection and/or quantification of the nucleic acid or protein.

[0088] The presence of the metalloprotease polynucleotide sequence can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes, portions or fragments of *B. subtilis* MP.

#### [0089] VII. Assay of Protease Activity

[0090] There are various assays known to those of skill in the art for detecting and measuring protease activity. There are assays based upon the release of acid-soluble peptides from casein or hemoglobin measured as absorbance at 280 nm or calorimetrically using the Folin method (Bergmeyer, et al., 1984, *Methods of Enzymatic Analysis*, Vol. 5, Peptidases, Proteinases and their Inhibitors, Verlag Chemie, Weinheim). Other assays involve the solubilization of chromogenic substrates (Ward, 1983, Proteinases, in *Microbial Enzymes and Biotechnology*, (W. M. Fogarty, ed.), Applied Science, London, pp. 251-317).

#### [0091] VIII. Secretion of Recombinant Proteins

[0092] Means for determining the levels of secretion of a heterologous or homologous protein in a gram positive host cell and detecting secreted proteins include using either polyclonal or monoclonal antibodies specific for the protein. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). These and other assays are described, among other places, in Hampton, R. et al., (1990, *Serological Methods, a Laboratory Manual*, APS Press, St. Paul Minn.) and Maddox, D E et al., (1983, *J. Exp. Med.*, 158:1211).

[0093] A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic and amino acid assays. -Means for producing labeled hybridization or PCR probes for detecting specific polynucleotide sequences include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the nucleotide sequence, or any portion of it, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides.

[0094] A number of companies such as Pharmacia Biotech (Piscataway N.J.), Promega (Madison Wis.), and US Biochemical Corp. (Cleveland Ohio) supply commercial kits and protocols for these procedures. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like. Patents teaching the use of such labels include U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immuno-

globulins may be produced as shown in U.S. Pat. No. 4,816,567 and incorporated herein by reference.

#### [0095] IX. Purification of Proteins

[0096] Gram positive host cells transformed with polynucleotide sequences encoding heterologous or homologous protein may be cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein produced by a recombinant gram positive host cell comprising a mutation or deletion of the metalloprotease activity will be secreted into the culture media. Other recombinant constructions may join the heterologous or homologous polynucleotide sequences to a nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins (Kroll, D J. et al., (1993), *DNA Cell Biol* 12:441-53).

[0097] Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals (Porath, J., (1992), *Protein Expr. Purif.* 3:263-281), protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle Wash.). The inclusion of a cleavable linker sequence such as Factor XA or enterokinase (Invitrogen, San Diego Calif.) between the purification domain and the heterologous protein can be used to facilitate purification.

#### [0098] X. Uses of The Present Invention

##### [0099] MP and Genetically Engineered Host Cells

[0100] The present invention provides genetically engineered host cells comprising mutations, preferably non-revertable mutations, or deletions in the naturally occurring gene encoding MP such that the proteolytic activity is diminished or deleted altogether. The host cell may contain additional protease deletions, such as deletions of the mature subtilisin prbtease and/or mature neutral protease disclosed in U.S. Pat. No. 5,264,366.

[0101] In a preferred embodiment, the host cell is further genetically engineered to produce a desired protein or polypeptide. In a preferred embodiment, the host cell is a *Bacillus*. In a further preferred embodiment, the host cell is a *Bacillus subtilis*.

[0102] In an alternative embodiment, a host cell is genetically engineered to produce a gram positive MP. In a preferred embodiment, the host cell is grown under large scale fermentation conditions. In another preferred embodiment, the MP is isolated and/or purified and used in the textile industry, the feed industry and in cleaning compositions such as detergents.

[0103] As noted, MP can be useful in formulating various cleaning compositions. A number of known compounds are suitable surfactants useful in compositions comprising the MP of the invention. These include nonionic, anionic, cationic, anionic or zwitterionic detergents, as disclosed in U.S. Pat. No. 4,404,128 and U.S. Pat. No. 4,261,868. A suitable detergent formulation is that described in Example 7 of U.S. Pat. No. 5,204,015. The art is familiar with the different formulations which can be used as cleaning compositions. In addition, MP can be used, for example, in bar or liquid soap applications, dishcare formulations, contact lens cleaning solutions or products, peptide hydrolysis, waste treatment,



textile applications, as fusion-cleavage enzymes in protein production, etc. MP may comprise enhanced performance in a detergent composition (as compared to another detergent protease). As used herein, enhanced performance in a detergent is defined as increasing cleaning of certain enzyme sensitive stains such as grass or blood, as determined by usual evaluation after a standard wash cycle.

[0104] MP can be formulated into known powdered and liquid detergents having pH between 6.5 and 12.0 at levels of about 0.01 to about 5% (preferably 0.1% to 0.5%) by weight. These detergent cleaning compositions can also include other enzymes such as known proteases, amylases, cellulases, lipases or endoglycosidases, as well as builders and stabilizers.

[0105] The addition of MP to conventional cleaning compositions does not create any special use limitation. In other words, any temperature and pH suitable for the detergent is also suitable for the present compositions as long as the pH is within the above range, and the temperature is below the described MP's denaturing temperature. In addition, MP can be used in a cleaning composition without detergents, again either alone or in combination with builders and stabilizers.

[0106] Proteases can be included in animal feed such as part of animal feed additives as described in, for example, U.S. Pat. No. 5,612,055; U.S. Pat. No. 5,314,692; and U.S. Pat. No. 5,147,642.

[0107] One aspect of the invention is a composition for the treatment of a textile that includes MP. The composition can be used to treat for example silk or wool as described in publications such as RD 216,034; EP 134,267; U.S. Pat. No. 4,533,359; and EP 344,259.

[0108] MP Polynucleotides

[0109] *AB. subtilis* MP polynucleotide, or any part thereof, provides the basis for detecting the presence of gram positive microorganism MP polynucleotide homologues through hybridization techniques and PCR technology.

[0110] Accordingly, one aspect of the present invention is to provide for nucleic acid hybridization and PCR probes which can be used to detect polynucleotide sequences, including genomic and cDNA sequences, encoding gram positive MP or portions thereof.

[0111] The manner and method of carrying out the present invention may be more fully understood by those of skill in the art by reference to the following examples, which examples are not intended in any manner to limit the scope of the present invention or of the claims directed thereto

#### EXAMPLE I

[0112] Preparation of a Genomic Library

[0113] The following example illustrates the preparation of a *Bacillus* genomic library.

[0114] Genomic DNA from *Bacillus* cells is prepared as taught in *Current Protocols In Molecular Biology*, Vol. 1, edited by Ausubel et al., John Wiley & Sons, Inc., 1987, Chapter 2. 4.1. Generally, *Bacillus* cells from a saturated liquid culture are lysed and the proteins removed by digestion with proteinase K. Cell wall debris, polysaccharides, and remaining proteins are removed by selective precipitation with CTAB, and high molecular weight genomic DNA

is recovered from the resulting supernatant by isopropanol precipitation. If exceptionally clean genomic DNA is desired, an additional step of purifying the *Bacillus* genomic DNA on a cesium chloride gradient is added.

[0115] After obtaining purified genomic DNA, the DNA is subjected to Sau3A digestion. Sau3A recognizes the 4 base pair site GATC and generates fragments compatible with several convenient phage lambda and cosmid vectors. The DNA is subjected to partial digestion to increase the chance of obtaining random fragments.

[0116] The partially digested *Bacillus* genomic DNA is subjected to size fractionation on a 1% agarose gel prior to cloning into a vector. Alternatively, size fractionation on a sucrose gradient can be used. The genomic DNA obtained from the size fractionation step is purified away from the agarose and ligated into a cloning vector appropriate for use in a host cell and transformed into the host cell.

#### EXAMPLE II

[0117] Detection of Gram Positive Microorganisms

[0118] The following example describes the detection of gram positive microorganism MP.

[0119] DNA derived from a gram positive microorganism is prepared according to the methods disclosed in *Current Protocols in Molecular Biology*, Chap. 2 or 3. The nucleic acid is subjected to hybridization and/or PCR amplification with a probe or primer derived from MP.

[0120] The nucleic acid probe is labeled by combining 50 pmol of the nucleic acid and 250 mCi of [ $\gamma$   $^{32}$ P] adenosine triphosphate (Amersham, Chicago Ill.) and T4 polynucleotide kinase (DuPont NEN®, Boston Mass.). The labeled probe is purified with Sephadex G-25 super fine resin column (Pharmacia). A portion containing  $10^7$  counts per minute of each is used in a typical membrane based hybridization analysis of nucleic acid sample of either genomic or cDNA origin.

[0121] The DNA sample which has been subjected to restriction endonuclease digestion is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham N.H.). Hybridization is carried out for 16 hours at 40 degrees C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1x saline sodium citrate and 0.5% sodium dodecyl sulfate. The blots are exposed to film for several hours, the film developed and hybridization patterns are compared visually to detect polynucleotide homologues of *B. subtilis* MP. The homologues are subjected to confirmatory nucleic acid sequencing. Methods for nucleic acid sequencing are well known in the art. Conventional enzymatic methods employ DNA polymerase Klenow fragment, SEQUENASE® (US Biochemical Corp, Cleveland, Ohio) or Taq polymerase to extend DNA chains from an oligonucleotide primer annealed to the DNA template of interest.

[0122] Various other examples and modifications of the foregoing description and examples will be apparent to a person skilled in the art after reading the disclosure without departing from the spirit and scope of the invention, and it is intended that all such examples or modifications be included within the scope of the appended claims. All publications and patents referenced herein are hereby incorporated by reference in their entirety.

## SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 2

<210> SEQ ID NO 1

<211> LENGTH: 1450

<212> TYPE: DNA

<213> ORGANISM: Bacillus subtilisin

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (112)...(1347)

<400> SEQUENCE: 1

```

tggatggata ctgtggctaa aaaataaacg tttttttggt taatctgtca atgaactatc      60
atgttgctcc ttctataatg aagtcaccat agcggccaga ggggggagta c atg gaa      117
                                     Met Glu
                                     1
aag cag aag ctt tca gta aaa aac agc atc aca gat tat att gaa tgg      165
Lys Gln Lys Leu Ser Val Lys Asn Ser Ile Thr Asp Tyr Ile Glu Trp
      5              10              15
ctc gcc caa tac ggt gcg tct gcg gat ggc ggc gta acg agg ctt tta      213
Leu Ala Gln Tyr Gly Ala Ser Ala Asp Gly Gly Val Thr Arg Leu Leu
      20              25              30
tac aca aaa gag tgg atg gac gca cag ctt gct gtg aaa acg gaa atg      261
Tyr Thr Lys Glu Trp Met Asp Ala Gln Leu Ala Val Lys Thr Glu Met
      35              40              45              50
tct tct ttc ggg ctt gaa aca aga ttc gac gat gtc ggc aat gtg ttt      309
Ser Ser Phe Gly Leu Glu Thr Arg Phe Asp Asp Val Gly Asn Val Phe
      55              60              65
gga agg ctt tcg ggc acc cag tct ccg gat gaa gtc att gta acc ggt      357
Gly Arg Leu Ser Gly Thr Gln Ser Pro Asp Glu Val Ile Val Thr Gly
      70              75              80
tca cat atc gat act gtc atc aat gga gga aaa tac gac ggc gcc tat      405
Ser His Ile Asp Thr Val Ile Asn Gly Gly Lys Tyr Asp Gly Ala Tyr
      85              90              95
ggc gtt ctg gcg gca atg ctc gcg tta aag cag ctc aag gaa aca tac      453
Gly Val Leu Ala Ala Met Leu Ala Leu Lys Gln Leu Lys Glu Thr Tyr
      100             105             110
gga gcg cca aaa aaa acg ctc gag gcg gtc tcc tta tgt gag gag gaa      501
Gly Ala Pro Lys Lys Thr Leu Glu Ala Val Ser Leu Cys Glu Glu Glu
      115             120             125             130
ggc agc cgt ttt cca atg acc tat tgg ggg tcg ggg aat atg acg ggt      549
Gly Ser Arg Phe Pro Met Thr Tyr Trp Gly Ser Gly Asn Met Thr Gly
      135             140             145
gtt ttt tca gaa caa gat gca aaa gag cca aga gac gag tcg gga gtc      597
Val Phe Ser Glu Gln Asp Ala Lys Glu Pro Arg Asp Glu Ser Gly Val
      150             155             160
tcc ctc caa aca gcg atg cat gag agc ggt ttt ggc aaa ggt gtg ttt      645
Ser Leu Gln Thr Ala Met His Glu Ser Gly Phe Gly Lys Gly Val Phe
      165             170             175
cag tca gca tat agg aca gat atc agc gcg ttt gtt gag ctg cat att      693
Gln Ser Ala Tyr Arg Thr Asp Ile Ser Ala Phe Val Glu Leu His Ile
      180             185             190
gaa cag ggg aag acg ttg gaa atg tca ggc cgg gat ctc ggc atc gtg      741
Glu Gln Gly Lys Thr Leu Glu Met Ser Gly Arg Asp Leu Gly Ile Val
      195             200             205             210
aca agt att gcg ggg cag aga cga tat ctc gtc acg ctc gaa gga gaa      789
Thr Ser Ile Ala Gly Gln Arg Arg Tyr Leu Val Thr Leu Glu Gly Glu
      215             220             225

```

-continued

---

```

tgc aat cac gca gga acc acc tcc atg aaa tgg cgc aag gac ccg ctc      837
Cys Asn His Ala Gly Thr Thr Ser Met Lys Trp Arg Lys Asp Pro Leu
      230                      235                      240

gca gcc agc agc cgt atc att cat gag ctg ctg ctg cgg tcg gat gag      885
Ala Ala Ser Ser Arg Ile Ile His Glu Leu Leu Leu Arg Ser Asp Glu
      245                      250                      255

ctg ccg gat gag ctc cgt ctg aca tgc gga aaa ata acg gca gag ccc      933
Leu Pro Asp Glu Leu Arg Leu Thr Cys Gly Lys Ile Thr Ala Glu Pro
      260                      265                      270

aat gta gcc aat gtc ata ccg ggc cgc gtc cag ttt tca atc gat att      981
Asn Val Ala Asn Val Ile Pro Gly Arg Val Gln Phe Ser Ile Asp Ile
      275                      280                      285                      290

cgc cat cag cat cag cat gtg ctg gaa cag ttt cat caa gac atg gtt      1029
Arg His Gln His Gln His Val Leu Glu Gln Phe His Gln Asp Met Val
      295                      300                      305

gct ttg atc aac ggc att tgc cta caa aaa gga att cgc gct gtg att      1077
Ala Leu Ile Asn Gly Ile Cys Leu Gln Lys Gly Ile Arg Ala Val Ile
      310                      315                      320

gat gaa tat atg cgg ata gag cct gtg ccg atg gac gaa agg ctg aag      1125
Asp Glu Tyr Met Arg Ile Glu Pro Val Pro Met Asp Glu Arg Leu Lys
      325                      330                      335

gct gcg gct ttt gaa aca gca tta gaa aac ggc ttc agc tgt gag gaa      1173
Ala Ala Ala Phe Glu Thr Ala Leu Glu Asn Gly Phe Ser Cys Glu Glu
      340                      345                      350

atg gtg agc gga gca ggg cat gac gcg caa atg atc gga agg cgc tat      1221
Met Val Ser Gly Ala Gly His Asp Ala Gln Met Ile Gly Arg Arg Tyr
      355                      360                      365                      370

cct gct tgt atg ctg ttt gtg ccg agc cga ggc ggc gtc agc cac tca      1269
Pro Ala Cys Met Leu Phe Val Pro Ser Arg Gly Gly Val Ser His Ser
      375                      380                      385

ccg aag gaa tat acg tca gcc aga cag ctt gag atc ggc gtc cgc gca      1317
Pro Lys Glu Tyr Thr Ser Ala Arg Gln Leu Glu Ile Gly Val Arg Ala
      390                      395                      400

ctg act gat tta tta tac aaa ctg gct tac tgataaagga ggaatggctg      1367
Leu Thr Asp Leu Leu Tyr Lys Leu Ala Tyr
      405                      410

tgtcaggcag aagagaatta tgcaccccgt taagaacgat tatgacaccg gggcctgttg      1427

aggttgatcc gcgtgtatta aga                                          1450

```

```

<210> SEQ ID NO 2
<211> LENGTH: 412
<212> TYPE: PRT
<213> ORGANISM: Bacillus subtilisin

```

```

<400> SEQUENCE: 2

```

```

Met Glu Lys Gln Lys Leu Ser Val Lys Asn Ser Ile Thr Asp Tyr Ile
 1                      5                      10                      15

Glu Trp Leu Ala Gln Tyr Gly Ala Ser Ala Asp Gly Gly Val Thr Arg
 20                      25                      30

Leu Leu Tyr Thr Lys Glu Trp Met Asp Ala Gln Leu Ala Val Lys Thr
 35                      40                      45

Glu Met Ser Ser Phe Gly Leu Thr Arg Phe Asp Asp Val Gly Asn
 50                      55                      60

Val Phe Gly Arg Leu Ser Gly Thr Gln Ser Pro Asp Glu Val Ile Val
 65                      70                      75                      80

```

-continued

---

Thr Gly Ser His Ile Asp Thr Val Ile Asn Gly Gly Lys Tyr Asp Gly  
 85 90 95

Ala Tyr Gly Val Leu Ala Ala Met Leu Ala Leu Lys Gln Leu Lys Glu  
 100 105 110

Thr Tyr Gly Ala Pro Lys Lys Thr Leu Glu Ala Val Ser Leu Cys Glu  
 115 120 125

Glu Glu Gly Ser Arg Phe Pro Met Thr Tyr Trp Gly Ser Gly Asn Met  
 130 135 140

Thr Gly Val Phe Ser Glu Gln Asp Ala Lys Glu Pro Arg Asp Glu Ser  
 145 150 155 160

Gly Val Ser Leu Gln Thr Ala Met His Glu Ser Gly Phe Gly Lys Gly  
 165 170 175

Val Phe Gln Ser Ala Tyr Arg Thr Asp Ile Ser Ala Phe Val Glu Leu  
 180 185 190

His Ile Glu Gln Gly Lys Thr Leu Glu Met Ser Gly Arg Asp Leu Gly  
 195 200 205

Ile Val Thr Ser Ile Ala Gly Gln Arg Arg Tyr Leu Val Thr Leu Glu  
 210 215 220

Gly Glu Cys Asn His Ala Gly Thr Thr Ser Met Lys Trp Arg Lys Asp  
 225 230 235 240

Pro Leu Ala Ala Ser Ser Arg Ile Ile His Glu Leu Leu Leu Arg Ser  
 245 250 255

Asp Glu Leu Pro Asp Glu Leu Arg Leu Thr Cys Gly Lys Ile Thr Ala  
 260 265 270

Glu Pro Asn Val Ala Asn Val Ile Pro Gly Arg Val Gln Phe Ser Ile  
 275 280 285

Asp Ile Arg His Gln His Gln His Val Leu Glu Gln Phe His Gln Asp  
 290 295 300

Met Val Ala Leu Ile Asn Gly Ile Cys Leu Gln Lys Gly Ile Arg Ala  
 305 310 315 320

Val Ile Asp Glu Tyr Met Arg Ile Glu Pro Val Pro Met Asp Glu Arg  
 325 330 335

Leu Lys Ala Ala Ala Phe Glu Thr Ala Leu Glu Asn Gly Phe Ser Cys  
 340 345 350

Glu Glu Met Val Ser Gly Ala Gly His Asp Ala Gln Met Ile Gly Arg  
 355 360 365

Arg Tyr Pro Ala Cys Met Leu Phe Val Pro Ser Arg Gly Gly Val Ser  
 370 375 380

His Ser Pro Lys Glu Tyr Thr Ser Ala Arg Gln Leu Glu Ile Gly Val  
 385 390 395 400

Arg Ala Leu Thr Asp Leu Leu Tyr Lys Leu Ala Tyr  
 405 410

---

1-10. (canceled)

11. A cleaning composition comprising a metalloprotease encoded by a gene that hybridizes with the nucleic acid shown in SEQ ID NO:1.

12-18. (canceled)

19. A composition for the treatment of a textile comprising a metalloprotease encoded by a gene that hybridizes with the nucleic acid shown in SEQ ID NO:1.

20. The composition for the treatment of a textile of claim 19, wherein the hybridization takes place under low stringency conditions.

21-25. (canceled)

\* \* \* \* \*