POLYPEPTIDES HAVING ANTIMICROBIAL ACTIVITY AND POLYNUCLEOTIDES ENCODING SAME

FIELD OF THE INVENTION

The present invention relates to isolated polypeptides having antimicrobial activity and isolated polynucleotides encoding the polypeptides. The invention also relates to nucleic acid constructs, vectors, and host cells comprising the polynucleotides as well as methods for producing and using the polypeptides.

BACKGROUND OF THE INVENTION

It is an object of the present invention to provide polypeptides having antimicrobial activity and polynucleotides encoding the polypeptides.

SUMMARY OF THE INVENTION

The present invention relates to polypeptides having antimicrobial activity, which comprises an amino acid sequence represented by: C-x(3)-C-x(7,9)-C-C; C-C-x(8)-C-x-C; or C-x-C-x(8,11)-C.

In an embodiment, the polypeptides are defensins.

In another embodiment, the polypeptides are selected from the group consisting of:

(a) a polypeptide having an amino acid sequence which has at least 60% identity with:
amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;
amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;
amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;
amino acids 1 to 45 of SEQ ID NO:22;
amino acids 1 to 49 of SEQ ID NO:24;
amino acids 1 to 44 of SEQ ID NO:26; or
amino acids 1 to 59 of SEQ ID NO:28.

(b) a polypeptide which is encoded by a nucleotide sequence which hybridizes under at least medium stringency conditions with (i)
nucleotides 151 to 297 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9;
nucleotides 118 to 255 of SEQ ID NO:11, SEQ ID NO:13;
nucleotides 112 to 255 of SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19;
nucleotides 118 to 252 of SEQ ID NO:21;
nucleotides 151 to 297 of SEQ ID NO:23;
nucleotides 121 to 252 of SEQ ID NO:25; or

nucleotides 145 to 321 of SEQ ID NO:27; or (ii) a complementary strand of (i); and

(c) a variant comprising a conservative substitution, deletion, and/or insertion of one or more amino acids of:

amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;

amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;

amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;

amino acids 1 to 45 of SEQ ID NO:22;

amino acids 1 to 49 of SEQ ID NO:24;

amino acids 1 to 44 of SEQ ID NO:26; or

amino acids 1 to 59 of SEQ ID NO:28.

The present invention also relates to isolated polynucleotides encoding polypeptides having antimicrobial activity, selected from the group consisting of:

(a) a polynucleotide encoding a polypeptide having an amino acid sequence which has at least 60% identity with:

amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;

amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;

amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;

amino acids 1 to 45 of SEQ ID NO:22;

amino acids 1 to 49 of SEQ ID NO:24;

amino acids 1 to 44 of SEQ ID NO:26; or

amino acids 1 to 59 of SEQ ID NO:28.

(b) a polynucleotide having at least 60% identity with

nucleotides 151 to 297 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9;

nucleotides 118 to 255 of SEQ ID NO:11, SEQ ID NO:13;

nucleotides 112 to 255 of SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19;

nucleotides 118 to 252 of SEQ ID NO:21;

nucleotides 151 to 297 of SEQ ID NO:23;

nucleotides 121 to 252 of SEQ ID NO:25; or

nucleotides 145 to 321 of SEQ ID NO:27; and

(c) a polynucleotide which hybridizes under at least medium stringency conditions with (i):

nucleotides 151 to 297 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9;

nucleotides 118 to 255 of SEQ ID NO:11, SEQ ID NO:13;

nucleotides 112 to 255 of SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19;
nucleotides 118 to 252 of SEQ ID NO:21;
nucleotides 151 to 297 of SEQ ID NO:23;
nucleotides 121 to 252 of SEQ ID NO:25; or
nucleotides 145 to 321 of SEQ ID NO:27; or (ii) a complementary strand of (i).

The present invention also relates to nucleic acid constructs, recombinant expression vectors, and recombinant host cells comprising the polynucleotides.

The present invention also relates to methods for producing such polypeptides having antimicrobial activity comprising (a) cultivating a recombinant host cell comprising a nucleic acid construct comprising a polynucleotide encoding the polypeptide under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide.

The present invention also relates to methods of using the polypeptides and polynucleotides of the invention.

DEFINITIONS

Antimicrobial activity: The term "antimicrobial activity" is defined herein as an activity which is capable of killing or inhibiting growth of microbial cells. In the context of the present invention the term "antimicrobial" is intended to mean that there is a bactericidal and/or a bacteriostatic and/or fungicidal and/or fungistatic effect and/or a virucidal effect, wherein the term "bactericidal" is to be understood as capable of killing bacterial cells. The term "bacteriostatic" is to be understood as capable of inhibiting bacterial growth, i.e. inhibiting growing bacterial cells. The term "fungicidal" is to be understood as capable of killing fungal cells. The term "fungistatic" is to be understood as capable of inhibiting fungal growth, i.e. inhibiting growing fungal cells. The term "virucidal" is to be understood as capable of inactivating virus. The term "microbial cells" denotes bacterial or fungal cells (including yeasts).

In the context of the present invention the term "inhibiting growth of microbial cells" is intended to mean that the cells are in the non-growing state, i.e., that they are not able to propagate.

For purposes of the present invention, antimicrobial activity may be determined according to the procedure described by Lehrer *et al.*, Journal of Immunological Methods, Vol. 137 (2) pp. 167-174 (1991). Alternatively, antimicrobial activity may be determined according to the NCCLS guidelines from CLSI (Clinical and Laboratory Standards Institute; formerly known as National Committee for Clinical and Laboratory Standards).

Polypeptides having antimicrobial activity may be capable of reducing the number of living cells of *Escherichia coli* (DSM 1576) to 1/100 after 8 hours (preferably after 4 hours, more preferably after 2 hours, most preferably after 1 hour, and in particular after 30 minutes)

incubation at 20°C in an aqueous solution of 25%(w/w); preferably in an aqueous solution of 10%(w/w); more preferably in an aqueous solution of 5%(w/w); even more preferably in an aqueous solution of 1%(w/w); most preferably in an aqueous solution of 0.5%(w/w); and in particular in an aqueous solution of 0.1%(w/w) of the polypeptides having antimicrobial activity.

Polypeptides having antimicrobial activity may also be capable of inhibiting the outgrowth of *Escherichia coli* (DSM 1576) for 24 hours at 25°C in a microbial growth substrate, when added in a concentration of 1000 ppm; preferably when added in a concentration of 500 ppm; more preferably when added in a concentration of 250 ppm; even more preferably when added in a concentration of 100 ppm; most preferably when added in a concentration of 50 ppm; and in particular when added in a concentration of 25 ppm.

Polypeptides having antimicrobial activity may be capable of reducing the number of living cells of *Bacillus subtilis* (ATCC 6633) to 1/100 after 8 hours (preferably after 4 hours, more preferably after 2 hours, most preferably after 1 hour, and in particular after 30 minutes) incubation at 20°C in an aqueous solution of 25%(w/w); preferably in an aqueous solution of 10%(w/w); more preferably in an aqueous solution of 5%(w/w); even more preferably in an aqueous solution of 1%(w/w); most preferably in an aqueous solution of 0.5%(w/w); and in particular in an aqueous solution of 0.1%(w/w) of the polypeptides having antimicrobial activity.

Polypeptides having antimicrobial activity may also be capable of inhibiting the outgrowth of *Bacillus subtilis* (ATCC 6633) for 24 hours at 25°C in a microbial growth substrate, when added in a concentration of 1000 ppm; preferably when added in a concentration of 500 ppm; more preferably when added in a concentration of 250 ppm; even more preferably when added in a concentration of 100 ppm; most preferably when added in a concentration of 50 ppm; and in particular when added in a concentration of 25 ppm.

The polypeptides of the present invention have at least 20%, preferably at least 40%, more preferably at least 50%, more preferably at least 60%, more preferably at least 70%, more preferably at least 80%, even more preferably at least 90%, most preferably at least 95%, and even most preferably at least 100% of the antimicrobial activity of the polypeptide consisting of the amino acid sequence shown as

amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;

amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;

amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;

amino acids 1 to 45 of SEQ ID NO:22;

amino acids 1 to 49 of SEQ ID NO:24;

amino acids 1 to 44 of SEQ ID NO:26; or

amino acids 1 to 59 of SEQ ID NO:28.

Defensin: The term "defensin" as used herein refers to polypeptides recognized by a

person skilled in the art as belonging to the defensin class of antimicrobial peptides. To determine if a polypeptide is a defensin according to the invention, the amino acid sequence is preferably compared with the hidden markov model profiles (HMM profiles) of the PFAM database by using the freely available HMMER software package (see Example 6).

The PFAM defensin families include Defensin_1 or "Mammalian defensin" (accession no. PF00323), Defensin_2 or "Arthropod defensin" (accession no. PF01097), Defensin_beta or "Beta Defensin" (accession no. PF00711), Defensin_propep or "Defensin propeptide" (accession no. PF00879) and Gamma-thionin or "Gamma-thionins family" (accession no. PF00304).

The defensins may belong to the alpha-defensin class, the beta-defensin class, the theta-defensin class, the insect or arthropod defensin classes, or the plant defensin class.

In an embodiment, the amino acid sequence of a defensin according to the invention comprises 4, 5, 6, 7, 8, 9, or 10 cysteine residues, preferably 6, 7, 8, 9, or 10 cysteine residues, more preferably 6, 8, or 10 cysteine residues, and most preferably 6 or 8 cysteine residues.

The defensins may also be synthetic defensins sharing the characteristic features of any of the defensin classes.

Examples of defensins include, but are not limited to, α-Defensin HNP-1 (human neutrophil peptide) HNP-2 and HNP-3; β-Defensin-12, Drosomycin, Heliomicin, γ1-purothionin, Insect defensin A, and the defensins disclosed in PCT applications WO 99/53053, WO 02/085934, WO 03/044049, PCT/DK2005/000725 and PCT/DK2005/000735.

Isolated polypeptide: The term "isolated polypeptide" as used herein refers to a polypeptide which is at least 20% pure, preferably at least 40% pure, more preferably at least 60% pure, even more preferably at least 80% pure, most preferably at least 90% pure, and even most preferably at least 95% pure, as determined by SDS-PAGE.

Substantially pure polypeptide: The term "substantially pure polypeptide" denotes herein a polypeptide preparation which contains at most 10%, preferably at most 8%, more preferably at most 6%, more preferably at most 5%, more preferably at most 4%, at most 3%, even more preferably at most 2%, most preferably at most 1%, and even most preferably at most 0.5% by weight of other polypeptide material with which it is natively associated. It is, therefore, preferred that the substantially pure polypeptide is at least 92% pure, preferably at least 94% pure, more preferably at least 95% pure, more preferably at least 96% pure, more preferably at least 96% pure, more preferably at least 97% pure, more preferably at least 98% pure, even more preferably at least 99%, most preferably at least 99.5% pure, and even most preferably 100% pure by weight of the total polypeptide material present in the preparation.

The polypeptides of the present invention are preferably in a substantially pure form. In particular, it is preferred that the polypeptides are in "essentially pure form", *i.e.*, that the

polypeptide preparation is essentially free of other polypeptide material with which it is natively associated. This can be accomplished, for example, by preparing the polypeptide by means of well-known recombinant methods or by classical purification methods.

Herein, the term "substantially pure polypeptide" is synonymous with the terms "isolated polypeptide" and "polypeptide in isolated form."

Identity: The relatedness between two amino acid sequences or between two nucleotide sequences is described by the parameter "identity".

For purposes of the present invention, the degree of identity between two amino acid sequences is determined by using the program FASTA included in version 2.0x of the FASTA program package (see W. R. Pearson and D. J. Lipman (1988), "Improved Tools for Biological Sequence Analysis", PNAS 85:2444-2448; and W. R. Pearson (1990) "Rapid and Sensitive Sequence Comparison with FASTP and FASTA", Methods in Enzymology 183:63-98). The scoring matrix used was BLOSUM50, gap penalty was -12, and gap extension penalty was -2.

The degree of identity between two nucleotide sequences is determined using the same algorithm and software package as described above. The scoring matrix used was the identity matrix, gap penalty was -16, and gap extension penalty was -4.

Alternatively, an alignment of two amino acid sequences is determined by using the Needle program from the EMBOSS package (http://emboss.org) version 2.8.0. The Needle program implements the global alignment algorithm described in Needleman, S. B. and Wunsch, C. D. (1970) J. Mol. Biol. 48, 443-453. The substitution matrix used is BLOSUM62, gap opening penalty is 10, and gap extension penalty is 0.5.

The degree of identity between an amino acid sequence of the present invention ("invention sequence"; e.g. amino acids 1 to 49 of SEQ ID NO:2) and a different amino acid sequence ("foreign sequence") is calculated as the number of exact matches in the overlap of an alignment of the two sequences, divided by the length of the "invention sequence" or the length of the "foreign sequence", whichever is the shortest. The result is expressed in percent identity.

An exact match occurs when the "invention sequence" and the "foreign sequence" have identical amino acid residues in the same positions of the overlap. The length of a sequence is the number of amino acid residues in the sequence (e.g. the length of amino acids 1 to 49 of SEQ ID NO:2 is 49).

Polypeptide Fragment: The term "polypeptide fragment" is defined herein as a polypeptide having one or more amino acids deleted from the amino and/or carboxyl terminus of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28, or a homologous sequence thereof, wherein the fragment has antimicrobial activity.

Subsequence: The term "subsequence" is defined herein as a nucleotide sequence having one or more nucleotides deleted from the 5' and/or 3' end of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27 or a homologous sequence thereof, wherein the subsequence encodes a polypeptide fragment having antimicrobial activity.

Allelic variant: The term "allelic variant" denotes herein any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequences. An allelic variant of a polypeptide is a polypeptide encoded by an allelic variant of a gene.

Substantially pure polynucleotide: The term "substantially pure polynucleotide" as used herein refers to a polynucleotide preparation free of other extraneous or unwanted nucleotides and in a form suitable for use within genetically engineered protein production systems. Thus, a substantially pure polynucleotide contains at most 10%, preferably at most 8%, more preferably at most 6%, more preferably at most 5%, more preferably at most 4%, more preferably at most 3%, even more preferably at most 2%, most preferably at most 1%, and even most preferably at most 0.5% by weight of other polynucleotide material with which it is natively associated. A substantially pure polynucleotide may, however, include naturally occurring 5' and 3' untranslated regions, such as promoters and terminators. It is preferred that the substantially pure polynucleotide is at least 90% pure, preferably at least 92% pure, more preferably at least 94% pure, more preferably at least 95% pure, more preferably at least 96% pure, more preferably at least 97% pure, even more preferably at least 98% pure, most preferably at least 99%, and even most preferably at least 99.5% pure by weight. The polynucleotides of the present invention are preferably in a substantially pure form. In particular, it is preferred that the polynucleotides disclosed herein are in "essentially pure form", *i.e.*, that the polynucleotide preparation is essentially free of other polynucleotide material with which it is natively associated. Herein, the term "substantially pure polynucleotide" is synonymous with the terms "isolated polynucleotide" and "polynucleotide in isolated form." The polynucleotides may be of genomic, cDNA, RNA, semisynthetic, synthetic origin, or any combinations thereof.

cDNA: The term "cDNA" is defined herein as a DNA molecule which can be prepared by reverse transcription from a mature, spliced, mRNA molecule obtained from a eukaryotic cell. cDNA lacks intron sequences that are usually present in the corresponding genomic DNA. The initial, primary RNA transcript is a precursor to mRNA which is processed through a series of steps before appearing as mature spliced mRNA. These steps include the removal of intron

sequences by a process called splicing. cDNA derived from mRNA lacks, therefore, any intron sequences.

Nucleic acid construct: The term "nucleic acid construct" as used herein refers to a nucleic acid molecule, either single- or double-stranded, which is isolated from a naturally occurring gene or which is modified to contain segments of nucleic acids in a manner that would not otherwise exist in nature. The term nucleic acid construct is synonymous with the term "expression cassette" when the nucleic acid construct contains the control sequences required for expression of a coding sequence of the present invention.

Control sequence: The term "control sequences" is defined herein to include all components, which are necessary or advantageous for the expression of a polynucleotide encoding a polypeptide of the present invention. Each control sequence may be native or foreign to the nucleotide sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleotide sequence encoding a polypeptide.

Operably linked: The term "operably linked" denotes herein a configuration in which a control sequence is placed at an appropriate position relative to the coding sequence of the polynucleotide sequence such that the control sequence directs the expression of the coding sequence of a polypeptide.

Coding sequence: When used herein the term "coding sequence" means a nucleotide sequence, which directly specifies the amino acid sequence of its protein product. The boundaries of the coding sequence are generally determined by an open reading frame, which usually begins with the ATG start codon or alternative start codons such as GTG and TTG. The coding sequence may a DNA, cDNA, or recombinant nucleotide sequence.

Expression: The term "expression" includes any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

Expression vector: The term "expression vector" is defined herein as a linear or circular DNA molecule that comprises a polynucleotide encoding a polypeptide of the invention, and which is operably linked to additional nucleotides that provide for its expression.

Host cell: The term "host cell", as used herein, includes any cell type which is susceptible to transformation, transfection, transduction, and the like with a nucleic acid construct comprising a polynucleotide of the present invention.

Modification: The term "modification" means herein any chemical modification of the

polypeptide consisting of

amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;

amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;

amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;

amino acids 1 to 45 of SEQ ID NO:22;

amino acids 1 to 49 of SEQ ID NO:24;

amino acids 1 to 44 of SEQ ID NO:26; or

amino acids 1 to 59 of SEQ ID NO:28; as well as genetic manipulation of the DNA encoding that polypeptide. The modification(s) can be substitution(s), deletion(s) and/or insertions(s) of the amino acid(s) as well as replacement(s) of amino acid side chain(s); or use of unnatural amino acids with similar characteristics in the amino acid sequence. In particular the modification(s) can be amidations, such as amidation of the C-terminus.

Artificial variant: When used herein, the term "artificial variant" means a polypeptide having antimicrobial activity produced by an organism expressing a modified nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27. The modified nucleotide sequence is obtained through human intervention by modification of the nucleotide sequence disclosed in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27.

DETAILED DESCRIPTION OF THE INVENTION

Polypeptides Having Antimicrobial Activity

In a first aspect, the present invention relates to polypeptides having antimicrobial activity, which comprise an amino acid sequence represented by: C-x(3)-C-x(7,9)-C-C; C-C-x(8)-C-x-C; or C-x-C-x(8,11)-C.

In an embodiment, the polypeptides comprise an amino acid sequence represented by:

C-x(3)-C-x(7,9)-C-C and C-C-x(8)-C-x-C; or

C-C-x(8)-C-x-C and C-x-C-x(8,11)-C; or

C-x(3)-C-x(7,9)-C-C and C-x-C-x(8,11)-C; or

C-x(3)-C-x(7,9)-C-C and C-C-x(8)-C-x-C and C-x-C-x(8,11)-C.

The patterns C-x(3)-C-x(7,9)-C-C; C-C-x(8)-C-x-C; and C-x-C-x(8,11)-C shall be interpreted using the PROSITE (www.expasy.org/prosite/) pattern definition format:

- the standard IUPAC one-letter codes for the amino acids are used;

- the symbol "x" is used for a position where any amino acid is accepted;
- each element in a pattern is separated from its neighbor by a "-"; and
- repetition of an element of the pattern is indicated by following that element with a numerical value or a numerical range between parenthesis. For example: x(3) corresponds to x-x-x, x(2,4) corresponds to x-x or x-x-x or x-x-x-x.

For more information on use of the PROSITE methodology, please refer to Sigrist et al. PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinform.* 3:265-274 (2002).

In a second aspect, which may be an embodiment of the first aspect, the invention relates to polypeptides comprising an amino acid sequence which has a degree of identity to amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;
amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;
amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;
amino acids 1 to 45 of SEQ ID NO:22;
amino acids 1 to 49 of SEQ ID NO:24;
amino acids 1 to 44 of SEQ ID NO:26; or
amino acids 1 to 59 of SEQ ID NO:28 (*i.e.*, the mature polypeptides) of at least 60%, preferably at least 65%, more preferably at least 70%, more preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, and even most preferably at least 97%, which have antimicrobial activity (hereinafter "homologous polypeptides"). In a preferred aspect, the homologous polypeptides have an amino acid sequence which differs by ten amino acids, preferably by five amino acids, more preferably by four amino acids, even more preferably by three amino acids, most preferably by two amino acids, and even most preferably by one amino acid from
amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;
amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;
amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;
amino acids 1 to 45 of SEQ ID NO:22;
amino acids 1 to 49 of SEQ ID NO:24;
amino acids 1 to 44 of SEQ ID NO:26; or
amino acids 1 to 59 of SEQ ID NO:28.

A polypeptide of the present invention preferably comprises the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28, or an allelic variant thereof; or a fragment thereof

that has antimicrobial activity. In a preferred aspect, a polypeptide comprises the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28. In another preferred aspect, a polypeptide comprises amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;

amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;

amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;

amino acids 1 to 45 of SEQ ID NO:22;

amino acids 1 to 49 of SEQ ID NO:24;

amino acids 1 to 44 of SEQ ID NO:26; or

amino acids 1 to 59 of SEQ ID NO:28, or an allelic variant thereof; or a fragment thereof that has antimicrobial activity. In another preferred aspect, a polypeptide comprises

amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;

amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;

amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;

amino acids 1 to 45 of SEQ ID NO:22;

amino acids 1 to 49 of SEQ ID NO:24;

amino acids 1 to 44 of SEQ ID NO:26; or

amino acids 1 to 59 of SEQ ID NO:28.

In another preferred aspect, a polypeptide consists of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28, or an allelic variant thereof; or a fragment thereof that has antimicrobial activity. In another preferred aspect, a polypeptide consists of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28. In another preferred aspect, a polypeptide consists of amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;

amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;

amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;

amino acids 1 to 45 of SEQ ID NO:22;

amino acids 1 to 49 of SEQ ID NO:24;

amino acids 1 to 44 of SEQ ID NO:26; or

amino acids 1 to 59 of SEQ ID NO:28 or an allelic variant thereof; or a fragment thereof that

has antimicrobial activity. In another preferred aspect, a polypeptide consists of
amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;
amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;
amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;
amino acids 1 to 45 of SEQ ID NO:22;
amino acids 1 to 49 of SEQ ID NO:24;
amino acids 1 to 44 of SEQ ID NO:26; or
amino acids 1 to 59 of SEQ ID NO:28.

In a second aspect, the present invention relates to isolated polypeptides having antimicrobial activity which are encoded by polynucleotides which hybridize under very low stringency conditions, preferably low stringency conditions, more preferably medium stringency conditions, more preferably medium-high stringency conditions, even more preferably high stringency conditions, and most preferably very high stringency conditions with (i) nucleotides 151 to 297 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9;
nucleotides 118 to 255 of SEQ ID NO:11, SEQ ID NO:13;
nucleotides 112 to 255 of SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19;
nucleotides 118 to 252 of SEQ ID NO:21;
nucleotides 151 to 297 of SEQ ID NO:23;
nucleotides 121 to 252 of SEQ ID NO:25; or
nucleotides 145 to 321 of SEQ ID NO:27;
(ii) the cDNA sequence contained in nucleotides 1 to 297 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9;
nucleotides 1 to 255 of SEQ ID NO:11, SEQ ID NO:13;
nucleotides 1 to 255 of SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19;
nucleotides 1 to 252 of SEQ ID NO:21;
nucleotides 1 to 297 of SEQ ID NO:23;
nucleotides 1 to 252 of SEQ ID NO:25; or
nucleotides 1 to 321 of SEQ ID NO:27;
(iii) a subsequence of (i) or (ii), or (iv) a complementary strand of (i), (ii), or (iii) (J. Sambrook, E.F. Fritsch, and T. Maniatus, 1989, *Molecular Cloning, A Laboratory Manual*, 2d edition, Cold Spring Harbor, New York). A subsequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27 contains at least 100 contiguous nucleotides or preferably at least 200 continguous nucleotides. Moreover, the subsequence may encode a polypeptide fragment which has antimicrobial

activity.

The nucleotide sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27, or a subsequence thereof, as well as the amino acid sequences of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28, or a fragment thereof, may be used to design a nucleic acid probe to identify and clone DNA encoding polypeptides having antimicrobial activity from strains of different genera or species according to methods well known in the art. In particular, such probes can be used for hybridization with the genomic or cDNA of the genus or species of interest, following standard Southern blotting procedures, in order to identify and isolate the corresponding gene therein. Such probes can be considerably shorter than the entire sequence, but should be at least 14, preferably at least 25, more preferably at least 35, and most preferably at least 70 nucleotides in length. It is, however, preferred that the nucleic acid probe is at least 100 nucleotides in length. For example, the nucleic acid probe may be at least 200 nucleotides, preferably at least 250 nucleotides. Both DNA and RNA probes can be used. The probes are typically labeled for detecting the corresponding gene (for example, with ^{32}P, ^{3}H, ^{35}S, biotin, or avidin). Such probes are encompassed by the present invention.

A genomic DNA or cDNA library prepared from such other organisms may, therefore, be screened for DNA which hybridizes with the probes described above and which encodes a polypeptide having antimicrobial activity. Genomic or other DNA from such other organisms may be separated by agarose or polyacrylamide gel electrophoresis, or other separation techniques. DNA from the libraries or the separated DNA may be transferred to and immobilized on nitrocellulose or other suitable carrier material. In order to identify a clone or DNA which is homologous with SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27, or a subsequence thereof, the carrier material is used in a Southern blot.

For purposes of the present invention, hybridization indicates that the nucleotide sequence hybridizes to a labeled nucleic acid probe corresponding to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27, its complementary strand, or a subsequence thereof, under very low to very high stringency conditions. Molecules to which the nucleic acid probe hybridizes under these conditions can be detected using X-ray film.

In a preferred aspect, the nucleic acid probe is a polynucleotide sequence which

encodes the polypeptide of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28, or a subsequence thereof. In another preferred aspect, the nucleic acid probe is SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27. In another preferred aspect, the nucleic acid probe is the mature polypeptide coding region of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27.

For long probes of at least 100 nucleotides in length, very low to very high stringency conditions are defined as prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 µg/ml sheared and denatured salmon sperm DNA, and either 25% formamide for very low and low stringencies, 35% formamide for medium and medium-high stringencies, or 50% formamide for high and very high stringencies, following standard Southern blotting procedures for 12 to 24 hours optimally.

For long probes of at least 100 nucleotides in length, the carrier material is finally washed three times each for 15 minutes using 2X SSC, 0.2% SDS preferably at least at 45°C (very low stringency), more preferably at least at 50°C (low stringency), more preferably at least at 55°C (medium stringency), more preferably at least at 60°C (medium-high stringency), even more preferably at least at 65°C (high stringency), and most preferably at least at 70°C (very high stringency).

For short probes which are about 15 nucleotides to about 70 nucleotides in length, stringency conditions are defined as prehybridization, hybridization, and washing post-hybridization at about 5°C to about 10°C below the calculated T_m using the calculation according to Bolton and McCarthy (1962, *Proceedings of the National Academy of Sciences USA* 48:1390) in 0.9 M NaCl, 0.09 M Tris-HCl pH 7.6, 6 mM EDTA, 0.5% NP-40, 1X Denhardt's solution, 1 mM sodium pyrophosphate, 1 mM sodium monobasic phosphate, 0.1 mM ATP, and 0.2 mg of yeast RNA per ml following standard Southern blotting procedures.

For short probes which are about 15 nucleotides to about 70 nucleotides in length, the carrier material is washed once in 6X SCC plus 0.1% SDS for 15 minutes and twice each for 15 minutes using 6X SSC at 5°C to 10°C below the calculated T_m.

In a third aspect, the present invention relates to artificial variants comprising a conservative substitution, deletion, and/or insertion of one or more amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28, or the mature polypeptides thereof. Preferably, amino acid

changes are of a minor nature, that is conservative amino acid substitutions or insertions that do not significantly affect the folding and/or activity of the protein; small deletions, typically of one to about 30 amino acids; small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue; a small linker peptide of up to about 20-25 residues; or a small extension that facilitates purification by changing net charge or another function, such as a poly-histidine tract, an antigenic epitope or a binding domain.

Examples of conservative substitutions are within the group of basic amino acids (arginine, lysine and histidine), acidic amino acids (glutamic acid and aspartic acid), polar amino acids (glutamine and asparagine), hydrophobic amino acids (leucine, isoleucine and valine), aromatic amino acids (phenylalanine, tryptophan and tyrosine), and small amino acids (glycine, alanine, serine, threonine and methionine). Amino acid substitutions which do not generally alter specific activity are known in the art and are described, for example, by H. Neurath and R.L. Hill, 1979, *In*, *The Proteins*, Academic Press, New York. The most commonly occurring exchanges are Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, and Asp/Gly.

In addition to the 20 standard amino acids, non-standard amino acids (such as 4-hydroxyproline, 6-*N*-methyl lysine, 2-aminoisobutyric acid, isovaline, and alpha-methyl serine) may be substituted for amino acid residues of a wild-type polypeptide. A limited number of non-conservative amino acids, amino acids that are not encoded by the genetic code, and unnatural amino acids may be substituted for amino acid residues. "Unnatural amino acids" have been modified after protein synthesis, and/or have a chemical structure in their side chain(s) different from that of the standard amino acids. Unnatural amino acids can be chemically synthesized, and preferably, are commercially available, and include pipecolic acid, thiazolidine carboxylic acid, dehydroproline, 3- and 4-methylproline, and 3,3-dimethylproline.

Alternatively, the amino acid changes are of such a nature that the physico-chemical properties of the polypeptides are altered. For example, amino acid changes may improve the thermal stability of the polypeptide, alter the substrate specificity, change the pH optimum, and the like.

Essential amino acids in the parent polypeptide can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, 1989, *Science* 244: 1081-1085). In the latter technique, single alanine mutations are introduced at every residue in the molecule, and the resultant mutant molecules are tested for biological activity (*i.e.*, antimicrobial activity) to identify amino acid residues that are critical to the activity of the molecule. See also, Hilton *et al.*, 1996, *J. Biol. Chem.* 271: 4699-4708. The active site of the enzyme or other biological interaction can also be determined by physical analysis of structure, as determined by such

techniques as nuclear magnetic resonance, crystallography, electron diffraction, or photoaffinity labeling, in conjunction with mutation of putative contact site amino acids. See, for example, de Vos *et al.*, 1992, *Science* 255: 306-312; Smith *et al.*, 1992, *J. Mol. Biol.* 224: 899-904; Wlodaver *et al.*, 1992, *FEBS Lett.* 309:59-64. The identities of essential amino acids can also be inferred from analysis of identities with polypeptides which are related to a polypeptide according to the invention.

Single or multiple amino acid substitutions can be made and tested using known methods of mutagenesis, recombination, and/or shuffling, followed by a relevant screening procedure, such as those disclosed by Reidhaar-Olson and Sauer, 1988, *Science* 241: 53-57; Bowie and Sauer, 1989, *Proc. Natl. Acad. Sci. USA* 86: 2152-2156; WO 95/17413; or WO 95/22625. Other methods that can be used include error-prone PCR, phage display (e.g., Lowman *et al.*, 1991, *Biochem.* 30:10832-10837; U.S. Patent No. 5,223,409; WO 92/06204), and region-directed mutagenesis (Derbyshire *et al.*, 1986, *Gene* 46:145; Ner *et al.*, 1988, *DNA* 7:127).

Mutagenesis/shuffling methods can be combined with high-throughput, automated screening methods to detect activity of cloned, mutagenized polypeptides expressed by host cells. Mutagenized DNA molecules that encode active polypeptides can be recovered from the host cells and rapidly sequenced using standard methods in the art. These methods allow the rapid determination of the importance of individual amino acid residues in a polypeptide of interest, and can be applied to polypeptides of unknown structure.

The total number of amino acid substitutions, deletions and/or insertions of
amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;
amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;
amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;
amino acids 1 to 45 of SEQ ID NO:22;
amino acids 1 to 49 of SEQ ID NO:24;
amino acids 1 to 44 of SEQ ID NO:26; or
amino acids 1 to 59 of SEQ ID NO:28 is 10, preferably 9, more preferably 8, more preferably 7, more preferably at most 6, more preferably at most 5, more preferably 4, even more preferably 3, most preferably 2, and even most preferably 1.

In a preferred embodiment, the polypeptides of the invention are defensin polypeptides.

N-Terminal Extension

An N-terminal extension of the polypeptides of the invention may suitably consist of from 1 to 50 amino acids, preferably 2-20 amino acids, especially 3-15 amino acids. In one embodiment N-terminal peptide extension does not contain an Arg (R). In another

embodiment the N-terminal extension comprises a kex2 or kex2-like cleavage site as will be defined further below. In a preferred embodiment the N-terminal extension is a peptide, comprising at least two Glu (E) and/or Asp (D) amino acid residues, such as an N-terminal extension comprising one of the following sequences: EAE, EE, DE and DD.

Kex2 Sites

Kex2 sites (see, e.g., Methods in Enzymology Vol 185, ed. D. Goeddel, Academic Press Inc. (1990), San Diego, CA, "Gene Expression Technology") and kex2-like sites are di-basic recognition sites (i.e., cleavage sites) found between the pro-peptide encoding region and the mature region of some proteins.

Insertion of a kex2 site or a kex2-like site have in certain cases been shown to improve correct endopeptidase processing at the pro-peptide cleavage site resulting in increased protein secretion levels.

In the context of the invention insertion of a kex2 or kex2-like site result in the possibility to obtain cleavage at a certain position in the N-terminal extension resulting in an antimicrobial polypeptide being extended in comparison to the mature polypeptide shown as
amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;
amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;
amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;
amino acids 1 to 45 of SEQ ID NO:22;
amino acids 1 to 49 of SEQ ID NO:24;
amino acids 1 to 44 of SEQ ID NO:26; or
amino acids 1 to 59 of SEQ ID NO:28.

Fused Polypeptides

The polypeptides of the present invention also include fused polypeptides or cleavable fusion polypeptides in which another polypeptide is fused at the N-terminus or the C-terminus of the polypeptide of the invention or a fragment thereof. A fused polypeptide is produced by fusing a nucleotide sequence (or a portion thereof) encoding another polypeptide to a nucleotide sequence (or a portion thereof) of the present invention. Techniques for producing fusion polypeptides are known in the art, and include ligating the coding sequences encoding the polypeptides so that they are in frame and that expression of the fused polypeptide is under control of the same promoter(s) and terminator.

Sources of Polypeptides Having Antimicrobial Activity

A polypeptide of the present invention may be obtained from microorganisms of any

genus. For purposes of the present invention, the term "obtained from" as used herein in connection with a given source shall mean that the polypeptide encoded by a nucleotide sequence is produced by the source or by a strain in which the nucleotide sequence from the source has been inserted. In a preferred aspect, the polypeptide obtained from a given source is secreted extracellularly.

A polypeptide of the present invention may be a bacterial polypeptide. For example, the polypeptide may be a gram positive bacterial polypeptide such as a *Bacillus* polypeptide, *e.g.*, a *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis*, or *Bacillus thuringiensis* polypeptide; or a *Streptomyces* polypeptide, *e.g.*, a *Streptomyces lividans* or *Streptomyces murinus* polypeptide; or a gram negative bacterial polypeptide, *e.g.*, an *E. coli* or a *Pseudomonas* sp. polypeptide.

A polypeptide of the present invention may also be a fungal polypeptide, and more preferably a yeast polypeptide such as a *Candida*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* polypeptide; or more preferably a filamentous fungal polypeptide such as an *Acremonium*, *Aspergillus*, *Aureobasidium*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, or *Trichoderma* polypeptide.

In a preferred aspect, the polypeptide is a *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, or *Saccharomyces oviformis* polypeptide having antimicrobial activity.

In another preferred aspect, the polypeptide is an *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, *Fusarium venenatum*, *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium purpurogenum*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, or *Trichoderma viride* polypeptide.

In another preferred aspect, the polypeptide is an *Arenicola marina* polypeptide, *e.g.*, the polypeptide of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28.

It will be understood that for the aforementioned species, the invention encompasses both the perfect and imperfect states, and other taxonomic equivalents, *e.g.*, anamorphs, regardless of the species name by which they are known. Those skilled in the art will readily recognize the identity of appropriate equivalents.

Strains of these species are readily accessible to the public in a number of culture collections, such as the American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

Furthermore, such polypeptides may be identified and obtained from other sources including microorganisms isolated from nature (*e.g.*, soil, composts, water, etc.) using the above-mentioned probes. Techniques for isolating microorganisms from natural habitats are well known in the art. The polynucleotide may then be obtained by similarly screening a genomic or cDNA library of another microorganism. Once a polynucleotide sequence encoding a polypeptide has been detected with the probe(s), the polynucleotide can be isolated or cloned by utilizing techniques which are well known to those of ordinary skill in the art (see, *e.g.*, Sambrook *et al.*, 1989, *supra*).

Polypeptides of the present invention also include fused polypeptides or cleavable fusion polypeptides in which another polypeptide is fused at the N-terminus or the C-terminus of the polypeptide or fragment thereof. A fused polypeptide is produced by fusing a nucleotide sequence (or a portion thereof) encoding another polypeptide to a nucleotide sequence (or a portion thereof) of the present invention. Techniques for producing fusion polypeptides are known in the art, and include ligating the coding sequences encoding the polypeptides so that they are in frame and that expression of the fused polypeptide is under control of the same promoter(s) and terminator.

Polynucleotides

The present invention also relates to isolated polynucleotides having a nucleotide sequence which encode a polypeptide of the present invention. In a preferred aspect, the nucleotide sequence is set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27. In another preferred aspect, the nucleotide sequence is the mature polypeptide coding region of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27. The present invention also encompasses nucleotide sequences which encode a polypeptide having the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4,

SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28, or the mature polypeptide thereof, which differ from SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27 by virtue of the degeneracy of the genetic code. The present invention also relates to subsequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27 which encode fragments of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28, that have antimicrobial activity.

The present invention also relates to mutant polunucleotides comprising at least one mutation in the mature polypeptide coding sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27, in which the mutant nucleotide sequence encodes a polypeptide which consists of

amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;

amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;

amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;

amino acids 1 to 45 of SEQ ID NO:22;

amino acids 1 to 49 of SEQ ID NO:24;

amino acids 1 to 44 of SEQ ID NO:26; or

amino acids 1 to 59 of SEQ ID NO:28.

The techniques used to isolate or clone a polynucleotide encoding a polypeptide are known in the art and include isolation from genomic DNA, preparation from cDNA, or a combination thereof. The cloning of the polynucleotides of the present invention from such genomic DNA can be effected, *e.g.*, by using the well known polymerase chain reaction (PCR) or antibody screening of expression libraries to detect cloned DNA fragments with shared structural features. See, *e.g.*, Innis *et al.*, 1990, *PCR: A Guide to Methods and Application*, Academic Press, New York. Other nucleic acid amplification procedures such as ligase chain reaction (LCR), ligated activated transcription (LAT) and nucleotide sequence-based amplification (NASBA) may be used. The polynucleotides may be cloned from a strain of *Aspergillus*, or another or related organism and thus, for example, may be an allelic or species variant of the polypeptide encoding region of the nucleotide sequence.

The present invention also relates to polynucleotides having nucleotide sequences which

have a degree of identity to the mature polypeptide coding sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27 (*i.e.*, nucleotides 151 to 297 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9; nucleotides 118 to 255 of SEQ ID NO:11, SEQ ID NO:13; nucleotides 112 to 255 of SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19; nucleotides 118 to 252 of SEQ ID NO:21; nucleotides 151 to 297 of SEQ ID NO:23; nucleotides 121 to 252 of SEQ ID NO:25; or nucleotides 145 to 321 of SEQ ID NO:27) of at least 60%, preferably at least 65%, more preferably at least 70%, more preferably at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, even more preferably at least 95%, and most preferably at least 97% identity, which encode an active polypeptide.

Modification of a nucleotide sequence encoding a polypeptide of the present invention may be necessary for the synthesis of polypeptides substantially similar to the polypeptide. The term "substantially similar" to the polypeptide refers to non-naturally occurring forms of the polypeptide. These polypeptides may differ in some engineered way from the polypeptide isolated from its native source, *e.g.*, artificial variants that differ in specific activity, thermostability, pH optimum, or the like. The variant sequence may be constructed on the basis of the nucleotide sequence presented as the polypeptide encoding region of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27, *e.g.*, a subsequence thereof, and/or by introduction of nucleotide substitutions which do not give rise to another amino acid sequence of the polypeptide encoded by the nucleotide sequence, but which correspond to the codon usage of the host organism intended for production of the enzyme, or by introduction of nucleotide substitutions which may give rise to a different amino acid sequence. For a general description of nucleotide substitution, see, *e.g.*, Ford *et al.*, 1991, *Protein Expression and Purification* 2: 95-107.

It will be apparent to those skilled in the art that such substitutions can be made outside the regions critical to the function of the molecule and still result in an active polypeptide. Amino acid residues essential to the activity of the polypeptide encoded by an isolated polynucleotide of the invention, and therefore preferably not subject to substitution, may be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (see, *e.g.*, Cunningham and Wells, 1989, *Science* 244: 1081-1085). In the latter technique, mutations are introduced at every positively charged residue in the molecule, and the resultant mutant molecules are tested for antimicrobial activity to identify amino acid residues that are critical to the activity of the molecule. Sites of substrate-enzyme interaction can also be determined by analysis of the three-dimensional structure as

determined by such techniques as nuclear magnetic resonance analysis, crystallography or photoaffinity labelling (see, *e.g.*, de Vos *et al.*, 1992, *Science* 255: 306-312; Smith *et al.*, 1992, *Journal of Molecular Biology* 224: 899-904; Wlodaver *et al.*, 1992, *FEBS Letters* 309: 59-64).

The present invention also relates to isolated polynucleotides encoding a polypeptide of the present invention, which hybridize under low stringency conditions, preferably medium stringency conditions, more preferably medium-high stringency conditions, even more preferably high stringency conditions, and most preferably very high stringency conditions with (i) nucleotides 151 to 297 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9;

nucleotides 118 to 255 of SEQ ID NO:11, SEQ ID NO:13;

nucleotides 112 to 255 of SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19;

nucleotides 118 to 252 of SEQ ID NO:21;

nucleotides 151 to 297 of SEQ ID NO:23;

nucleotides 121 to 252 of SEQ ID NO:25; or

nucleotides 145 to 321 of SEQ ID NO:27;

(ii) the cDNA sequence contained in nucleotides 1 to 297 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9;

nucleotides 1 to 255 of SEQ ID NO:11, SEQ ID NO:13;

nucleotides 1 to 255 of SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19;

nucleotides 1 to 252 of SEQ ID NO:21;

nucleotides 1 to 297 of SEQ ID NO:23;

nucleotides 1 to 252 of SEQ ID NO:25; or

nucleotides 1 to 321 of SEQ ID NO:27; or

(iii) a complementary strand of (i) or (ii); or allelic variants and subsequences thereof (Sambrook *et al.*, 1989, *supra*), as defined herein.

The present invention also relates to isolated polynucleotides obtained by (a) hybridizing a population of DNA under low, medium, medium-high, high, or very high stringency conditions with (i) nucleotides 151 to 297 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9;

nucleotides 118 to 255 of SEQ ID NO:11, SEQ ID NO:13;

nucleotides 112 to 255 of SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19;

nucleotides 118 to 252 of SEQ ID NO:21;

nucleotides 151 to 297 of SEQ ID NO:23;

nucleotides 121 to 252 of SEQ ID NO:25; or

nucleotides 145 to 321 of SEQ ID NO:27;

(ii) the cDNA sequence contained in nucleotides 1 to 297 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9;

nucleotides 1 to 255 of SEQ ID NO:11, SEQ ID NO:13;

nucleotides 1 to 255 of SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19;

nucleotides 1 to 252 of SEQ ID NO:21;

nucleotides 1 to 297 of SEQ ID NO:23;

nucleotides 1 to 252 of SEQ ID NO:25; or

nucleotides 1 to 321 of SEQ ID NO:27; or

(iii) a complementary strand of (i) or (ii); and (b) isolating the hybridizing polynucleotide, which encodes a polypeptide having antimicrobial activity.

Nucleic Acid Constructs

The present invention also relates to nucleic acid constructs comprising an isolated polynucleotide of the present invention operably linked to one or more control sequences which direct the expression of the coding sequence in a suitable host cell under conditions compatible with the control sequences.

An isolated polynucleotide encoding a polypeptide of the present invention may be manipulated in a variety of ways to provide for expression of the polypeptide. Manipulation of the polynucleotide's sequence prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotide sequences utilizing recombinant DNA methods are well known in the art.

The control sequence may be an appropriate promoter sequence, a nucleotide sequence which is recognized by a host cell for expression of a polynucleotide encoding a polypeptide of the present invention. The promoter sequence contains transcriptional control sequences which mediate the expression of the polypeptide. The promoter may be any nucleotide sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

Examples of suitable promoters for directing the transcription of the nucleic acid constructs of the present invention, especially in a bacterial host cell, are the promoters obtained from the *E. coli lac* operon, *Streptomyces coelicolor* agarase gene (*dagA*), *Bacillus subtilis* levansucrase gene (*sacB*), *Bacillus licheniformis* alpha-amylase gene (*amyL*), *Bacillus stearothermophilus* maltogenic amylase gene (*amyM*), *Bacillus amyloliquefaciens* alpha-amylase gene (*amyQ*), *Bacillus licheniformis* penicillinase gene (*penP*), *Bacillus subtilis xylA* and *xylB* genes, and prokaryotic beta-lactamase gene (Villa-Kamaroff *et al.*, 1978, *Proceedings of the National Academy of Sciences USA* 75: 3727-3731), as well as the *tac* promoter (DeBoer *et al.*, 1983, *Proceedings of the National Academy of Sciences USA* 80: 21-25). Further promoters are described in "Useful proteins from recombinant bacteria" in *Scientific American*, 1980, 242: 74-94; and in Sambrook *et al.*, 1989, *supra*.

Examples of suitable promoters for directing the transcription of the nucleic acid constructs of the present invention in a filamentous fungal host cell are promoters obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (*glaA*), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, *Fusarium venenatum* amyloglucosidase (WO 00/56900), *Fusarium venenatum* Daria (WO 00/56900), *Fusarium venenatum* Quinn (WO 00/56900), *Fusarium oxysporum* trypsin-like protease (WO 96/00787), *Trichoderma reesei* beta-glucosidase, *Trichoderma reesei* cellobiohydrolase I, *Trichoderma reesei* endoglucanase I, *Trichoderma reesei* endoglucanase II, *Trichoderma reesei* endoglucanase III, *Trichoderma reesei* endoglucanase IV, *Trichoderma reesei* endoglucanase V, *Trichoderma reesei* xylanase I, *Trichoderma reesei* xylanase II, *Trichoderma reesei* beta-xylosidase, as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase); and mutant, truncated, and hybrid promoters thereof.

In a yeast host, useful promoters are obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* galactokinase (GAL1), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH1,ADH2/GAP), *Saccharomyces cerevisiae* triose phosphate isomerase (TPI), *Saccharomyces cerevisiae* metallothionine (CUP1), and *Saccharomyces cerevisiae* 3-phosphoglycerate kinase. Other useful promoters for yeast host cells are described by Romanos *et al.*, 1992, *Yeast* 8: 423-488.

The control sequence may also be a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleotide sequence encoding the polypeptide. Any terminator which is functional in the host cell of choice may be used in the present invention.

Preferred terminators for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease.

Preferred terminators for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYC1), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are described by Romanos *et al.*, 1992, *supra*.

The control sequence may also be a suitable leader sequence, a nontranslated region of an mRNA which is important for translation by the host cell. The leader sequence is operably

linked to the 5' terminus of the nucleotide sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice may be used in the present invention.

Preferred leaders for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase.

Suitable leaders for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP).

The control sequence may also be a polyadenylation sequence, a sequence operably linked to the 3' terminus of the nucleotide sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice may be used in the present invention.

Preferred polyadenylation sequences for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase.

Useful polyadenylation sequences for yeast host cells are described by Guo and Sherman, 1995, *Molecular Cellular Biology* 15: 5983-5990.

The control sequence may also be a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleotide sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region which encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region which is foreign to the coding sequence. The foreign signal peptide coding region may be required where the coding sequence does not naturally contain a signal peptide coding region. Alternatively, the foreign signal peptide coding region may simply replace the natural signal peptide coding region in order to enhance secretion of the polypeptide. However, any signal peptide coding region which directs the expressed polypeptide into the secretory pathway of a host cell of choice may be used in the present invention.

Effective signal peptide coding regions for bacterial host cells are the signal peptide coding regions obtained from the genes for *Bacillus* NCIB 11837 maltogenic amylase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* beta-lactamase, *Bacillus stearothermophilus* neutral proteases (*nprT*, *nprS*, *nprM*), and *Bacillus subtilis prsA*. Further signal peptides are described by Simonen and Palva, 1993, *Microbiological Reviews* 57: 109-137.

Effective signal peptide coding regions for filamentous fungal host cells are the signal peptide coding regions obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, and *Humicola lanuginosa* lipase.

In a preferred aspect, the signal peptide coding region is nucleotides 1 to 57 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 or SEQ ID NO:9 which encode amino acids -50 to -32 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10; nucleotides 1 to 63 of SEQ ID NO:11 or SEQ ID NO:13 which encode amino acids -39 to -19 of SEQ ID NO:12 or SEQ ID NO:14; nucleotides 1 to 63 of SEQ ID NO:15, SEQ ID NO:17 or SEQ ID NO:19 which encode amino acids -37 to -17 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20; nucleotides 1 to 63 of SEQ ID NO:21 which encode amino acids -39 to -19 of SEQ ID NO:22; nucleotides 1 to 57 of SEQ ID NO:23 which encode amino acids -50 to -32 of SEQ ID NO:24; nucleotides 1 to 63 of SEQ ID NO:25 which encode amino acids -40 to -20 of SEQ ID NO:26; or nucleotides 1 to 66 of SEQ ID NO:27 which encode amino acids -48 to -27 of SEQ ID NO:28.

Useful signal peptides for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase. Other useful signal peptide coding regions are described by Romanos *et al.*, 1992, *supra.*

The control sequence may also be a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to a mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region may be obtained from the genes for *Bacillus subtilis* alkaline protease (*aprE*), *Bacillus subtilis* neutral protease (*nprT*), *Saccharomyces cerevisiae* alpha-factor, *Rhizomucor miehei* aspartic proteinase, and *Myceliophthora thermophila* laccase (WO 95/33836).

In a preferred aspect, the propeptide coding region is nucleotides 58 to 150 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 or SEQ ID NO:9 which encode amino acids -31 to -1 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10; nucleotides 64 to 117 of SEQ ID NO:11 or SEQ ID NO:13 which encode amino acids -18 to -1 of SEQ ID NO:12 or SEQ ID NO:14; nucleotides 64 to 111 of SEQ ID NO:15, SEQ ID NO:17 or SEQ ID NO:19 which encode amino acids -16 to -1 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20; nucleotides 64 to 117 of SEQ ID NO:21 which encode amino acids -18 to -1 of SEQ ID NO:22; nucleotides 58 to 150 of SEQ ID NO:23 which encode amino acids -31 to -1 of SEQ ID NO:24; nucleotides 64 to 120 of SEQ ID NO:25 which encode amino acids -19 to -1 of SEQ ID NO:26; or nucleotides 67 to 144 of SEQ ID NO:27 which encode amino acids -26 to -1 of SEQ ID NO:28.

Where both signal peptide and propeptide regions are present at the amino terminus of a polypeptide, the propeptide region is positioned next to the amino terminus of a polypeptide and the signal peptide region is positioned next to the amino terminus of the propeptide region.

It may also be desirable to add regulatory sequences which allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Regulatory systems in prokaryotic systems include the *lac*, *tac*, and *trp* operator systems. In yeast, the ADH2 system or GAL1 system may be used. In filamentous fungi, the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter may be used as regulatory sequences. Other examples of regulatory sequences are those which allow for gene amplification. In eukaryotic systems, these include the dihydrofolate reductase gene which is amplified in the presence of methotrexate, and the metallothionein genes which are amplified with heavy metals. In these cases, the nucleotide sequence encoding the polypeptide would be operably linked with the regulatory sequence.

Expression Vectors

The present invention also relates to recombinant expression vectors comprising a polynucleotide of the present invention, a promoter, and transcriptional and translational stop signals. The various nucleic acids and control sequences described above may be joined together to produce a recombinant expression vector which may include one or more convenient restriction sites to allow for insertion or substitution of the nucleotide sequence encoding the polypeptide at such sites. Alternatively, a nucleotide sequence of the present invention may be expressed by inserting the nucleotide sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

The recombinant expression vector may be any vector (*e.g.*, a plasmid or virus) which can be conveniently subjected to recombinant DNA procedures and can bring about expression of the nucleotide sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids.

The vector may be an autonomously replicating vector, *i.e.*, a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively,

the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host cell, or a transposon may be used.

The vectors of the present invention preferably contain one or more selectable markers which permit easy selection of transformed cells. A selectable marker is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like.

Examples of bacterial selectable markers are the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol, or tetracycline resistance. Suitable markers for yeast host cells are ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Selectable markers for use in a filamentous fungal host cell include, but are not limited to, *amdS* (acetamidase), *argB* (ornithine carbamoyltransferase), *bar* (phosphinothricin acetyltransferase), *hph* (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (orotidine-5'-phosphate decarboxylase), *sC* (sulfate adenyltransferase), and *trpC* (anthranilate synthase), as well as equivalents thereof. Preferred for use in an *Aspergillus* cell are the *amdS* and *pyrG* genes of *Aspergillus nidulans* or *Aspergillus oryzae* and the *bar* gene of *Streptomyces hygroscopicus*.

The vectors of the present invention preferably contain an element(s) that permits integration of the vector into the host cell's genome or autonomous replication of the vector in the cell independent of the genome.

For integration into the host cell genome, the vector may rely on the polynucleotide's sequence encoding the polypeptide or any other element of the vector for integration into the genome by homologous or nonhomologous recombination. Alternatively, the vector may contain additional nucleotide sequences for directing integration by homologous recombination into the genome of the host cell at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements should preferably contain a sufficient number of nucleic acids, such as 100 to 10,000 base pairs, preferably 400 to 10,000 base pairs, and most preferably 800 to 10,000 base pairs, which have a high degree of identity with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleotide sequences. On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination.

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. The origin of

replication may be any plasmid replicator mediating autonomous replication which functions in a cell. The term "origin of replication" or "plasmid replicator" is defined herein as a nucleotide sequence that enables a plasmid or vector to replicate *in vivo*.

Examples of bacterial origins of replication are the origins of replication of plasmids pBR322, pUC19, pACYC177, and pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, and pAMβ1 permitting replication in *Bacillus*.

Examples of origins of replication for use in a yeast host cell are the 2 micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6.

Examples of origins of replication useful in a filamentous fungal cell are AMA1 and ANS1 (Gems *et al.*, 1991, *Gene* 98:61-67; Cullen *et al.*, 1987, *Nucleic Acids Research* 15: 9163-9175; WO 00/24883). Isolation of the AMA1 gene and construction of plasmids or vectors comprising the gene can be accomplished according to the methods disclosed in WO 00/24883.

More than one copy of a polynucleotide of the present invention may be inserted into the host cell to increase production of the gene product. An increase in the copy number of the polynucleotide can be obtained by integrating at least one additional copy of the sequence into the host cell genome or by including an amplifiable selectable marker gene with the polynucleotide where cells containing amplified copies of the selectable marker gene, and thereby additional copies of the polynucleotide, can be selected for by cultivating the cells in the presence of the appropriate selectable agent.

The procedures used to ligate the elements described above to construct the recombinant expression vectors of the present invention are well known to one skilled in the art (see, *e.g.*, Sambrook *et al.*, 1989, *supra*).

Host Cells

The present invention also relates to recombinant host cells, comprising a polynucleotide of the present invention, which are advantageously used in the recombinant production of the polypeptides. A vector comprising a polynucleotide of the present invention is introduced into a host cell so that the vector is maintained as a chromosomal integrant or as a self-replicating extra-chromosomal vector as described earlier. The term "host cell" encompasses any progeny of a parent cell that is not identical to the parent cell due to mutations that occur during replication. The choice of a host cell will to a large extent depend upon the gene encoding the polypeptide and its source.

The host cell may be a unicellular microorganism, *e.g.*, a prokaryote, or a non-unicellular microorganism, *e.g.*, a eukaryote.

Useful unicellular microorganisms are bacterial cells such as gram positive bacteria

including, but not limited to, a *Bacillus* cell, *e.g.*, *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis*, and *Bacillus thuringiensis*; or a *Streptomyces* cell, *e.g.*, *Streptomyces lividans* and *Streptomyces murinus*, or gram negative bacteria such as *E. coli* and *Pseudomonas* sp. In a preferred aspect, the bacterial host cell is a *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus stearothermophilus*, or *Bacillus subtilis* cell. In another preferred aspect, the *Bacillus* cell is an alkalophilic *Bacillus*.

The introduction of a vector into a bacterial host cell may, for instance, be effected by protoplast transformation (see, *e.g.*, Chang and Cohen, 1979, *Molecular General Genetics* 168: 111-115), using competent cells (see, *e.g.*, Young and Spizizin, 1961, *Journal of Bacteriology* 81: 823-829, or Dubnau and Davidoff-Abelson, 1971, *Journal of Molecular Biology* 56: 209-221), electroporation (see, *e.g.*, Shigekawa and Dower, 1988, *Biotechniques* 6: 742-751), or conjugation (see, *e.g.*, Koehler and Thorne, 1987, *Journal of Bacteriology* 169: 5771-5278).

The host cell may also be a eukaryote, such as a mammalian, insect, plant, or fungal cell.

In a preferred aspect, the host cell is a fungal cell. "Fungi" as used herein includes the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota (as defined by Hawksworth *et al.*, *In, Ainsworth and Bisby's Dictionary of The Fungi*, 8th edition, 1995, CAB International, University Press, Cambridge, UK) as well as the Oomycota (as cited in Hawksworth *et al.*, 1995, *supra*, page 171) and all mitosporic fungi (Hawksworth *et al.*, 1995, *supra*).

In a more preferred aspect, the fungal host cell is a yeast cell. "Yeast" as used herein includes ascosporogenous yeast (Endomycetales), basidiosporogenous yeast, and yeast belonging to the Fungi Imperfecti (Blastomycetes). Since the classification of yeast may change in the future, for the purposes of this invention, yeast shall be defined as described in *Biology and Activities of Yeast* (Skinner, F.A., Passmore, S.M., and Davenport, R.R., eds, *Soc. App. Bacteriol. Symposium Series* No. 9, 1980).

In an even more preferred aspect, the yeast host cell is a *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* cell.

In a most preferred aspect, the yeast host cell is a *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis* or *Saccharomyces oviformis* cell. In another most preferred aspect, the yeast host cell is a *Kluyveromyces lactis* cell. In another most preferred aspect, the yeast host cell is a *Yarrowia lipolytica* cell.

In another more preferred aspect, the fungal host cell is a filamentous fungal cell.

"Filamentous fungi" include all filamentous forms of the subdivision Eumycota and Oomycota (as defined by Hawksworth *et al.*, 1995, *supra*). The filamentous fungi are generally characterized by a mycelial wall composed of chitin, cellulose, glucan, chitosan, mannan, and other complex polysaccharides. Vegetative growth is by hyphal elongation and carbon catabolism is obligately aerobic. In contrast, vegetative growth by yeasts such as *Saccharomyces cerevisiae* is by budding of a unicellular thallus and carbon catabolism may be fermentative.

In an even more preferred aspect, the filamentous fungal host cell is an *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis*, *Coprinus*, *Coriolus*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*, *Phlebia*, *Piromyces*, *Pleurotus*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, *Trametes*, or *Trichoderma* cell.

In a most preferred aspect, the filamentous fungal host cell is an *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger* or *Aspergillus oryzae* cell. In another most preferred aspect, the filamentous fungal host cell is a *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, or *Fusarium venenatum* cell. In another most preferred aspect, the filamentous fungal host cell is a *Bjerkandera adusta*, *Ceriporiopsis aneirina*, *Ceriporiopsis aneirina*, *Ceriporiopsis caregiea*, *Ceriporiopsis gilvescens*, *Ceriporiopsis pannocinta*, *Ceriporiopsis rivulosa*, *Ceriporiopsis subrufa*, or *Ceriporiopsis subvermispora*, *Coprinus cinereus*, *Coriolus hirsutus*, *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium purpurogenum*, *Phanerochaete chrysosporium*, *Phlebia radiata*, *Pleurotus eryngii*, *Thielavia terrestris*, *Trametes villosa*, *Trametes versicolor*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, or *Trichoderma viride* strain cell.

Fungal cells may be transformed by a process involving protoplast formation, transformation of the protoplasts, and regeneration of the cell wall in a manner known *per se*. Suitable procedures for transformation of *Aspergillus* and *Trichoderma* host cells are described in EP 238 023 and Yelton *et al.*, 1984, *Proceedings of the National Academy of Sciences USA* 81: 1470-1474. Suitable methods for transforming *Fusarium* species are described by Malardier *et al.*, 1989, *Gene* 78: 147-156, and WO 96/00787. Yeast may be transformed using the procedures described by Becker and Guarente, *In* Abelson, J.N. and Simon, M.I., editors, *Guide to Yeast Genetics and Molecular Biology*, *Methods in Enzymology*,

Volume 194, pp 182-187, Academic Press, Inc., New York; Ito *et al.*, 1983, *Journal of Bacteriology* 153: 163; and Hinnen *et al.*, 1978, *Proceedings of the National Academy of Sciences USA* 75: 1920.

Methods of Production

The present invention also relates to methods for producing a polypeptide of the present invention, comprising (a) cultivating a cell, which in its wild-type form is capable of producing the polypeptide, under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide. Preferably, the cell is of the genus *Arenicola*, and more preferably *Arenicola marina*.

The present invention also relates to methods for producing a polypeptide of the present invention, comprising (a) cultivating a host cell under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide.

The present invention also relates to methods for producing a polypeptide of the present invention, comprising (a) cultivating a host cell under conditions conducive for production of the polypeptide, wherein the host cell comprises a mutant nucleotide sequence having at least one mutation in the mature polypeptide coding region of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27, wherein the mutant nucleotide sequence encodes a polypeptide which consists of
amino acids 1 to 49 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;
amino acids 1 to 46 of SEQ ID NO:12 or SEQ ID NO:14;
amino acids 1 to 48 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20;
amino acids 1 to 45 of SEQ ID NO:22;
amino acids 1 to 49 of SEQ ID NO:24;
amino acids 1 to 44 of SEQ ID NO:26; or
amino acids 1 to 59 of SEQ ID NO:28, and (b) recovering the polypeptide.

In the production methods of the present invention, the cells are cultivated in a nutrient medium suitable for production of the polypeptide using methods well known in the art. For example, the cell may be cultivated by shake flask cultivation, and small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors performed in a suitable medium and under conditions allowing the polypeptide to be expressed and/or isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable media are available from commercial suppliers or may be prepared according to published compositions (*e.g.*, in catalogues of the American Type Culture

Collection). If the polypeptide is secreted into the nutrient medium, the polypeptide can be recovered directly from the medium. If the polypeptide is not secreted, it can be recovered from cell lysates.

The polypeptides may be detected using methods known in the art that are specific for the polypeptides. These detection methods may include use of specific antibodies. For example, an antimicrobial activity assay may be used to determine the activity of the polypeptide as described herein.

The resulting polypeptide may be recovered using methods known in the art. For example, the polypeptide may be recovered from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, or precipitation.

The polypeptides of the present invention may be purified by a variety of procedures known in the art including, but not limited to, chromatography (*e.g.*, ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (*e.g.*, preparative isoelectric focusing), differential solubility (*e.g.*, ammonium sulfate precipitation), SDS-PAGE, or extraction (see, *e.g.*, *Protein Purification*, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

Plants

The present invention also relates to a transgenic plant, plant part, or plant cell which has been transformed with a nucleotide sequence encoding a polypeptide having antimicrobial activity of the present invention so as to express and produce the polypeptide in recoverable quantities. The polypeptide may be recovered from the plant or plant part. Alternatively, the plant or plant part containing the recombinant polypeptide may be used as such for improving the quality of a food or feed, *e.g.*, improving nutritional value, palatability, and rheological properties, or to destroy an antinutritive factor.

The transgenic plant can be dicotyledonous (a dicot) or monocotyledonous (a monocot). Examples of monocot plants are grasses, such as meadow grass (blue grass, Poa), forage grass such as Festuca, Lolium, temperate grass, such as Agrostis, and cereals, *e.g.*, wheat, oats, rye, barley, rice, sorghum, and maize (corn).

Examples of dicot plants are tobacco, legumes, such as lupins, potato, sugar beet, pea, bean and soybean, and cruciferous plants (family Brassicaceae), such as cauliflower, rape seed, and the closely related model organism *Arabidopsis thaliana.*

Examples of plant parts are stem, callus, leaves, root, fruits, seeds, and tubers as well as the individual tissues comprising these parts, *e.g.*, epidermis, mesophyll, parenchyme, vascular tissues, meristems. Specific plant cell compartments, such as chloroplasts, apoplasts, mitochondria, vacuoles, peroxisomes and cytoplasm are also considered to be a

plant part. Furthermore, any plant cell, whatever the tissue origin, is considered to be a plant part. Likewise, plant parts such as specific tissues and cells isolated to facilitate the utilisation of the invention are also considered plant parts, *e.g.*, embryos, endosperms, aleurone and seeds coats.

Also included within the scope of the present invention are the progeny of such plants, plant parts, and plant cells.

The transgenic plant or plant cell expressing a polypeptide of the present invention may be constructed in accordance with methods known in the art. In short, the plant or plant cell is constructed by incorporating one or more expression constructs encoding a polypeptide of the present invention into the plant host genome and propagating the resulting modified plant or plant cell into a transgenic plant or plant cell.

The expression construct is conveniently a nucleic acid construct which comprises a polynucleotide encoding a polypeptide of the present invention operably linked with appropriate regulatory sequences required for expression of the nucleotide sequence in the plant or plant part of choice. Furthermore, the expression construct may comprise a selectable marker useful for identifying host cells into which the expression construct has been integrated and DNA sequences necessary for introduction of the construct into the plant in question (the latter depends on the DNA introduction method to be used).

The choice of regulatory sequences, such as promoter and terminator sequences and optionally signal or transit sequences is determined, for example, on the basis of when, where, and how the polypeptide is desired to be expressed. For instance, the expression of the gene encoding a polypeptide of the present invention may be constitutive or inducible, or may be developmental, stage or tissue specific, and the gene product may be targeted to a specific tissue or plant part such as seeds or leaves. Regulatory sequences are, for example, described by Tague *et al.*, 1988, *Plant Physiology* 86: 506.

For constitutive expression, the 35S-CaMV, the maize ubiquitin 1, and the rice actin 1 promoter may be used (Franck *et al.*, 1980, *Cell* 21: 285-294, Christensen *et al.*, 1992, *Plant Mo. Biol.* 18: 675-689; Zhang *et al.*, 1991, *Plant Cell* 3: 1155-1165). Organ-specific promoters may be, for example, a promoter from storage sink tissues such as seeds, potato tubers, and fruits (Edwards & Coruzzi, 1990, *Ann. Rev. Genet.* 24: 275-303), or from metabolic sink tissues such as meristems (Ito *et al.*, 1994, *Plant Mol. Biol.* 24: 863-878), a seed specific promoter such as the glutelin, prolamin, globulin, or albumin promoter from rice (Wu *et al.*, 1998, *Plant and Cell Physiology* 39: 885-889), a *Vicia faba* promoter from the legumin B4 and the unknown seed protein gene from *Vicia faba* (Conrad *et al.*, 1998, *Journal of Plant Physiology* 152: 708-711), a promoter from a seed oil body protein (Chen *et al.*, 1998, *Plant and Cell Physiology* 39: 935-941), the storage protein napA promoter from *Brassica napus*, or any other seed specific promoter known in the art, *e.g.*, as described in WO 91/14772.

Furthermore, the promoter may be a leaf specific promoter such as the *rbcs* promoter from rice or tomato (Kyozuka *et al.*, 1993, *Plant Physiology* 102: 991-1000, the chlorella virus adenine methyltransferase gene promoter (Mitra and Higgins, 1994, *Plant Molecular Biology* 26: 85-93), or the *aldP* gene promoter from rice (Kagaya *et al.*, 1995, *Molecular and General Genetics* 248: 668-674), or a wound inducible promoter such as the potato pin2 promoter (Xu *et al.*, 1993, *Plant Molecular Biology* 22: 573-588). Likewise, the promoter may inducible by abiotic treatments such as temperature, drought, or alterations in salinity or induced by exogenously applied substances that activate the promoter, *e.g.*, ethanol, oestrogens, plant hormones such as ethylene, abscisic acid, and gibberellic acid, and heavy metals.

A promoter enhancer element may also be used to achieve higher expression of a polypeptide of the present invention in the plant. For instance, the promoter enhancer element may be an intron which is placed between the promoter and the nucleotide sequence encoding a polypeptide of the present invention. For instance, Xu *et al.*, 1993, *supra*, disclose the use of the first intron of the rice actin 1 gene to enhance expression.

The selectable marker gene and any other parts of the expression construct may be chosen from those available in the art.

The nucleic acid construct is incorporated into the plant genome according to conventional techniques known in the art, including *Agrobacterium*-mediated transformation, virus-mediated transformation, microinjection, particle bombardment, biolistic transformation, and electroporation (Gasser *et al.*, 1990, *Science* 244: 1293; Potrykus, 1990, *Bio/Technology* 8: 535; Shimamoto *et al.*, 1989, *Nature* 338: 274).

Presently, *Agrobacterium tumefaciens*-mediated gene transfer is the method of choice for generating transgenic dicots (for a review, see Hooykas and Schilperoort, 1992, *Plant Molecular Biology* 19: 15-38) and can also be used for transforming monocots, although other transformation methods are often used for these plants. Presently, the method of choice for generating transgenic monocots is particle bombardment (microscopic gold or tungsten particles coated with the transforming DNA) of embryonic calli or developing embryos (Christou, 1992, *Plant Journal* 2: 275-281; Shimamoto, 1994, *Current Opinion Biotechnology* 5: 158-162; Vasil *et al.*, 1992, *Bio/Technology* 10: 667-674). An alternative method for transformation of monocots is based on protoplast transformation as described by Omirulleh *et al.*, 1993, *Plant Molecular Biology* 21: 415-428.

Following transformation, the transformants having incorporated the expression construct are selected and regenerated into whole plants according to methods well-known in the art. Often the transformation procedure is designed for the selective elimination of selection genes either during regeneration or in the following generations by using, for example, co-transformation with two separate T-DNA constructs or site specific excision of the selection gene by a specific recombinase.

The present invention also relates to methods for producing a polypeptide of the present invention comprising (a) cultivating a transgenic plant or a plant cell comprising a polynucleotide encoding a polypeptide having antimicrobial activity of the present invention under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide.

Compositions

The present invention also relates to compositions, such as pharmaceutical compositions, comprising a polypeptide of the present invention. Preferably, the compositions are enriched in such a polypeptide. The term "enriched" indicates that the antimicrobial activity of the composition has been increased, *e.g.*, with an enrichment factor of 1.1.

The compositions may further comprise another pharmaceutically active agent, such as an additional biocidal agent, such as another antimicrobial polypeptide exhibiting antimicrobial activity as defined above. The biocidal agent may be an antibiotic, as known in the art. Classes of antibiotics include penicillins, e.g. penicillin G, penicillin V, methicillin, oxacillin, carbenicillin, nafcillin, ampicillin, etc.; penicillins in combination with beta-lactamase inhibitors, cephalosporins, e.g. cefaclor, cefazolin, cefuroxime, moxalactam, etc.; carbapenems; monobactams; aminoglycosides; tetracyclines; macrolides; lincomycins; polymyxins; sulfonamides; quinolones; cloramphenical; metronidazole; spectinomycin; trimethoprim; vancomycin; etc. The biocidal agent may also be an anti-mycotic agent, including polyenes, e.g. amphotericin B, nystatin; 5-flucosyn; and azoles, e.g. miconazol, ketoconazol, itraconazol and fluconazol.

In an embodiment the biocidal agent is a non-enzymatic chemical agent. In another embodiment the biocidal agent is a non-polypeptide chemical agent.

The compositions may comprise a suitable carrier material. The compositions may also comprise a suitable delivery vehicle capable of delivering the antimicrobial polypeptides of the invention to the desired locus when the compositions are used as a medicament.

The polypeptide compositions may be prepared in accordance with methods known in the art and may be in the form of a liquid or a dry composition. For instance, the polypeptide composition may be in the form of a granulate or a microgranulate. The polypeptide to be included in the composition may be stabilized in accordance with methods known in the art.

Examples are given below of preferred uses of the polypeptide compositions of the invention. The dosage of the polypeptide composition of the invention and other conditions under which the composition is used may be determined on the basis of methods known in the art.

Methods and Uses

The present invention is also directed to methods for using the polypeptides having antimicrobial activity. The antimicrobial polypeptides are typically useful at any locus subject to contamination by bacteria, fungi, yeast or algae. Typically, loci are in aqueous systems such as cooling water systems, laundry rinse water, oil systems such as cutting oils, lubricants, oil fields and the like, where microorganisms need to be killed or where their growth needs to be controlled. However, the present invention may also be used in all applications for which known antimicrobial compositions are useful, such as protection of wood, latex, adhesive, glue, paper, cardboard, textile, leather, plastics, caulking, and feed.

Other uses include preservation of foods, beverages, cosmetics such as lotions, creams, gels, ointments, soaps, shampoos, conditioners, antiperspirants, deodorants, mouth wash, contact lens products, enzyme formulations, or food ingredients.

Thus, the antimicrobial polypeptides of the invention may by useful as a disinfectant, e.g., in the treatment of infections in the eye or the mouth, skin infections; in antiperspirants or deodorants; for cleaning and disinfection of contact lenses and teeth (oral care).

In general it is contemplated that the antimicrobial polypeptides of the present invention are useful for cleaning, disinfecting or inhibiting microbial growth on any surface. Examples of surfaces, which may advantageously be contacted with the antimicrobial polypeptides of the invention are surfaces of process equipment used e.g. dairies, chemical or pharmaceutical process plants, water sanitation systems, oil processing plants, paper pulp processing plants, water treatment plants, and cooling towers. The antimicrobial polypeptides of the invention should be used in an amount, which is effective for cleaning, disinfecting or inhibiting microbial growth on the surface in question.

The antimicrobial polypeptides of the invention may additionally be used for cleaning surfaces and cooking utensils in food processing plants and in any area in which food is prepared or served such as hospitals, nursing homes and restaurants.

It may also be used as a preservation agent or a disinfection agent in water based paints.

The invention also relates to the use of an antimicrobial polypeptide or composition of the invention as a medicament. Further, an antimicrobial polypeptide or composition of the invention may also be used for the manufacture of a medicament for controlling or combating microorganisms, such as fungal organisms or bacteria, preferably gram positive bacteria.

The composition and antimicrobial polypeptide of the invention may be used as an antimicrobial veterinarian or human therapeutic or prophylactic agent. Thus, the composition and antimicrobial polypeptide of the invention may be used in the preparation of veterinarian or human therapeutic agents or prophylactic agents for the treatment of microbial infections, such as bacterial or fungal infections, preferably gram positive bacterial infections. In particular the microbial infections may be associated with lung diseases including, but not limited to,

tuberculosis, pneumonia and cystic fibrosis; and sexual transmitted diseases including, but not limited to, gonorrhea and chlamydia.

The composition of the invention comprises an effective amount of the antimicrobial polypeptide of the invention.

The term "effective amount" when used herein is intended to mean an amount of the antimicrobial polypeptides of the invention, which is sufficient to inhibit growth of the microorganisms in question.

The invention also relates to wound healing compositions or products such as bandages, medical devices such as, e.g., catheters and further to anti-dandruff hair products, such as shampoos.

Formulations of the antimicrobial polypeptides of the invention are administered to a host suffering from or predisposed to a microbial infection. Administration may be topical, localized or systemic, depending on the specific microorganism, preferably it will be localized. Generally the dose of the antimicrobial polypeptides of the invention will be sufficient to decrease the microbial population by at least about 50%, usually by at least 1 log, and may be by 2 or more logs of killing. The compounds of the present invention are administered at a dosage that reduces the microbial population while minimizing any side-effects. It is contemplated that the composition will be obtained and used under the guidance of a physician for *in vivo* use. The antimicrobial polypeptides of the invention are particularly useful for killing gram negative bacteria, including *Pseudomonas aeruginosa,* and *Chlamydia trachomatis*; and gram-positive bacteria, including streptococci such as *Streptococcus pneumonia*, *S. uberis*, *S. hyointestinalis*, *S. pyogenes* and *S. agalactiae;* and staphylococci such as *Staphylococcus aureus, S. epidermidis, S. simulans, S. xylosus,* and *S. carnosus.*

Formulations of the antimicrobial polypeptides of the invention may be administered to a host suffering from or predisposed to a microbial lung infection, such as pneumonia; or to a microbial wound infection, such as a bacterial wound infection.

Formulations of the antimicrobial polypeptides of the invention may also be administered to a host suffering from or predisposed to a skin infection, such as acne, atopic dermatitis or seborrheic dermatitis; preferably the skin infection is a bacterial skin infection, e.g. caused by *Staphylococcus epidermidis*, *Staphylococcus aureus*, *Propionibacterium acnes*, *Pityrosporum ovale* or *Malassezia furfur.*

The antimicrobial polypeptides of the invention are also useful for *in vitro* formulations to kill microbes, particularly where one does not wish to introduce quantities of conventional antibiotics. For example, the antimicrobial polypeptides of the invention may be added to animal and/or human food preparations; or they may be included as an additive for *in vitro* cultures of cells, to prevent the overgrowth of microbes in tissue culture.

The susceptibility of a particular microbe to killing with the antimicrobial polypeptides of

the invention may be determined by *in vitro* testing, as detailed in the experimental section. Typically a culture of the microbe is combined with the antimicrobial polypeptide at varying concentrations for a period of time sufficient to allow the protein to act, usually between about one hour and one day. The viable microbes are then counted, and the level of killing determined.

Microbes of interest include, but are not limited to, Gram-negative bacteria, for example: *Citrobacter sp.; Enterobacter sp.; Escherichia sp., e.g. E. coli; Klebsiella sp.; Morganella sp.; Proteus sp.; Providencia sp.; Salmonella sp., e.g. S. typhi, S. typhimurium; Serratia sp.; Shigella sp.; Pseudomonas sp., e.g. P. aeruginosa; Yersinia sp., e.g. Y. pestis, Y. pseudotuberculosis, Y. enterocolitica; Franciscella sp.; Pasturella sp.; Vibrio sp., e.g. V. cholerae, V. parahemolyticus; Campylobacter sp., e.g. C. jejuni; Haemophilus sp., e.g. H. influenzae, H. ducreyi; Bordetella sp., e.g. B. pertussis, B. bronchiseptica, B. parapertussis; Brucella sp., Neisseria sp., e.g. N. gonorrhoeae, N. meningitidis, etc.* Other bacteria of interest include *Legionella sp., e.g. L. pneumophila; Listeria sp., e.g. L. monocytogenes; Mycoplasma sp., e.g. M. hominis, M. pneumoniae; Mycobacterium sp., e.g. M. tuberculosis, M. leprae; Treponema sp., e.g. T. pallidum; Borrelia sp., e.g. B. burgdorferi; Leptospirae sp.; Rickettsia sp., e.g. R. rickettsii, R. typhi; Chlamydia sp., e.g. C. trachomatis, C. pneumoniae, C. psittaci; Helicobacter sp., e.g. H. pylori, etc.*

Non-bacterial pathogens of interest include fungal and protozoan pathogens, *e.g. Plasmodia sp., e.g. P. falciparum, Trypanosoma sp., e.g. T. brucei;* shistosomes; *Entaemoeba sp., Cryptococcus sp., Candida sp., e.g. C. albicans; etc.*

Various methods for administration may be employed. The polypeptide formulation may be given orally, or may be injected intravascularly, subcutaneously, peritoneally, by aerosol, opthalmically, intra-bladder, topically, *etc.* For example, methods of administration by inhalation are well-known in the art. The dosage of the therapeutic formulation will vary widely, depending on the specific antimicrobial polypeptide to be administered, the nature of the disease, the frequency of administration, the manner of administration, the clearance of the agent from the host, and the like. The initial dose may be larger, followed by smaller maintenance doses. The dose may be administered as infrequently as weekly or biweekly, or fractionated into smaller doses and administered once or several times daily, semi-weekly, *etc.* to maintain an effective dosage level. In many cases, oral administration will require a higher dose than if administered intravenously. The amide bonds, as well as the amino and carboxy termini, may be modified for greater stability on oral administration. For example, the carboxy terminus may be amidated.

Formulations

The compounds of this invention can be incorporated into a variety of formulations for

therapeutic administration. More particularly, the compounds of the present invention can be formulated into pharmaceutical compositions by combination with appropriate, pharmaceutically acceptable carriers or diluents, and may be formulated into preparations in solid, semi-solid, liquid or gaseous forms, such as tablets, capsules, powders, granules, ointments, creams, foams, solutions, suppositories, injections, inhalants, gels, microspheres, lotions, and aerosols. As such, administration of the compounds can be achieved in various ways, including oral, buccal, rectal, parenteral, intraperitoneal, intradermal, transdermal, intracheal, *etc.,* administration. The antimicrobial polypeptides of the invention may be systemic after administration or may be localized by the use of an implant or other formulation that acts to retain the active dose at the site of implantation.

In one embodiment, a formulation for topical use comprises a chelating agent that decreases the effective concentration of divalent cations, particularly calcium and magnesium. For example, agents such as citrate, EGTA or EDTA may be included, where citrate is preferred. The concentration of citrate will usually be from about 1 to 10 mM.

The compounds of the present invention can be administered alone, in combination with each other, or they can be used in combination with other known compounds *(e.g.,* perforin, anti-inflammatory agents, antibiotics, *etc.)* In pharmaceutical dosage forms, the compounds may be administered in the form of their pharmaceutically acceptable salts. The following methods and excipients are merely exemplary and are in no way limiting.

For oral preparations, the compounds can be used alone or in combination with appropriate additives to make tablets, powders, granules or capsules, for example, with conventional additives, such as lactose, mannitol, corn starch or potato starch; with binders, such as crystalline cellulose, cellulose derivatives, acacia, corn starch or gelatins; with disintegrators, such as corn starch, potato starch or sodium carboxymethylcellulose; with lubricants, such as talc or magnesium stearate; and if desired, with diluents, buffering agents, moistening agents, preservatives and flavoring agents.

The compounds can be formulated into preparations for injections by dissolving, suspending or emulsifying them in an aqueous or nonaqueous solvent, such as vegetable or other similar oils, synthetic aliphatic acid glycerides, esters of higher aliphatic acids or propylene glycol; and if desired, with conventional additives such as solubilizers, isotonic agents, suspending agents, emulsifying agents, stabilizers and preservatives.

The compounds can be utilized in aerosol formulation to be administered via inhalation. The compounds of the present invention can be formulated into pressurized acceptable propellants such as dichlorodifluoromethane, propane, nitrogen and the like.

The compounds can be used as lotions, for example to prevent infection of burns, by formulation with conventional additives such as solubilizers, isotonic agents, suspending agents, emulsifying agents, stabilizers and preservatives.

Furthermore, the compounds can be made into suppositories by mixing with a variety of bases such as emulsifying bases or water-soluble bases. The compounds of the present invention can be administered rectally via a suppository. The suppository can include vehicles such as cocoa butter, carbowaxes and polyethylene glycols, which melt at body temperature, yet are solidified at room temperature.

Unit dosage forms for oral or rectal administration such as syrups, elixirs, and suspensions may be provided wherein each dosage unit, for example, teaspoonful, tablespoonful, tablet or suppository, contains a predetermined amount of the composition containing one or more compounds of the present invention. Similarly, unit dosage forms for injection or intravenous administration may comprise the compound of the present invention in a composition as a solution in sterile water, normal saline or another pharmaceutically acceptable carrier.

Implants for sustained release formulations are well-known in the art. Implants are formulated as microspheres, slabs, *etc.* with biodegradable or non-biodegradable polymers. For example, polymers of lactic acid and/or glycolic acid form an erodible polymer that is well-tolerated by the host. The implant containing the antimicrobial polypeptides of the invention is placed in proximity to the site of infection, so that the local concentration of active agent is increased relative to the rest *of* the body.

The term "unit dosage form", as used herein, refers to physically discrete units suitable as unitary dosages for human and animal subjects, each unit containing a predetermined quantity of compounds of the present invention calculated in an amount sufficient to produce the desired effect in association with a pharmaceutically acceptable diluent, carrier or vehicle. The specifications for the unit dosage forms of the present invention depend on the particular compound employed and the effect to be achieved, and the pharmacodynamics associated with the compound in the host.

The pharmaceutically acceptable excipients, such as vehicles, adjuvants, carriers or diluents, are readily available to the public. Moreover, pharmaceutically acceptable auxiliary substances, such as pH adjusting and buffering agents, tonicity adjusting agents, stabilizers, wetting agents and the like, are readily available to the public.

Typical dosages for systemic administration range from 0.1 pg to 100 milligrams per kg weight of subject per administration. A typical dosage may be one tablet taken from two to six times daily, or one time-release capsule or tablet taken once a day and containing a proportionally higher content of active ingredient. The time-release effect may be obtained by capsule materials that dissolve at different pH values, by capsules that release slowly by osmotic pressure, or by any other known means of controlled release.

Those of skill will readily appreciate that dose levels can vary as a function of the specific compound, the severity of the symptoms and the susceptibility of the subject to side effects.

Some of the specific compounds are more potent than others. Preferred dosages for a given compound are readily determinable by those of skill in the art by a variety of means. A preferred means is to measure the physiological potency of a given compound.

The use of liposomes as a delivery vehicle is one method of interest. The liposomes fuse with the cells of the target site and deliver the contents of the lumen intracellularly. The liposomes are maintained in contact with the cells for sufficient time for fusion, using various means to maintain contact, such as isolation, binding agents, and the like. In one aspect of the invention, liposomes are designed to be aerosolized for pulmonary administration. Liposomes may be prepared with purified proteins or peptides that mediate fusion of membranes, such as Sendai virus or influenza virus, etc. The lipids may be any useful combination of known liposome forming lipids, including cationic or zwitterionic lipids, such as phosphatidylcholine. The remaining lipid will be normally be neutral or acidic lipids, such as cholesterol, phosphatidyl serine, phosphatidyl glycerol, and the like.

For preparing the liposomes, the procedure described by Kato *et al.* (1991) *J. Biol. Chem.* **266**:3361 may be used. Briefly, the lipids and lumen composition containing peptides are combined in an appropriate aqueous medium, conveniently a saline medium where the total solids will be in the range of about 1-10 weight percent. After intense agitation for short periods of time, from about 5-60 sec., the tube is placed in a warm water bath, from about 25-40°C and this cycle repeated from about 5-10 times. The composition is then sonicated for a convenient period of time, generally from about 1-10 sec. and may be further agitated by vortexing. The volume is then expanded by adding aqueous medium, generally increasing the volume by about from 1-2 fold, followed by shaking and cooling. This method allows for the incorporation into the lumen of high molecular weight molecules.

Formulations with Other Active Agents

For use in the subject methods, the antimicrobial polypeptides of the invention may be formulated with other pharmaceutically active agents, particularly other antimicrobial agents. Other agents of interest include a wide variety of antibiotics, as known in the art. Classes of antibiotics include penicillins, e.g. penicillin G, penicillin V, methicillin, oxacillin, carbenicillin, nafcillin, ampicillin, etc.; penicillins in combination with beta-lactamase inhibitors, cephalosporins, e.g. cefaclor, cefazolin, cefuroxime, moxalactam, etc.; carbapenems; monobactams; aminoglycosides; tetracyclines; macrolides; lincomycins; polymyxins; sulfonamides; quinolones; cloramphenical; metronidazole; spectinomycin; trimethoprim; vancomycin; etc.

Anti-mycotic agents are also useful, including polyenes, e.g. amphotericin B, nystatin; 5-flucosyn; and azoles, e.g. miconazol, ketoconazol, itraconazol and fluconazol. Antituberculotic drugs include isoniazid, ethambutol, streptomycin and rifampin. Cytokines may also be

included in a formulation of the antimicrobial polypeptides of the invention, e.g. interferon gamma, tumor necrosis factor alpha, interleukin 12, etc.

***In vitro* synthesis**

The antimicrobial peptides of the invention may be prepared by *in vitro* synthesis, using conventional methods as known in the art. Various commercial synthetic apparatuses are available, for example automated synthesizers by Applied Biosystems Inc., Beckman, etc. By using synthesizers, naturally occurring amino acids may be substituted with unnatural amino acids, particularly D-isomers (or D-forms) e.g. D-alanine and D-isoleucine, diastereoisomers, side chains having different lengths or functionalities, and the like. The particular sequence and the manner of preparation will be determined by convenience, economics, purity required, and the like.

Chemical linking may be provided to various peptides or proteins comprising convenient functionalities for bonding, such as amino groups for amide or substituted amine formation, e.g. reductive amination, thiol groups for thioether or disulfide formation, carboxyl groups for amide formation, and the like.

If desired, various groups may be introduced into the peptide during synthesis or during expression, which allow for linking to other molecules or to a surface. Thus cysteines can be used to make thioethers, histidines for linking to a metal ion complex, carboxyl groups for forming amides or esters, amino groups for forming amides, and the like.

The polypeptides may also be isolated and purified in accordance with conventional methods of recombinant synthesis. A lysate may be prepared of the expression host and the lysate purified using HPLC, exclusion chromatography, gel electrophoresis, affinity chromatography, or other purification technique. For the most part, the compositions which are used will comprise at least 20% by weight of the desired product, more usually at least about 75% by weight, preferably at least about 95% by weight, and for therapeutic purposes, usually at least about 99.5% by weight, in relation to contaminants related to the method of preparation of the product and its purification. Usually, the percentages will be based upon total protein

Animal Feed

The present invention is also directed to methods for using the polypeptides having antimicrobial activity in animal feed, as well as to feed compositions and feed additives comprising the antimicrobial polypeptides of the invention.

The term animal includes all animals, including human beings. Examples of animals are non-ruminants, and ruminants, such as cows, sheep and horses. In a particular embodiment, the animal is a non-ruminant animal. Non-ruminant animals include mono-gastric animals, e.g.

pigs or swine (including, but not limited to, piglets, growing pigs, and sows); poultry such as turkeys and chicken (including but not limited to broiler chicks, layers); young calves; and fish (including but not limited to salmon).

The term feed or feed composition means any compound, preparation, mixture, or composition suitable for, or intended for intake by an animal.

In the use according to the invention the antimicrobial polypeptide can be fed to the animal before, after, or simultaneously with the diet. The latter is preferred.

In a particular embodiment, the antimicrobial polypeptide, in the form in which it is added to the feed, or when being included in a feed additive, is well defined. Well-defined means that the antimicrobial polypeptide preparation is at least 50% pure as determined by Size-exclusion chromatography (see Example 12 of WO 01/58275). In other particular embodiments the antimicrobial polypeptide preparation is at least 60, 70, 80, 85, 88, 90, 92, 94, or at least 95% pure as determined by this method.

A well-defined antimicrobial polypeptide preparation is advantageous. For instance, it is much easier to dose correctly to the feed an antimicrobial polypeptide that is essentially free from interfering or contaminating other antimicrobial polypeptides. The term dose correctly refers in particular to the objective of obtaining consistent and constant results, and the capability of optimising dosage based upon the desired effect.

For the use in animal feed, however, the antimicrobial polypeptide need not be that pure; it may e.g. include other enzymes, in which case it could be termed an antimicrobial polypeptide preparation.

The antimicrobial polypeptide preparation can be (a) added directly to the feed (or used directly in a treatment process of vegetable proteins), or (b) it can be used in the production of one or more intermediate compositions such as feed additives or premixes that is subsequently added to the feed (or used in a treatment process). The degree of purity described above refers to the purity of the original antimicrobial polypeptide preparation, whether used according to (a) or (b) above.

Antimicrobial polypeptide preparations with purities of this order of magnitude are in particular obtainable using recombinant methods of production, whereas they are not so easily obtained and also subject to a much higher batch-to-batch variation when the antimicrobial polypeptide is produced by traditional fermentation methods.

Such antimicrobial polypeptide preparation may of course be mixed with other enzymes.

The term vegetable proteins as used herein refers to any compound, composition, preparation or mixture that includes at least one protein derived from or originating from a vegetable, including modified proteins and protein-derivatives. In particular embodiments, the protein content of the vegetable proteins is at least 10, 20, 30, 40, 50, or 60% (w/w).

Vegetable proteins may be derived from vegetable protein sources, such as legumes

and cereals, for example materials from plants of the families *Fabaceae* (*Leguminosae*), *Cruciferaceae*, *Chenopodiaceae*, and *Poaceae*, such as soy bean meal, lupin meal and rapeseed meal.

In a particular embodiment, the vegetable protein source is material from one or more plants of the family *Fabaceae*, e.g. soybean, lupine, pea, or bean.

In another particular embodiment, the vegetable protein source is material from one or more plants of the family *Chenopodiaceae*, e.g. beet, sugar beet, spinach or quinoa.

Other examples of vegetable protein sources are rapeseed, and cabbage.

Soybean is a preferred vegetable protein source.

Other examples of vegetable protein sources are cereals such as barley, wheat, rye, oat, maize (corn), rice, and sorghum.

The antimicrobial polypeptide can be added to the feed in any form, be it as a relatively pure antimicrobial polypeptide, or in admixture with other components intended for addition to animal feed, i.e. in the form of animal feed additives, such as the so-called pre-mixes for animal feed.

In a further aspect the present invention relates to compositions for use in animal feed, such as animal feed, and animal feed additives, e.g. premixes.

Apart from the antimicrobial polypeptide of the invention, the animal feed additives of the invention contain at least one fat soluble vitamin, and/or at least one water soluble vitamin, and/or at least one trace mineral, and/or at least one macro mineral.

Further, optional, feed-additive ingredients are colouring agents, aroma compounds, stabilisers, and/or at least one other enzyme selected from amongst phytases EC 3.1.3.8 or 3.1.3.26; xylanases EC 3.2.1.8; galactanases EC 3.2.1.89; and/or beta-glucanases EC 3.2.1.4.

In a particular embodiment these other enzymes are well defined (as defined above for antimicrobial polypeptide preparations).

Examples of other antimicrobial peptides (AMPs) are CAP18, Leucocin A, Tritrpticin, Protegrin-1, Thanatin, Defensin, Ovispirin such as Novispirin (Robert Lehrer, 2000), and variants, or fragments thereof which retain antimicrobial activity.

Examples of other antifungal polypeptides (AFPs) are the *Aspergillus giganteus*, and *Aspergillus niger* peptides, as well as variants and fragments thereof which retain antifungal activity, as disclosed in WO 94/01459 and WO 02/090384.

Usually fat and water soluble vitamins, as well as trace minerals form part of a so-called premix intended for addition to the feed, whereas macro minerals are usually separately added to the feed. Either of these composition types, when enriched with an antimicrobial polypeptide of the invention, is an animal feed additive of the invention.

In a particular embodiment, the animal feed additive of the invention is intended for

being included (or prescribed as having to be included) in animal diets or feed at levels of 0.01 to 10.0%; more particularly 0.05 to 5.0%; or 0.2 to 1.0% (% meaning g additive per 100 g feed). This is so in particular for premixes.

The following are non-exclusive lists of examples of these components:

Examples of fat soluble vitamins are vitamin A, vitamin D3, vitamin E, and vitamin K, e.g. vitamin K3.

Examples of water soluble vitamins are vitamin B12, biotin and choline, vitamin B1, vitamin B2, vitamin B6, niacin, folic acid and panthothenate, e.g. Ca-D-panthothenate.

Examples of trace minerals are manganese, zinc, iron, copper, iodine, selenium, and cobalt.

Examples of macro minerals are calcium, phosphorus and sodium.

The nutritional requirements of these components (exemplified with poultry and piglets/pigs) are listed in Table A of WO 01/58275. Nutritional requirement means that these components should be provided in the diet in the concentrations indicated.

In the alternative, the animal feed additive of the invention comprises at least one of the individual components specified in Table A of WO 01/58275. At least one means either of, one or more of, one, or two, or three, or four and so forth up to all thirteen, or up to all fifteen individual components. More specifically, this at least one individual component is included in the additive of the invention in such an amount as to provide an in-feed-concentration within the range indicated in column four, or column five, or column six of Table A.

The present invention also relates to animal feed compositions. Animal feed compositions or diets have a relatively high content of protein. Poultry and pig diets can be characterised as indicated in Table B of WO 01/58275, columns 2-3. Fish diets can be characterised as indicated in column 4 of this Table B. Furthermore such fish diets usually have a crude fat content of 200-310 g/kg.

An animal feed composition according to the invention has a crude protein content of 50-800 g/kg, and furthermore comprises at least one antimicrobial polypeptide as claimed herein.

Furthermore, or in the alternative (to the crude protein content indicated above), the animal feed composition of the invention has a content of metabolisable energy of 10-30 MJ/kg; and/or a content of calcium of 0.1-200 g/kg; and/or a content of available phosphorus of 0.1-200 g/kg; and/or a content of methionine of 0.1-100 g/kg; and/or a content of methionine plus cysteine of 0.1-150 g/kg; and/or a content of lysine of 0.5-50 g/kg.

In particular embodiments, the content of metabolisable energy, crude protein, calcium, phosphorus, methionine, methionine plus cysteine, and/or lysine is within any one of ranges 2, 3, 4 or 5 in Table B of WO 01/58275 (R. 2-5).

Crude protein is calculated as nitrogen (N) multiplied by a factor 6.25, i.e. Crude protein (g/kg)= N (g/kg) x 6.25. The nitrogen content is determined by the Kjeldahl method (A.O.A.C.,

1984, Official Methods of Analysis 14th ed., Association of Official Analytical Chemists, Washington DC).

Metabolisable energy can be calculated on the basis of the NRC publication Nutrient requirements in swine, ninth revised edition 1988, subcommittee on swine nutrition, committee on animal nutrition, board of agriculture, national research council. National Academy Press, Washington, D.C., pp. 2-6, and the European Table of Energy Values for Poultry Feed-stuffs, Spelderholt centre for poultry research and extension, 7361 DA Beekbergen, The Netherlands. Grafisch bedrijf Ponsen & looijen bv, Wageningen. ISBN 90-71463-12-5.

The dietary content of calcium, available phosphorus and amino acids in complete animal diets is calculated on the basis of feed tables such as Veevoedertabel 1997, gegevens over chemische samenstelling, verteerbaarheid en voederwaarde van voedermiddelen, Central Veevoederbureau, Runderweg 6, 8219 pk Lelystad. ISBN 90-72839-13-7.

In a particular embodiment, the animal feed composition of the invention contains at least one vegetable protein or protein source as defined above.

In still further particular embodiments, the animal feed composition of the invention contains 0-80% maize; and/or 0-80% sorghum; and/or 0-70% wheat; and/or 0-70% Barley; and/or 0-30% oats; and/or 0-40% soybean meal; and/or 0-10% fish meal; and/or 0-20% whey. Animal diets can e.g. be manufactured as mash feed (non pelleted) or pelleted feed. Typically, the milled feed-stuffs are mixed and sufficient amounts of essential vitamins and minerals are added according to the specifications for the species in question. Enzymes can be added as solid or liquid enzyme formulations. For example, a solid enzyme formulation is typically added before or during the mixing step; and a liquid enzyme preparation is typically added after the pelleting step. The enzyme may also be incorporated in a feed additive or premix.

The final enzyme concentration in the diet is within the range of 0.01-200 mg enzyme protein per kg diet, for example in the range of 5-30 mg enzyme protein per kg animal diet.

The antimicrobial polypeptide may be administered in one or more of the following amounts (dosage ranges): 0.01-200; or 0.01-100; or 0.05-100; or 0.05-50; or 0.10-10 - all these ranges being in mg antimicrobial polypeptide protein per kg feed (ppm).

For determining mg antimicrobial polypeptide protein per kg feed, the antimicrobial polypeptide is purified from the feed composition, and the specific activity of the purified antimicrobial polypeptide is determined using a relevant assay (see under antimicrobial activity, substrates, and assays). The antimicrobial activity of the feed composition as such is also determined using the same assay, and on the basis of these two determinations, the dosage in mg antimicrobial polypeptide protein per kg feed is calculated.

The same principles apply for determining mg antimicrobial polypeptide protein in feed additives. Of course, if a sample is available of the antimicrobial polypeptide used for preparing the feed additive or the feed, the specific activity is determined from this sample (no

need to purify the antimicrobial polypeptide from the feed composition or the additive).

Signal Peptide and Propeptide

The present invention also relates to nucleic acid constructs comprising a gene encoding a protein operably linked to one or both of a first nucleotide sequence consisting of nucleotides 1 to 57 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 or SEQ ID NO:9 encoding a signal peptide consisting of amino acids -50 to -32 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10; nucleotides 1 to 63 of SEQ ID NO:11 or SEQ ID NO:13 encoding a signal peptide consisting of amino acids -39 to -19 of SEQ ID NO:12 or SEQ ID NO:14; nucleotides 1 to 63 of SEQ ID NO:15, SEQ ID NO:17 or SEQ ID NO:19 encoding a signal peptide consisting of amino acids -37 to -17 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20; nucleotides 1 to 63 of SEQ ID NO:21 encoding a signal peptide consisting of amino acids -39 to -19 of SEQ ID NO:22; nucleotides 1 to 57 of SEQ ID NO:23 encoding a signal peptide consisting of amino acids -50 to -32 of SEQ ID NO:24; nucleotides 1 to 63 of SEQ ID NO:25 encoding a signal peptide consisting of amino acids -40 to -20 of SEQ ID NO:26; or nucleotides 1 to 66 of SEQ ID NO:27 encoding a signal peptide consisting of amino acids -48 to -27 of SEQ ID NO:28;

and a second nucleotide sequence consisting of nucleotides 58 to 150 of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 or SEQ ID NO:9 encoding a propeptide consisting of amino acids -31 to -1 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10; nucleotides 64 to 117 of SEQ ID NO:11 or SEQ ID NO:13 encoding a propeptide consisting of amino acids -18 to -1 of SEQ ID NO:12 or SEQ ID NO:14; nucleotides 64 to 111 of SEQ ID NO:15, SEQ ID NO:17 or SEQ ID NO:19 encoding a propeptide consisting of amino acids -16 to -1 of SEQ ID NO:16, SEQ ID NO:18 or SEQ ID NO:20; nucleotides 64 to 117 of SEQ ID NO:21 encoding a propeptide consisting of amino acids -18 to -1 of SEQ ID NO:22; nucleotides 58 to 150 of SEQ ID NO:23 encoding a propeptide consisting of amino acids -31 to -1 of SEQ ID NO:24; nucleotides 64 to 120 of SEQ ID NO:25 encoding a propeptide consisting of amino acids -19 to -1 of SEQ ID NO:26; or nucleotides 67 to 144 of SEQ ID NO:27 encoding a propeptide consisting of amino acids -26 to -1 of SEQ ID NO:28;

wherein the gene is foreign to the first and second nucleotide sequences.

The present invention also relates to recombinant expression vectors and recombinant host cells comprising such nucleic acid constructs.

The present invention also relates to methods for producing a protein comprising (a) cultivating such a recombinant host cell under conditions suitable for production of the protein; and (b) recovering the protein.

The first and second nucleotide sequences may be operably linked to foreign genes individually with other control sequences or in combination with other control sequences. Such

other control sequences are described *supra*. As described earlier, where both signal peptide and propeptide regions are present at the amino terminus of a protein, the propeptide region is positioned next to the amino terminus of a protein and the signal peptide region is positioned next to the amino terminus of the propeptide region.

The protein may be native or heterologous to a host cell. The term "protein" is not meant herein to refer to a specific length of the encoded product and, therefore, encompasses peptides, oligopeptides, and proteins. The term "protein" also encompasses two or more polypeptides combined to form the encoded product. The proteins also include hybrid polypeptides which comprise a combination of partial or complete polypeptide sequences obtained from at least two different proteins wherein one or more may be heterologous or native to the host cell. Proteins further include naturally occurring allelic and engineered variations of the above mentioned proteins and hybrid proteins.

Preferably, the protein is a hormone or variant thereof, enzyme, receptor or portion thereof, antibody or portion thereof, or reporter. In a more preferred aspect, the protein is an oxidoreductase, transferase, hydrolase, lyase, isomerase, or ligase. In an even more preferred aspect, the protein is an aminopeptidase, amylase, carbohydrase, carboxypeptidase, catalase, cellulase, chitinase, cutinase, cyclodextrin glycosyltransferase, deoxyribonuclease, esterase, alpha-galactosidase, beta-galactosidase, glucoamylase, alpha-glucosidase, beta-glucosidase, invertase, laccase, lipase, mannosidase, mutanase, oxidase, pectinolytic enzyme, peroxidase, phytase, polyphenoloxidase, proteolytic enzyme, ribonuclease, transglutaminase or xylanase.

The gene may be obtained from any prokaryotic, eukaryotic, or other source.

The present invention is further described by the following examples which should not be construed as limiting the scope of the invention.

EXAMPLES

Chemicals used as buffers and substrates were commercial products of at least reagent grade.

EXAMPLE 1

Preparation of cDNA library of *Arenicola marina*

A cDNA library was prepared from *A. marina* gut. PolyA enriched RNA was purified, cDNA was synthesized and the library made according to standard molecular biology procedures. A detailed protocol on the general process can be found in the examples of international patent application WO 01/12794. Vector used for cloning was pMHAs7i. Plasmid pMHas7i is a derivative of pMhas5 (see WO 03/044049) in which SfiI cloning sites compatible

with SfiI adapted cDNA created in the SMART protocol. Briefly, the EcoRI-NotI human aldolase cDNA fragment of lambdaTripIX2 (Clontech) was cloned into the Eco-RI cloning sites of pMHas5. The resulting plasmid, pMHas7i contains the human aldolase cDNA flanked by the appropriate SfiI sites needed for cloning SMART adapted cDNAs.

Preparation of pMHas7i for cloning of cDNAs: The vector was restricted with SfiI and the plasmid gel purified from the aldolase insert by agarose gel electrophoresis. The band containing the restricted plasmid was purified by GFX treatment (GE Healthcare).

EXAMPLE 2

Discovery of the Marinasin peptide family by signal trapping of the *A. marina* cDNA library

A cDNA plasmid pool was prepared from 20,000 total transformants of the original cDNA-pMHas5 vector ligation and signal trapped as described in WO 01/77315 resulting in 381 sequence contigs.

All 381 contigs were independently blasted and the results analyzed. One contig (ZY151351) shared some amino acid homology with known antimicrobial polypeptides (defensin-like polypeptides). The derived polypeptide was termed Marinasin 1A (SEQ ID NO:6).

Upon re-examination of the 381 contigs by blasting with the Marinasin 1A polypeptide sequence several other "Marinasins" were identified (SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28).

EXAMPLE 3

Construction of an *Aspergillus* expression vector for Marinasins

The cDNA encoding the predicted mature region of Marinasin 1B (amino acids 1 to 49 of SEQ ID NO:8), Marinasin 2A (amino acids 1 to 46 of SEQ ID NO:12), Marinasin 3B (amino acids 1 to 48 of SEQ ID NO:18), Marinasin 4 (amino acids 1 to 45 of SEQ ID NO:22), Marinasin 5 (amino acids 1 to 49 of SEQ ID NO:24), Marinasin 6 (amino acids 1 to 44 of SEQ ID NO:26) and Marinasin 7 (amino acids 1 to 59 of SEQ ID NO:28) were amplified from the *A. marina* gut cDNA library and fused to a DNA fragment encoding the Plectasin pre-pro region in the following manner: 10 ng cDNA was used as template in PCR reactions using the oligonucleotides listed in Table 1.

Table 1

Marinasin1B-F	5' GGATGCGAACCAACTTCAGAAACGTAGCTGGCTATGTAACTGGTTGGGC 3'
Marinasin1B-R	5' CCCAAGCTTCACATGGTGTATGGTTGATCTCTCC 3'

Marinasin2A-F	5' GGATGCGAACCAACTTCAGAAACGTGGTTGGTGCTGGCAGTGGACATGT 3'
Marinasin2B-R	5' CCCAAGCTTGCCCTTCTAGCGTTCAGCTCTT 3'
Marinasin3B-F	5' GGATGCGAACCAACTTCAGAAACGTCATTGGTGTTTCGAGTGGTCATGT 3'
Marinasin3B-R	5' CCCAAGCTTTCAGTGGAGAACCGTTACAGCA 3'
Marinasin4-F	5' GGATGCGAACCAACTTCAGAAACGTATCCCCTGTTGGACTCCGACATGT 3'
Marinasin4-R	5' CCCAAGCTTAGCCCAGTATGCTCTGCACGT 3'
Marinasin5-F	5' GGATGCGAACCAACTTCAGAAACGTGGGGGCCGGCCCTGTCATAGGCAT 3'
Marinasin5-R	5' CCCAAGCTTAGTTGTTCGCTCCATTAGCACCT 3'
Marinasin6-F	5' ACCAACTTCAGAAACGTGGCAGGTCGTGTAATTTCTGGTT 3'
Marinasin6-R	5' CCCAAGCTTTGGGCGATTCTATCTGCCTCTTA 3'
Marinasin7-F	5' ACCAACTTCAGAAACGTGGTGGGATGTGTGGTGACGA 3'
Marinasin7-R	5' CCCAAGCTTGCACATCGTTTCCACCGCAA 3'

5 pmol of each primer F and R (eg. Marinasin1B-F and Marinasin1B-R) were used in a 25 µl reaction volume with the Expand High Fidelity PCR System (Roche). After denaturation at 94°C for 2 minutes, 35 cycles of PCR was carried out with the following program: 94°C for 15 sec and 60°C for 60 sec.

The Plectasin pre-pro region including the 58 bp intron (see Examples of WO 03/044049) was amplified from a plasmid template using the oligonucleotides listed in Table 2.

Table 2

PlecBHI-F	5' CGCGGATCCCACCATGCAATTTACCACCATCCTCTC 3'
Plec-Mar1B-R	5' GCCCAACCAGTTACATAGCCAGCTACGTTTCTGAAGTTGGTTCGCATCC 3'
Plec-Mar2A-R	5' ACATGTCCACTGCCAGCACCAACCACGTTTCTGAAGTTGGTTCGCATCC 3'
Plec-Mar3B-R	5' ACATGACCACTCGAAACACCAATGACGTTTCTGAAGTTGGTTCGCATCC 3'
Plec-Mar4-R	5' TCCAACAGGGGATACGTTTCTGAAGTTGGTTCGCATCC 3'
Plec-Mar5-R	5' ATGCCTATGACAGGGCCGGCCCCCACGTTTCTGAAGTTGGTTCGCATCC 3'
Plec-Mar6-R	5' AAATTACACGACCTGCCACGTTTCTGAAGTTGGTTCGCATCC 3'
Plec-Mar7-R	5' CACACATCCCACCACGTTTCTGAAGTTGGTTCGCATCC 3'

5 pmol of each primer F and R (eg. PlecBHI-F and Marinasin1B-R) were used in a 25 µl reaction volume with the Expand High Fidelity PCR System (Roche). After denaturation at 94°C for 2 minutes, 35 cycles of PCR was carried out with the following program: 94°C for 15 sec and 60°C for 60 sec.

The PCR reactions were separated on 2% agarose gel. Bands of the expected sizes for both the Marinasin cDNAs and the Plectasin pre-pro region were isolated using the GFX DNA

purification kit (Amersham) according to the manufacturer's protocol. 1/50 of the purified material was used as template for fusion PCR using the Plectasin forward primer and the Marinasin reverse primer (eg. for Marinasin 1B: The product of Marinasin1B-F and Marinasin1B-R combined with the product of PlecBHI-F and Plec-Mar1B-R as template using PlecBHI-F and Marinasin1B-R as primers). 5 pmol of each primer F and R (eg. Marinasin1B-F and Marinasin1B-R) were used in a 25 µl reaction volume with the Expand High Fidelity PCR System (Roche). After denaturation at 94°C for 2 minutes, 25 cycles of PCR was carried out with the following program: 94°C for 15 sec and 60°C for 60 sec. The PCR reactions were separated on 2% agarose gel. Bands of the expected sizes encoding the Plectasin/Marinasin fusion constructs were isolated using the GFX DNA purification kit (Amersham) according to the manufacturer's protocol. The fragments were digested with BamHI and HindIII which cut in the overhangs introduced by the PCR primers. The digested fragments were isolated and cloned into the expression plasmid pDAu109 (see Examples of WO 2005/042735)

EXAMPLE 4

Expression of Marinasins in *Aspergillus*

The pDAu109 based Marinasin expression plasmids were transformed into *Aspergillus oryzae* strain BECh2 (see WO 00/393229). Ten to twenty transformants of each strain were re-isolated twice under selective and noninducing conditions on Cove minimal plates with sucrose and acetamide. To test expression of the Marinasins, transformants were grown for 3 days at 26 degrees C in tubes with 10 ml Dap2C-1 (see 'Media' in the Examples of WO 2004/032648). Expression was verified by Maldi-TOF and SDS-page of culture supernatants on NuPage 16% Tricine SDS gels (Invitrogen) as recommended by the manufacturer.

EXAMPLE 5

Antimicrobial activity by radial diffusion assay

The supernatants described above in Example 4 were analyzed in a radial diffusion assay following a previously published protocol for detection of antimicrobial activity (Lehrer *et al.*, (1991) Ultra sensitive assays for endogenous antimicrobial polypeptides *J Immunol Methods* 137: 167-173). Target bacteria (10^6 colony forming units (CFU)) were added to 10 ml of underlay agarose (1% low electro-endosmosis agarose, 0.03% Trypticase soy broth, 10 mM sodium phosphate, pH 7.4, 37 degrees Celsius). Suspension was solidified on an INTEGRID Petri Dish (Becton Dickinson Labware, NJ). A 3 mm Gel Puncher was used to make holes in the underlay agarose (Amersham Pharmacia Biotech, Sweden). Samples were added to the holes and incubated at 37 degrees Celsius, 3 hours. An overlay was poured on top and the plate was incubated overnight (LB media, 7.5% Agar). Antimicrobial activity was seen as bacterial clearing zones around the wells. Living cells was counterstained by adding 10 ml, 0.2

mM MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide Thiazolyl blue). All standard protocols have been described elsewhere (Sambrook, Fritsch, and Maniatis, 1989). An electronic picture is taken. The bacterial growth inhibition is seen as clearing zones, which are measured to the nearest 0.1 mm and subtracted the diameter of the hole. The results are shown as zones of inhibition of native samples (Table 3) and concentrated samples (Table 4) against a range of test strains:

A: *Staphylococcus simulans* (ATCC11631)

B: *Staphylococcus carnosus* (ATCC51365)

C: *Streptococcus hyointestinalis* (ATCC49169)

D: *Bacillus subtilis* (ATCC23857)

E: *Enterococcus saccharolyticus* (ATCC43076)

F: *Escherichia coli* (ATCC10536)

G: *Acinetobacter anitratus* (ATCC17903)

H: *Erwinia chrysanthemi* (ATCC11663)

J: *Pseudomonas boreopolis* (ATCC15452)

Concentration of samples was performed by precipitation of 45 ml culture supernatant with 2.5 ml 100 % TCA and resuspension of the pelleted precipitate in 450 µl resuspension buffer (100 mM Tris pH 8, 100 mM NaCl).

Table 3

Radial diffusion assay results with native culture supernatants of recombinant *A. oryzae* clones expressing Marinasins. 10 units in radial diffusion assay correspond to 1 mm diameter clear zone around the 3 mm sample well.

	Test strain			
	B	**C**	**D**	**E**
Marinasin 1B	21	10	10	9
Marinasin 2A	26	12	16	14
Marinasin 4	60	54	32	43
Marinasin 6	-	45	39	40
Control	0	0	0	0

Table 4

Radial diffusion assay results with TCA precipitated culture supernatants of recombinant *A. oryzae* clones expressing Marinasins (100 x concentrated). 10 units in radial diffusion assay correspond to 1 mm diameter clear zone around the 3 mm sample well.

	Test strain

	A	B	C	D	E	F	G	H	J
Marinasin 1B	22	25	22	28	30	25	26	25	28
Marinasin 2A	34	35	35	38	41	25	39	40	39
Marinasin 4	14	81	59	32	68	0	6	0	0
Marinasin 5	22	13	24	25	23	19	26	26	22
Control	0	0	0	0	0	0	0	0	0

EXAMPLE 6

Using the HMM files from the PFAM database to identify a defensin

Sequence analysis using hidden markov model profiles (HMM profiles) may be carried out either online on the Internet or locally on a computer using the well-known HMMER freely available software package. The current version is HMMER 2.3.2 from October 2003.

The HMM profiles may be obtained from the well-known PFAM database. The current version is PFAM 16.0 from November 2004. Both HMMER and PFAM are available for all computer platforms from e.g. Washington University in St. Louis (USA), School of Medicine (http://pfam.wustl.edu and http://hmmer.wustl.edu).

If a query amino acid sequence, or a fragment thereof, belongs to one of the following five PFAM families, the amino acid sequence is a defensin according to the present invention:

- Defensin_beta or "Beta Defensin", accession number: PF00711;
- Defensin_propep or "Defensin propeptide", accession number: PF00879;
- Defensin_1 or "Mammalian defensin", accession number: PF00323;
- Defensin_2 or "Arthropod defensin", accession number: PF01097;
- Gamma-thionin or "Gamma-thionins family", accession number: PF00304.

An amino acid sequence belongs to a PFAM family, according to the present invention, if it generates an E-value which is greater than 0.1, and a score which is larger or equal to zero, when the PFAM database is used online, or when the hmmpfam program (from the HMMER software package) is used locally.

When the sequence analysis is carried out locally using the hmmpfam program, it is necessary to obtain (download) the HMM profiles from the PFAM database. Two profiles exist for each family; "xxx_ls.hmm" for glocal searches, and "xxx_fs.hmm" for local searches ("xxx" is the name of the family). That makes a total of ten profiles for the five families mentioned above.

These ten profiles may be used individually, or joined (appended) into a single profile (using a text editor - the profiles are ASCII files) that could be named e.g. defensin.hmm. A query amino acid sequence can then be evaluated by using the following command line:

hmmpfam -E 0.1 defensin.hmm sequence_file

- wherein "sequence_file" is a file with the query amino acid sequence in any of the formats recognized by the HMMER software package.

If the score is larger or equal to zero (0.0), and the E-value is greater than 0.1, the query amino acid sequence is a defensin according to the present invention.

The PFAM database is further described in Bateman et al. (2004) "The Pfam Protein Families Database", Nucleic Acids Research, Vol. 32 (Database Issue) pp. D138-D141.

CLAIMS

1. A polypeptide having antimicrobial activity, which comprises the amino acid sequence represented by: C-x(3)-C-x(7,9)-C-C; C-C-x(8)-C-x-C; or C-x-C-x(8,11)-C.

2. The polypeptide of claim 1, which is a defensin.

3. The polypeptide of anyone of claims 1 or 2, which comprises 5, 6, 7, 8, 9 or 10 cysteine residues.

4. The polypeptide of anyone of claims 1-3, which is:

(a) a polypeptide comprising an amino acid sequence which has at least 60% identity with:
amino acids 1 to 49 of SEQ ID NO:2,
amino acids 1 to 49 of SEQ ID NO:4,
amino acids 1 to 49 of SEQ ID NO:6,
amino acids 1 to 49 of SEQ ID NO:8,
amino acids 1 to 49 of SEQ ID NO:10,
amino acids 1 to 46 of SEQ ID NO:12,
amino acids 1 to 46 of SEQ ID NO:14,
amino acids 1 to 48 of SEQ ID NO:16,
amino acids 1 to 48 of SEQ ID NO:18,
amino acids 1 to 48 of SEQ ID NO:20,
amino acids 1 to 45 of SEQ ID NO:22,
amino acids 1 to 49 of SEQ ID NO:24,
amino acids 1 to 44 of SEQ ID NO:26, or
amino acids 1 to 59 of SEQ ID NO:28;

(b) a polypeptide which is encoded by a nucleotide sequence which hybridizes under at least medium stringency conditions with (i):
nucleotides 151 to 297 of SEQ ID NO:1,
nucleotides 151 to 297 of SEQ ID NO:3,
nucleotides 151 to 297 of SEQ ID NO:5,
nucleotides 151 to 297 of SEQ ID NO:7,
nucleotides 151 to 297 of SEQ ID NO:9,
nucleotides 118 to 255 of SEQ ID NO:11,
nucleotides 118 to 255 of SEQ ID NO:13,
nucleotides 112 to 255 of SEQ ID NO:15,
nucleotides 112 to 255 of SEQ ID NO:17,

nucleotides 112 to 255 of SEQ ID NO:19,
nucleotides 118 to 252 of SEQ ID NO:21,
nucleotides 151 to 297 of SEQ ID NO:23,
nucleotides 121 to 252 of SEQ ID NO:25, or
nucleotides 145 to 321 of SEQ ID NO:27; or (ii) a complementary strand of (i); or

(c) a variant comprising a conservative substitution, deletion, and/or insertion of one or more amino acids of:
amino acids 1 to 49 of SEQ ID NO:2,
amino acids 1 to 49 of SEQ ID NO:4,
amino acids 1 to 49 of SEQ ID NO:6,
amino acids 1 to 49 of SEQ ID NO:8,
amino acids 1 to 49 of SEQ ID NO:10,
amino acids 1 to 46 of SEQ ID NO:12,
amino acids 1 to 46 of SEQ ID NO:14,
amino acids 1 to 48 of SEQ ID NO:16,
amino acids 1 to 48 of SEQ ID NO:18,
amino acids 1 to 48 of SEQ ID NO:20,
amino acids 1 to 45 of SEQ ID NO:22,
amino acids 1 to 49 of SEQ ID NO:24,
amino acids 1 to 44 of SEQ ID NO:26, or
amino acids 1 to 59 of SEQ ID NO:28.

5. The polypeptide of claim 4, comprising an amino acid sequence which has at least 65% identity, preferably at least 70% identity, at least 80% identity, at least 85% identity, at least 90% identity, at least 95% identity, or at least 99% identity with:
amino acids 1 to 49 of SEQ ID NO:2,
amino acids 1 to 49 of SEQ ID NO:4,
amino acids 1 to 49 of SEQ ID NO:6,
amino acids 1 to 49 of SEQ ID NO:8,
amino acids 1 to 49 of SEQ ID NO:10,
amino acids 1 to 46 of SEQ ID NO:12,
amino acids 1 to 46 of SEQ ID NO:14,
amino acids 1 to 48 of SEQ ID NO:16,
amino acids 1 to 48 of SEQ ID NO:18,
amino acids 1 to 48 of SEQ ID NO:20,
amino acids 1 to 45 of SEQ ID NO:22,
amino acids 1 to 49 of SEQ ID NO:24,

amino acids 1 to 44 of SEQ ID NO:26, or
amino acids 1 to 59 of SEQ ID NO:28.

6. The polypeptide of claim 4, which is encoded by a nucleotide sequence which hybridizes under medium stringency conditions, preferably under medium-high, or high stringency conditions with (i):
nucleotides 151 to 297 of SEQ ID NO:1,
nucleotides 151 to 297 of SEQ ID NO:3,
nucleotides 151 to 297 of SEQ ID NO:5,
nucleotides 151 to 297 of SEQ ID NO:7,
nucleotides 151 to 297 of SEQ ID NO:9,
nucleotides 118 to 255 of SEQ ID NO:11,
nucleotides 118 to 255 of SEQ ID NO:13,
nucleotides 112 to 255 of SEQ ID NO:15,
nucleotides 112 to 255 of SEQ ID NO:17,
nucleotides 112 to 255 of SEQ ID NO:19,
nucleotides 118 to 252 of SEQ ID NO:21,
nucleotides 151 to 297 of SEQ ID NO:23,
nucleotides 121 to 252 of SEQ ID NO:25, or
nucleotides 145 to 321 of SEQ ID NO:27; or (ii) a complementary strand of (i).

7. The polypeptide of anyone of claims 1-6, which consists of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28, or a fragment thereof having antimicrobial activity.

8. The polypeptide of claim 7, which consists of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, or SEQ ID NO:28.

9. The polypeptide of claim 7, which consists of
amino acids 1 to 49 of SEQ ID NO:2,
amino acids 1 to 49 of SEQ ID NO:4,
amino acids 1 to 49 of SEQ ID NO:6,
amino acids 1 to 49 of SEQ ID NO:8,
amino acids 1 to 49 of SEQ ID NO:10,
amino acids 1 to 46 of SEQ ID NO:12,

amino acids 1 to 46 of SEQ ID NO:14,
amino acids 1 to 48 of SEQ ID NO:16,
amino acids 1 to 48 of SEQ ID NO:18,
amino acids 1 to 48 of SEQ ID NO:20,
amino acids 1 to 45 of SEQ ID NO:22,
amino acids 1 to 49 of SEQ ID NO:24,
amino acids 1 to 44 of SEQ ID NO:26, or
amino acids 1 to 59 of SEQ ID NO:28.

10. A polynucleotide comprising a nucleotide sequence which encodes the polypeptide of anyone of claims 1-9.

11. The polynucleotide of claim 10, having at least one mutation in the mature polypeptide coding sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27, in which the mutant nucleotide sequence encodes a polypeptide consisting of
amino acids 1 to 49 of SEQ ID NO:2,
amino acids 1 to 49 of SEQ ID NO:4,
amino acids 1 to 49 of SEQ ID NO:6,
amino acids 1 to 49 of SEQ ID NO:8,
amino acids 1 to 49 of SEQ ID NO:10,
amino acids 1 to 46 of SEQ ID NO:12,
amino acids 1 to 46 of SEQ ID NO:14,
amino acids 1 to 48 of SEQ ID NO:16,
amino acids 1 to 48 of SEQ ID NO:18,
amino acids 1 to 48 of SEQ ID NO:20,
amino acids 1 to 45 of SEQ ID NO:22,
amino acids 1 to 49 of SEQ ID NO:24,
amino acids 1 to 44 of SEQ ID NO:26, or
amino acids 1 to 59 of SEQ ID NO:28.

12. A nucleic acid construct comprising the polynucleotide of claim 10 operably linked to one or more control sequences that direct the production of the polypeptide in an expression host.

13. A recombinant expression vector comprising the nucleic acid construct of claim 12.

14. A recombinant host cell comprising the nucleic acid construct of claim 12.

15. A method for producing the polypeptide of any of claims 1-9 comprising (a) cultivating a cell, which in its wild-type form is capable of producing the polypeptide, under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide.

16. A method for producing the polypeptide of any of claims 1-9 comprising (a) cultivating a host cell comprising a nucleic acid construct comprising a nucleotide sequence encoding the polypeptide under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide.

17. An isolated polynucleotide obtained by (a) hybridizing a population of DNA under medium stringency conditions, preferably under medium-high, or high stringency conditions with (i)
nucleotides 151 to 297 of SEQ ID NO:1,
nucleotides 151 to 297 of SEQ ID NO:3,
nucleotides 151 to 297 of SEQ ID NO:5,
nucleotides 151 to 297 of SEQ ID NO:7,
nucleotides 151 to 297 of SEQ ID NO:9,
nucleotides 118 to 255 of SEQ ID NO:11,
nucleotides 118 to 255 of SEQ ID NO:13,
nucleotides 112 to 255 of SEQ ID NO:15,
nucleotides 112 to 255 of SEQ ID NO:17,
nucleotides 112 to 255 of SEQ ID NO:19,
nucleotides 118 to 252 of SEQ ID NO:21,
nucleotides 151 to 297 of SEQ ID NO:23,
nucleotides 121 to 252 of SEQ ID NO:25, or
nucleotides 145 to 321 of SEQ ID NO:27; or (ii) a complementary strand of (i); and (b) isolating the hybridizing polynucleotide, which encodes a polypeptide having antimicrobial activity.

18. A method for producing a polynucleotide having a mutant nucleotide sequence, comprising (a) introducing at least one mutation into the mature polypeptide coding sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, or SEQ ID NO:27, wherein the mutant nucleotide sequence encodes a polypeptide consisting of
amino acids 1 to 49 of SEQ ID NO:2,

amino acids 1 to 49 of SEQ ID NO:4,
amino acids 1 to 49 of SEQ ID NO:6,
amino acids 1 to 49 of SEQ ID NO:8,
amino acids 1 to 49 of SEQ ID NO:10,
amino acids 1 to 46 of SEQ ID NO:12,
amino acids 1 to 46 of SEQ ID NO:14,
amino acids 1 to 48 of SEQ ID NO:16,
amino acids 1 to 48 of SEQ ID NO:18,
amino acids 1 to 48 of SEQ ID NO:20,
amino acids 1 to 45 of SEQ ID NO:22,
amino acids 1 to 49 of SEQ ID NO:24,
amino acids 1 to 44 of SEQ ID NO:26, or
amino acids 1 to 59 of SEQ ID NO:28; and (b) recovering the polynucleotide comprising the mutant nucleotide sequence.

19. A mutant polynucleotide produced by the method of claim 18.

20. A method for producing a polypeptide, comprising (a) cultivating a cell comprising the mutant polynucleotide of claim 19 encoding the polypeptide under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide.

21. A method for producing the polypeptide of any of claims 1-9, comprising (a) cultivating a transgenic plant or a plant cell comprising a polynucleotide encoding a polypeptide having antimicrobial activity of the present invention under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide.

22. A transgenic plant, plant part or plant cell, which has been transformed with a polynucleotide encoding the polypeptide of any of claims 1-9.

23. A method for killing or inhibiting growth of microbial cells comprising contacting the microbial cells with an antimicrobial polypeptide as defined in any of claims 1-9.

24. An antimicrobial polypeptide as defined in any of claims 1-9 for use as a medicament, or an antimicrobial veterinarian or human therapeutic or prophylactic agent.

25. Use of an antimicrobial polypeptide as defined in any of claims 1-9 in the preparation of a veterinarian or human therapeutic agent for the treatment of a microbial infection or for

prophylactic use.

26. Use of at least one antimicrobial polypeptide as defined in any of claims 1-9 in animal feed.

SEQUENCE LISTING

<110> Novozymes A/S

<120> Polypeptides Having Antimicrobial Activity

<130> 10803.204-WO

<160> 50

<170> PatentIn version 3.3

<210> 1
<211> 297
<212> DNA
<213> Arenicola marina

<220>
<221> CDS
<222> (1)..(297)

<220>
<221> sig_peptide
<222> (1)..(57)

<220>
<221> mat_peptide
<222> (151)..(297)

<400> 1

```
atg aag ttt ttc tta ccg att atg atc gcc ttg gca ttt gct gcc gtc       48
Met Lys Phe Phe Leu Pro Ile Met Ile Ala Leu Ala Phe Ala Ala Val
-50                 -45                 -40                 -35

gcc atg gca act tct gat act gag ccc gtt gaa ccg gaa gag gaa ctt       96
Ala Met Ala Thr Ser Asp Thr Glu Pro Val Glu Pro Glu Glu Glu Leu
                -30                 -25                 -20

tcc atc atg cta ccg ttt gty gaa gat gac ttg ctg gag aaa cct atc      144
Ser Ile Met Leu Pro Phe Xaa Glu Asp Asp Leu Leu Glu Lys Pro Ile
            -15                 -10                 -5

ccc cgg agc tgg cta tgt aac tgg ttg ggc cat gac wtc ggc tgt ata      192
Pro Arg Ser Trp Leu Cys Asn Trp Leu Gly His Asp Xaa Gly Cys Ile
    -1  1                   5                   10

rct tat tgc aag ctg ttg ggt aac agc cga ggt ggt tgc tgt gct ggg      240
Xaa Tyr Cys Lys Leu Leu Gly Asn Ser Arg Gly Gly Cys Cys Ala Gly
15                  20                  25                  30

grc gac tgg aas ggr tac tgt tac tgc cac gac ggt mgc agc ccw acc      288
Xaa Asp Trp Xaa Xaa Tyr Cys Tyr Cys His Asp Gly Xaa Ser Xaa Thr
            35                  40                  45

gat agg tgt                                                          297
Asp Arg Cys
```

<210> 2
<211> 99
<212> PRT
<213> Arenicola marina

<220>
<221> misc_feature

<222> (-12)..(-12)
<223> The 'Xaa' at location -12 stands for Val.

<220>
<221> misc_feature
<222> (11)..(11)
<223> The 'Xaa' at location 11 stands for Ile, or Phe.

<220>
<221> misc_feature
<222> (15)..(15)
<223> The 'Xaa' at location 15 stands for Ala, or Thr.

<220>
<221> misc_feature
<222> (31)..(31)
<223> The 'Xaa' at location 31 stands for Gly, or Asp.

<220>
<221> misc_feature
<222> (34)..(34)
<223> The 'Xaa' at location 34 stands for Lys, or Asn.

<220>
<221> misc_feature
<222> (35)..(35)
<223> The 'Xaa' at location 35 stands for Gly.

<220>
<221> misc_feature
<222> (43)..(43)
<223> The 'Xaa' at location 43 stands for Ser, or Arg.

<220>
<221> misc_feature
<222> (45)..(45)
<223> The 'Xaa' at location 45 stands for Pro.

<400> 2

Met Lys Phe Phe Leu Pro Ile Met Ile Ala Leu Ala Phe Ala Ala Val
-50 -45 -40 -35

Ala Met Ala Thr Ser Asp Thr Glu Pro Val Glu Pro Glu Glu Glu Leu
-30 -25 -20

Ser Ile Met Leu Pro Phe Xaa Glu Asp Asp Leu Leu Glu Lys Pro Ile
-15 -10 -5

Pro Arg Ser Trp Leu Cys Asn Trp Leu Gly His Asp Xaa Gly Cys Ile
-1 1 5 10

Xaa Tyr Cys Lys Leu Leu Gly Asn Ser Arg Gly Gly Cys Cys Ala Gly
15 20 25 30

Xaa Asp Trp Xaa Xaa Tyr Cys Tyr Cys His Asp Gly Xaa Ser Xaa Thr
35 40 45

Asp Arg Cys

<210> 3
<211> 297
<212> DNA
<213> Arenicola marina

<220>
<221> CDS
<222> (1)..(297)

<220>
<221> sig_peptide
<222> (1)..(57)

<220>
<221> mat_peptide
<222> (151)..(297)

<400> 3

```
atg aag ttt ttc tta ccg att atg atc gcc ttg gca ttt gct gcc gtc       48
Met Lys Phe Phe Leu Pro Ile Met Ile Ala Leu Ala Phe Ala Ala Val
-50                 -45                 -40                 -35

gcc atg gca act tct gat act gag ccc gtt gaa ccg gaa gag gaa ctt       96
Ala Met Ala Thr Ser Asp Thr Glu Pro Val Glu Pro Glu Glu Glu Leu
                -30                 -25                 -20

tcc atc atg cta ccg ttt gtt gaa gat gac ttg ctg gag aaa cct atc      144
Ser Ile Met Leu Pro Phe Val Glu Asp Asp Leu Leu Glu Lys Pro Ile
            -15                 -10                 -5

ccc cgg agc tgg cta tgt aac tgg ttg ggc cat gac ttc ggc tgt ata      192
Pro Arg Ser Trp Leu Cys Asn Trp Leu Gly His Asp Phe Gly Cys Ile
    -1  1               5                   10

act tat tgc aag ctg ttg ggt aac agc cga ggt ggt tgc tgt gct ggg      240
Thr Tyr Cys Lys Leu Leu Gly Asn Ser Arg Gly Gly Cys Cys Ala Gly
15                  20                  25                  30

ggc gac tgg aag gga tac tgt tac tgc cac gac ggt cgc agc cca acc      288
Gly Asp Trp Lys Gly Tyr Cys Tyr Cys His Asp Gly Arg Ser Pro Thr
                35                  40                  45

gat agg tgt                                                         297
Asp Arg Cys
```

<210> 4
<211> 99
<212> PRT
<213> Arenicola marina

<400> 4

```
Met Lys Phe Phe Leu Pro Ile Met Ile Ala Leu Ala Phe Ala Ala Val
-50                 -45                 -40                 -35

Ala Met Ala Thr Ser Asp Thr Glu Pro Val Glu Pro Glu Glu Glu Leu
                -30                 -25                 -20

Ser Ile Met Leu Pro Phe Val Glu Asp Asp Leu Leu Glu Lys Pro Ile
            -15                 -10                 -5
```

```
Pro Arg Ser Trp Leu Cys Asn Trp Leu Gly His Asp Phe Gly Cys Ile
    -1  1               5                   10

Thr Tyr Cys Lys Leu Leu Gly Asn Ser Arg Gly Gly Cys Cys Ala Gly
15                  20                  25                  30

Gly Asp Trp Lys Gly Tyr Cys Tyr Cys His Asp Gly Arg Ser Pro Thr
                35                  40                  45

Asp Arg Cys
```

<210> 5
<211> 297
<212> DNA
<213> Arenicola marina

<220>
<221> CDS
<222> (1)..(297)

<220>
<221> sig_peptide
<222> (1)..(57)

<220>
<221> mat_peptide
<222> (151)..(297)

<400> 5

```
atg aag ttt ttc tta ccg att atg atc gcc ttg gca ttt gct gcc gtc      48
Met Lys Phe Phe Leu Pro Ile Met Ile Ala Leu Ala Phe Ala Ala Val
-50                 -45                 -40                 -35

gcc atg gca act gct gat act gag ccc gtt gaa ccg gaa gag gaa ctt      96
Ala Met Ala Thr Ala Asp Thr Glu Pro Val Glu Pro Glu Glu Glu Leu
                -30                 -25                 -20

tcc atc atg cta ccg ttt gtt gaa gat gac ttg ctg gag aaa cct atc     144
Ser Ile Met Leu Pro Phe Val Glu Asp Asp Leu Leu Glu Lys Pro Ile
            -15                 -10                 -5

ccc cgg agc tgg cta tgt aac tgg ttg ggc cat gac ttc ggc tgt ata     192
Pro Arg Ser Trp Leu Cys Asn Trp Leu Gly His Asp Phe Gly Cys Ile
    -1  1               5                   10

act tat tgc aag ctg ttg ggt aac agc cga ggt ggt tgc tgt gct ggg     240
Thr Tyr Cys Lys Leu Leu Gly Asn Ser Arg Gly Gly Cys Cys Ala Gly
15                  20                  25                  30

gac gac tgg aag gga tac tgt tac tgc cac gac ggt agc agc cct acc     288
Asp Asp Trp Lys Gly Tyr Cys Tyr Cys His Asp Gly Ser Ser Pro Thr
                35                  40                  45

gat agg tgt                                                        297
Asp Arg Cys
```

<210> 6
<211> 99

<212> PRT
<213> Arenicola marina

<400> 6

```
Met Lys Phe Phe Leu Pro Ile Met Ile Ala Leu Ala Phe Ala Ala Val
-50                 -45                 -40                 -35

Ala Met Ala Thr Ala Asp Thr Glu Pro Val Glu Pro Glu Glu Glu Leu
                -30                 -25                 -20

Ser Ile Met Leu Pro Phe Val Glu Asp Asp Leu Leu Glu Lys Pro Ile
            -15                 -10                 -5

Pro Arg Ser Trp Leu Cys Asn Trp Leu Gly His Asp Phe Gly Cys Ile
    -1  1               5                   10

Thr Tyr Cys Lys Leu Leu Gly Asn Ser Arg Gly Gly Cys Cys Ala Gly
15                  20                  25                  30

Asp Asp Trp Lys Gly Tyr Cys Tyr Cys His Asp Gly Ser Ser Pro Thr
                35                  40                  45

Asp Arg Cys
```

<210> 7
<211> 297
<212> DNA
<213> Arenicola marina

<220>
<221> CDS
<222> (1)..(297)

<220>
<221> sig_peptide
<222> (1)..(57)

<220>
<221> mat_peptide
<222> (151)..(297)

<400> 7

```
atg aag ttt ttc tta ccg att atg atc gcc ttg gca ttt gct gcc gtc      48
Met Lys Phe Phe Leu Pro Ile Met Ile Ala Leu Ala Phe Ala Ala Val
-50                 -45                 -40                 -35

gcc atg gca act gct gat act gag ccc gtt gaa ccg gaa gag gaa ctt      96
Ala Met Ala Thr Ala Asp Thr Glu Pro Val Glu Pro Glu Glu Glu Leu
                -30                 -25                 -20

tcc atc atg cta ccg ttt gtt gaa gat gac ttg ctg gag aaa cct atc     144
Ser Ile Met Leu Pro Phe Val Glu Asp Asp Leu Leu Glu Lys Pro Ile
            -15                 -10                 -5

ccc cgg agc tgg cta tgt aac tgg ttg ggc cat gac ttc ggc tgt ata     192
Pro Arg Ser Trp Leu Cys Asn Trp Leu Gly His Asp Phe Gly Cys Ile
    -1  1               5                   10
```

```
act tat tgc aag ctg ttg ggt aac agc cga ggt ggt tgc tgt gct ggg      240
Thr Tyr Cys Lys Leu Leu Gly Asn Ser Arg Gly Gly Cys Cys Ala Gly
15                  20                  25                  30

ggc gac tgg aac ggg tac tgt tac tgc cac gac ggt cgc agc cca acc      288
Gly Asp Trp Asn Gly Tyr Cys Tyr Cys His Asp Gly Arg Ser Pro Thr
                35                  40                  45

gat agg tgt                                                         297
Asp Arg Cys
```

<210> 8
<211> 99
<212> PRT
<213> Arenicola marina

<400> 8

```
Met Lys Phe Phe Leu Pro Ile Met Ile Ala Leu Ala Phe Ala Ala Val
-50                 -45                 -40                 -35

Ala Met Ala Thr Ala Asp Thr Glu Pro Val Glu Pro Glu Glu Glu Leu
                -30                 -25                 -20

Ser Ile Met Leu Pro Phe Val Glu Asp Asp Leu Leu Glu Lys Pro Ile
            -15                 -10                 -5

Pro Arg Ser Trp Leu Cys Asn Trp Leu Gly His Asp Phe Gly Cys Ile
    -1  1                   5                   10

Thr Tyr Cys Lys Leu Leu Gly Asn Ser Arg Gly Gly Cys Cys Ala Gly
15                  20                  25                  30

Gly Asp Trp Asn Gly Tyr Cys Tyr Cys His Asp Gly Arg Ser Pro Thr
                35                  40                  45

Asp Arg Cys
```

<210> 9
<211> 297
<212> DNA
<213> Arenicola marina

<220>
<221> CDS
<222> (1)..(297)

<220>
<221> sig_peptide
<222> (1)..(57)

<220>
<221> mat_peptide
<222> (151)..(297)

<400> 9

```
atg aag ttt ttc tta ccg att atg atc gcc ttg gca ttt gct gcc gtc        48
Met Lys Phe Phe Leu Pro Ile Met Ile Ala Leu Ala Phe Ala Ala Val
-50                 -45                 -40                 -35

gcc atg gca act gct gat act gag ccc gtt gaa ccg gaa gag gaa ctt        96
Ala Met Ala Thr Ala Asp Thr Glu Pro Val Glu Pro Glu Glu Glu Leu
                -30                 -25                 -20

tcc atc atg cta ccg ttt gtc gaa gat gac ttg ctg gag aaa cct atc       144
Ser Ile Met Leu Pro Phe Val Glu Asp Asp Leu Leu Glu Lys Pro Ile
            -15                 -10                 -5

ccc cgg agc tgg cta tgt aac tgg ttg ggc cat gac atc ggc tgt ata       192
Pro Arg Ser Trp Leu Cys Asn Trp Leu Gly His Asp Ile Gly Cys Ile
    -1  1                   5                   10

gct tat tgc aag ctg ttg ggt aac agc cga ggt ggt tgc tgt gct ggg       240
Ala Tyr Cys Lys Leu Leu Gly Asn Ser Arg Gly Gly Cys Cys Ala Gly
15                  20                  25                  30

ggc gac tgg aag gga tac tgt tac tgc cac gac ggt cgc agc cct acc       288
Gly Asp Trp Lys Gly Tyr Cys Tyr Cys His Asp Gly Arg Ser Pro Thr
                35                  40                  45

gat agg tgt                                                          297
Asp Arg Cys
```

<210> 10
<211> 99
<212> PRT
<213> Arenicola marina

<400> 10

```
Met Lys Phe Phe Leu Pro Ile Met Ile Ala Leu Ala Phe Ala Ala Val
-50                 -45                 -40                 -35

Ala Met Ala Thr Ala Asp Thr Glu Pro Val Glu Pro Glu Glu Glu Leu
                -30                 -25                 -20

Ser Ile Met Leu Pro Phe Val Glu Asp Asp Leu Leu Glu Lys Pro Ile
            -15                 -10                 -5

Pro Arg Ser Trp Leu Cys Asn Trp Leu Gly His Asp Ile Gly Cys Ile
    -1  1                   5                   10

Ala Tyr Cys Lys Leu Leu Gly Asn Ser Arg Gly Gly Cys Cys Ala Gly
15                  20                  25                  30

Gly Asp Trp Lys Gly Tyr Cys Tyr Cys His Asp Gly Arg Ser Pro Thr
                35                  40                  45

Asp Arg Cys
```

<210> 11
<211> 255

<212> DNA
<213> Arenicola marina

<220>
<221> CDS
<222> (1)..(255)

<220>
<221> sig_peptide
<222> (1)..(63)

<220>
<221> mat_peptide
<222> (118)..(255)

<400> 11

```
atg aac caa atc aca ttg ctg atg ttg cta gtc gcc gtc gct gtc atg       48
Met Asn Gln Ile Thr Leu Leu Met Leu Leu Val Ala Val Ala Val Met
                -35                 -30                 -25

gca acc gtt acg gga caa tct ctg gat gac gaa agc gat gtt gaa aac       96
Ala Thr Val Thr Gly Gln Ser Leu Asp Asp Glu Ser Asp Val Glu Asn
            -20                 -15                 -10

gtt tcc caa cct gaa cag cgc ggt tgg tgc tgg cag tgg aca tgt gat      144
Val Ser Gln Pro Glu Gln Arg Gly Trp Cys Trp Gln Trp Thr Cys Asp
        -5              -1  1                   5

gcc cac tgc tat ctc aag cag tat gac agg ggc tgc tgt ggt gtt gga      192
Ala His Cys Tyr Leu Lys Gln Tyr Asp Arg Gly Cys Cys Gly Val Gly
10                  15                  20                  25

gaa cac agc ggt aaa tgc ctt tgc tac gac cat ggt ggc ccc ctc gcc      240
Glu His Ser Gly Lys Cys Leu Cys Tyr Asp His Gly Gly Pro Leu Ala
                30                  35                  40

gat ggc tgc aag agc                                                 255
Asp Gly Cys Lys Ser
            45
```

<210> 12
<211> 85
<212> PRT
<213> Arenicola marina

<400> 12

```
Met Asn Gln Ile Thr Leu Leu Met Leu Leu Val Ala Val Ala Val Met
                -35                 -30                 -25

Ala Thr Val Thr Gly Gln Ser Leu Asp Asp Glu Ser Asp Val Glu Asn
            -20                 -15                 -10

Val Ser Gln Pro Glu Gln Arg Gly Trp Cys Trp Gln Trp Thr Cys Asp
        -5              -1  1                   5

Ala His Cys Tyr Leu Lys Gln Tyr Asp Arg Gly Cys Cys Gly Val Gly
10                  15                  20                  25

Glu His Ser Gly Lys Cys Leu Cys Tyr Asp His Gly Gly Pro Leu Ala
                30                  35                  40
```

```
Asp Gly Cys Lys Ser
            45


<210>  13
<211>  255
<212>  DNA
<213>  Arenicola marina


<220>
<221>  CDS
<222>  (1)..(255)

<220>
<221>  sig_peptide
<222>  (1)..(63)

<220>
<221>  mat_peptide
<222>  (118)..(255)

<400>  13
atg aag caa ctc atc ttg ctg atg ttg ctg gtc gcc atc gct gtc atg       48
Met Lys Gln Leu Ile Leu Leu Met Leu Leu Val Ala Ile Ala Val Met
                -35                 -30                 -25

gca aca gtt acg gca cag tcc ctg gat gac gaa agc gat gtt gat atc       96
Ala Thr Val Thr Ala Gln Ser Leu Asp Asp Glu Ser Asp Val Asp Ile
            -20                 -15                 -10

gtt tcc caa tct gaa cag cgc cag tgg tgc tgg gag tgg tca tgt gat      144
Val Ser Gln Ser Glu Gln Arg Gln Trp Cys Trp Glu Trp Ser Cys Asp
        -5              -1  1               5

gcc aac tgc ttt ttc aag cag ttt gat aac ggc tgc tgt ggt gtt ggg      192
Ala Asn Cys Phe Phe Lys Gln Phe Asp Asn Gly Cys Cys Gly Val Gly
10                  15                  20                  25

gaa cac agc ggt aaa tgc ctt tgc tac gac cat gga ggc ccc ctc gcc      240
Glu His Ser Gly Lys Cys Leu Cys Tyr Asp His Gly Gly Pro Leu Ala
                30                  35                  40

gcc ggc tgc acg cgc                                                 255
Ala Gly Cys Thr Arg
            45


<210>  14
<211>  85
<212>  PRT
<213>  Arenicola marina

<400>  14

Met Lys Gln Leu Ile Leu Leu Met Leu Leu Val Ala Ile Ala Val Met
                -35                 -30                 -25


Ala Thr Val Thr Ala Gln Ser Leu Asp Asp Glu Ser Asp Val Asp Ile
            -20                 -15                 -10


Val Ser Gln Ser Glu Gln Arg Gln Trp Cys Trp Glu Trp Ser Cys Asp
        -5              -1  1               5
```

Ala Asn Cys Phe Phe Lys Gln Phe Asp Asn Gly Cys Cys Gly Val Gly
10 15 20 25

Glu His Ser Gly Lys Cys Leu Cys Tyr Asp His Gly Gly Pro Leu Ala
30 35 40

Ala Gly Cys Thr Arg
45

<210> 15
<211> 255
<212> DNA
<213> Arenicola marina

<220>
<221> CDS
<222> (1)..(255)

<220>
<221> sig_peptide
<222> (1)..(63)

<220>
<221> mat_peptide
<222> (112)..(255)

<400> 15

```
atg aaa aac ata aca ttg ctg atg ctg ctg gtt gcc gtc ttc gtc atg      48
Met Lys Asn Ile Thr Leu Leu Met Leu Leu Val Ala Val Phe Val Met
        -35                 -30                 -25

gca acc gtt atg gca acg cct cta gat ggc gat att gaa gcc gtt tcc      96
Ala Thr Val Met Ala Thr Pro Leu Asp Gly Asp Ile Glu Ala Val Ser
    -20                 -15                 -10

cag cca gag cag cgg ctt tgg tgt ttc gag tgg tca tgt gat gtc aga     144
Gln Pro Glu Gln Arg Leu Trp Cys Phe Glu Trp Ser Cys Asp Val Arg
-5              -1  1               5                   10

tgc tgg tgg tcg cat cag aac ggc tgt tgt ggc gtt ggc gaa aac aaa     192
Cys Trp Trp Ser His Gln Asn Gly Cys Cys Gly Val Gly Glu Asn Lys
            15                  20                  25

cat aaa tgc ctc tgc tat gtg agt ggt ggc ccc ctc gcc gct ggc tgc     240
His Lys Cys Leu Cys Tyr Val Ser Gly Gly Pro Leu Ala Ala Gly Cys
        30                  35                  40

att agg gaa gtg ctg                                                 255
Ile Arg Glu Val Leu
    45
```

<210> 16
<211> 85
<212> PRT
<213> Arenicola marina

<400> 16

```
Met Lys Asn Ile Thr Leu Leu Met Leu Leu Val Ala Val Phe Val Met
        -35                 -30                 -25
```

```
Ala Thr Val Met Ala Thr Pro Leu Asp Gly Asp Ile Glu Ala Val Ser
    -20                 -15                 -10

Gln Pro Glu Gln Arg Leu Trp Cys Phe Glu Trp Ser Cys Asp Val Arg
-5              -1  1                5                   10

Cys Trp Trp Ser His Gln Asn Gly Cys Cys Gly Val Gly Glu Asn Lys
            15                  20                  25

His Lys Cys Leu Cys Tyr Val Ser Gly Gly Pro Leu Ala Ala Gly Cys
        30                  35                  40

Ile Arg Glu Val Leu
    45
```

<210> 17
<211> 255
<212> DNA
<213> Arenicola marina

<220>
<221> CDS
<222> (1)..(255)

<220>
<221> sig_peptide
<222> (1)..(63)

<220>
<221> mat_peptide
<222> (112)..(255)

<400> 17

```
atg aag aac ata acg ttg ctg atg ctg ctg gtt gcc gtc gtc gtc atg       48
Met Lys Asn Ile Thr Leu Leu Met Leu Leu Val Ala Val Val Val Met
        -35                 -30                 -25

gca acc gtt atg gca acg cct cta gat ggc gat att gaa gcc gtt tcc       96
Ala Thr Val Met Ala Thr Pro Leu Asp Gly Asp Ile Glu Ala Val Ser
    -20                 -15                 -10

cag cca gag cag cgt cat tgg tgt ttc gag tgg tca tgt gat gtc aaa      144
Gln Pro Glu Gln Arg His Trp Cys Phe Glu Trp Ser Cys Asp Val Lys
-5              -1  1                5                   10

tgc tgg tgg tcg cat cag aac ggc tgt tgt ggc gtt ggc gaa aac aaa      192
Cys Trp Trp Ser His Gln Asn Gly Cys Cys Gly Val Gly Glu Asn Lys
            15                  20                  25

cat aac tgc ctc tgc tat gag agt ggt ggc ccc ctc gcc gct ggc tgc      240
His Asn Cys Leu Cys Tyr Glu Ser Gly Gly Pro Leu Ala Ala Gly Cys
        30                  35                  40

att agg gaa gtg ctg                                                 255
Ile Arg Glu Val Leu
    45
```

<210> 18

<211> 85
<212> PRT
<213> Arenicola marina

<400> 18

```
Met Lys Asn Ile Thr Leu Leu Met Leu Leu Val Ala Val Val Val Met
        -35                 -30                 -25

Ala Thr Val Met Ala Thr Pro Leu Asp Gly Asp Ile Glu Ala Val Ser
    -20                 -15                 -10

Gln Pro Glu Gln Arg His Trp Cys Phe Glu Trp Ser Cys Asp Val Lys
-5              -1  1                5                   10

Cys Trp Trp Ser His Gln Asn Gly Cys Cys Gly Val Gly Glu Asn Lys
            15                  20                  25

His Asn Cys Leu Cys Tyr Glu Ser Gly Gly Pro Leu Ala Ala Gly Cys
        30                  35                  40

Ile Arg Glu Val Leu
    45
```

<210> 19
<211> 255
<212> DNA
<213> Arenicola marina

<220>
<221> CDS
<222> (1)..(255)

<220>
<221> sig_peptide
<222> (1)..(63)

<220>
<221> mat_peptide
<222> (112)..(255)

<400> 19

```
atg aag aat ata aca atg ctg atg ctg ctg gtt gcc gtc gtc gtc atg       48
Met Lys Asn Ile Thr Met Leu Met Leu Leu Val Ala Val Val Val Met
        -35                 -30                 -25

gca acc gtt atg gca acg cct cta gat ggc gat att gaa gcc gtt tac       96
Ala Thr Val Met Ala Thr Pro Leu Asp Gly Asp Ile Glu Ala Val Tyr
    -20                 -15                 -10

cag ccg gag cag cgt cat tgg tgt tgg gaa tgg aca tgt gat gtc aga      144
Gln Pro Glu Gln Arg His Trp Cys Trp Glu Trp Thr Cys Asp Val Arg
-5              -1  1                5                   10

tgc tgg tgg tcg cat cgg gat ggc tgt tgt ggc gtt ggc gaa aac aaa      192
Cys Trp Trp Ser His Arg Asp Gly Cys Cys Gly Val Gly Glu Asn Lys
            15                  20                  25

aat aac tgc ctc tgc tat gag agt ggt ggc ccc ctc gcc gat ggc tgc      240
Asn Asn Cys Leu Cys Tyr Glu Ser Gly Gly Pro Leu Ala Asp Gly Cys
```

30 35 40

att agc aaa gtg ctg 255
Ile Ser Lys Val Leu
45

<210> 20
<211> 85
<212> PRT
<213> Arenicola marina

<400> 20

Met Lys Asn Ile Thr Met Leu Met Leu Leu Val Ala Val Val Val Met
-35 -30 -25

Ala Thr Val Met Ala Thr Pro Leu Asp Gly Asp Ile Glu Ala Val Tyr
-20 -15 -10

Gln Pro Glu Gln Arg His Trp Cys Trp Glu Trp Thr Cys Asp Val Arg
-5 -1 1 5 10

Cys Trp Trp Ser His Arg Asp Gly Cys Cys Gly Val Gly Glu Asn Lys
15 20 25

Asn Asn Cys Leu Cys Tyr Glu Ser Gly Gly Pro Leu Ala Asp Gly Cys
30 35 40

Ile Ser Lys Val Leu
45

<210> 21
<211> 252
<212> DNA
<213> Arenicola marina

<220>
<221> CDS
<222> (1)..(252)

<220>
<221> sig_peptide
<222> (1)..(63)

<220>
<221> mat_peptide
<222> (118)..(252)

<400> 21
atg aag cca atc ata ttg ctg atg ctg ctg gtc gcc gtt gtc gtc atg 48
Met Lys Pro Ile Ile Leu Leu Met Leu Leu Val Ala Val Val Val Met
-35 -30 -25

gca aca gtt acg gca gag ccc ctg gat gaa gaa agc gat gtt gaa ata 96
Ala Thr Val Thr Ala Glu Pro Leu Asp Glu Glu Ser Asp Val Glu Ile
-20 -15 -10

gtt tcc cag cct gaa cag cgc atc ccc tgt tgg act ccg aca tgt agc 144
Val Ser Gln Pro Glu Gln Arg Ile Pro Cys Trp Thr Pro Thr Cys Ser

```
        -5              -1  1                5

aca cgc tgc tgg tgg aag ggc aag agc ggc tgt tgt ggc gtt aaa gaa      192
Thr Arg Cys Trp Trp Lys Gly Lys Ser Gly Cys Cys Gly Val Lys Glu
10                  15                  20                  25

cac tat ggt gta tgc atc tgc tat cag cat ctc ggc ccg aag gac agc      240
His Tyr Gly Val Cys Ile Cys Tyr Gln His Leu Gly Pro Lys Asp Ser
                30                  35                  40

aga tgc ctg gcc                                                     252
Arg Cys Leu Ala
            45
```

<210> 22
<211> 84
<212> PRT
<213> Arenicola marina

<400> 22

```
Met Lys Pro Ile Ile Leu Leu Met Leu Leu Val Ala Val Val Val Met
                -35                 -30                 -25

Ala Thr Val Thr Ala Glu Pro Leu Asp Glu Glu Ser Asp Val Glu Ile
            -20                 -15                 -10

Val Ser Gln Pro Glu Gln Arg Ile Pro Cys Trp Thr Pro Thr Cys Ser
        -5              -1  1                5

Thr Arg Cys Trp Trp Lys Gly Lys Ser Gly Cys Cys Gly Val Lys Glu
10                  15                  20                  25

His Tyr Gly Val Cys Ile Cys Tyr Gln His Leu Gly Pro Lys Asp Ser
                30                  35                  40

Arg Cys Leu Ala
            45
```

<210> 23
<211> 297
<212> DNA
<213> Arenicola marina

<220>
<221> CDS
<222> (1)..(297)

<220>
<221> sig_peptide
<222> (1)..(57)

<220>
<221> mat_peptide
<222> (151)..(297)

<400> 23

```
atg aag ttt tta tta ccg ttg atg gtc gtc atg gcc ttt gct gcc gtc      48
Met Lys Phe Leu Leu Pro Leu Met Val Val Met Ala Phe Ala Ala Val
```

```
-50                 -45                 -40                 -35

gcc atg gca act gct gat act gag ccg gtt gag agg gag gag ggt gtt      96
Ala Met Ala Thr Ala Asp Thr Glu Pro Val Glu Arg Glu Glu Gly Val
                -30                 -25                 -20

ttc atc atg cta ccg ttt gtt gaa gac gat atg ctg gag aaa cca att     144
Phe Ile Met Leu Pro Phe Val Glu Asp Asp Met Leu Glu Lys Pro Ile
            -15                 -10                 -5

ctt cgg ggg ggc cgg ccc tgt cat agg cat ggc cat gac atg ggt tgc     192
Leu Arg Gly Gly Arg Pro Cys His Arg His Gly His Asp Met Gly Cys
    -1  1                   5                   10

aca gct gct tgc aac gaa agg ggg cac agt cta ggt agt tgc tgt cct     240
Thr Ala Ala Cys Asn Glu Arg Gly His Ser Leu Gly Ser Cys Cys Pro
15                  20                  25                  30

gtg ggc cac tat agt gga tac tgt tac tgc cac gac agt ggc act ccc     288
Val Gly His Tyr Ser Gly Tyr Cys Tyr Cys His Asp Ser Gly Thr Pro
                35                  40                  45

gat agg tgc                                                        297
Asp Arg Cys
```

<210> 24
<211> 99
<212> PRT
<213> Arenicola marina

<400> 24

```
Met Lys Phe Leu Leu Pro Leu Met Val Val Met Ala Phe Ala Ala Val
-50                 -45                 -40                 -35

Ala Met Ala Thr Ala Asp Thr Glu Pro Val Glu Arg Glu Glu Gly Val
                -30                 -25                 -20

Phe Ile Met Leu Pro Phe Val Glu Asp Asp Met Leu Glu Lys Pro Ile
            -15                 -10                 -5

Leu Arg Gly Gly Arg Pro Cys His Arg His Gly His Asp Met Gly Cys
    -1  1                   5                   10

Thr Ala Ala Cys Asn Glu Arg Gly His Ser Leu Gly Ser Cys Cys Pro
15                  20                  25                  30

Val Gly His Tyr Ser Gly Tyr Cys Tyr Cys His Asp Ser Gly Thr Pro
                35                  40                  45

Asp Arg Cys
```

<210> 25
<211> 252
<212> DNA
<213> Arenicola marina

<220>
<221> CDS
<222> (1)..(252)

<220>
<221> sig_peptide
<222> (1)..(63)

<220>
<221> mat_peptide
<222> (121)..(252)

<400> 25

```
atg aca acg ctg ctc cta atg ttt ggg gtc gcc atc gtc gtc ctg gca        48
Met Thr Thr Leu Leu Leu Met Phe Gly Val Ala Ile Val Val Leu Ala
-40                 -35                 -30                 -25

act gta gca gcc gtg cct cta caa gat gac gat gtc ctt gac cca tca        96
Thr Val Ala Ala Val Pro Leu Gln Asp Asp Asp Val Leu Asp Pro Ser
                -20                 -15                 -10

ggt ctt gtg gaa ggg gaa aca cgg ggc agg tcg tgt aat ttc tgg ttg       144
Gly Leu Val Glu Gly Glu Thr Arg Gly Arg Ser Cys Asn Phe Trp Leu
            -5                  -1  1                   5

tgc cac gca tcc tgc atc gct aaa gtc gct gaa cga ggc tgc tgt gga       192
Cys His Ala Ser Cys Ile Ala Lys Val Ala Glu Arg Gly Cys Cys Gly
    10                  15                  20

gtt gga aac tac ctt ggc tac tgt tat tgt tat gac agt ggc tat gaa       240
Val Gly Asn Tyr Leu Gly Tyr Cys Tyr Cys Tyr Asp Ser Gly Tyr Glu
25                  30                  35                  40

tac agg tgc cgc                                                      252
Tyr Arg Cys Arg
```

<210> 26
<211> 84
<212> PRT
<213> Arenicola marina

<400> 26

```
Met Thr Thr Leu Leu Leu Met Phe Gly Val Ala Ile Val Val Leu Ala
-40                 -35                 -30                 -25

Thr Val Ala Ala Val Pro Leu Gln Asp Asp Asp Val Leu Asp Pro Ser
                -20                 -15                 -10

Gly Leu Val Glu Gly Glu Thr Arg Gly Arg Ser Cys Asn Phe Trp Leu
            -5                  -1  1                   5

Cys His Ala Ser Cys Ile Ala Lys Val Ala Glu Arg Gly Cys Cys Gly
    10                  15                  20

Val Gly Asn Tyr Leu Gly Tyr Cys Tyr Cys Tyr Asp Ser Gly Tyr Glu
25                  30                  35                  40

Tyr Arg Cys Arg
```

<210> 27
<211> 321
<212> DNA
<213> Arenicola marina

<220>
<221> CDS
<222> (1)..(321)

<220>
<221> sig_peptide
<222> (1)..(66)

<220>
<221> mat_peptide
<222> (145)..(321)

<400> 27

```
atg aag ttg ttc tta ccc ttg gtg gtc acc gtg gca ttt gcc gcc gtc        48
Met Lys Leu Phe Leu Pro Leu Val Val Thr Val Ala Phe Ala Ala Val
            -45                 -40                 -35

gtc atg act gat acg gag cca gtt gag gct gag gat gaa gtt tcc atc        96
Val Met Thr Asp Thr Glu Pro Val Glu Ala Glu Asp Glu Val Ser Ile
        -30                 -25                 -20

atg aac cct tat gta aaa gac ggt gtg caa ggg aga cag acc ctc gag       144
Met Asn Pro Tyr Val Lys Asp Gly Val Gln Gly Arg Gln Thr Leu Glu
    -15                 -10                 -5                  -1

ggt ggg atg tgt ggt gac gac ttt ggc tcg tgg tct gac atc cgt gac       192
Gly Gly Met Cys Gly Asp Asp Phe Gly Ser Trp Ser Asp Ile Arg Asp
1               5                   10                  15

aac cag tgc agg gat tat tgt cgt tcg agg ggt agc ttc gga ggt tgc       240
Asn Gln Cys Arg Asp Tyr Cys Arg Ser Arg Gly Ser Phe Gly Gly Cys
            20                  25                  30

tgt ggt att ggt ctc tgg cag gga caa tgt tac tgc tat tac ttc tac       288
Cys Gly Ile Gly Leu Trp Gln Gly Gln Cys Tyr Cys Tyr Tyr Phe Tyr
        35                  40                  45

gga ccc agt ctt acc tat agg tgc tgg tcg aaa                           321
Gly Pro Ser Leu Thr Tyr Arg Cys Trp Ser Lys
    50                  55
```

<210> 28
<211> 107
<212> PRT
<213> Arenicola marina

<400> 28

```
Met Lys Leu Phe Leu Pro Leu Val Val Thr Val Ala Phe Ala Ala Val
            -45                 -40                 -35

Val Met Thr Asp Thr Glu Pro Val Glu Ala Glu Asp Glu Val Ser Ile
        -30                 -25                 -20

Met Asn Pro Tyr Val Lys Asp Gly Val Gln Gly Arg Gln Thr Leu Glu
```

-15 -10 -5 -1

Gly Gly Met Cys Gly Asp Asp Phe Gly Ser Trp Ser Asp Ile Arg Asp
1 5 10 15

Asn Gln Cys Arg Asp Tyr Cys Arg Ser Arg Gly Ser Phe Gly Gly Cys
20 25 30

Cys Gly Ile Gly Leu Trp Gln Gly Gln Cys Tyr Cys Tyr Tyr Phe Tyr
35 40 45

Gly Pro Ser Leu Thr Tyr Arg Cys Trp Ser Lys
50 55

<210> 29
<211> 49
<212> DNA
<213> Artificial

<220>
<223> Marinasin1B-F

<400> 29
ggatgcgaac caacttcaga aacgtagctg gctatgtaac tggttgggc 49

<210> 30
<211> 34
<212> DNA
<213> Artificial

<220>
<223> Marinasin1B-R

<400> 30
cccaagcttc acatggtgta tggttgatct ctcc 34

<210> 31
<211> 49
<212> DNA
<213> Artificial

<220>
<223> Marinasin2A-F

<400> 31
ggatgcgaac caacttcaga aacgtggttg gtgctggcag tggacatgt 49

<210> 32
<211> 31
<212> DNA
<213> Artificial

<220>
<223> Marinasin2A-R

<400> 32
cccaagcttg cccttctagc gttcagctct t 31

<210> 33
<211> 49
<212> DNA
<213> Artificial

<220>
<223> Marinasin3B-F

<400> 33
ggatgcgaac caacttcaga aacgtcattg gtgtttcgag tggtcatgt 49

<210> 34
<211> 31
<212> DNA
<213> Artificial

<220>
<223> Marinasin3B-R

<400> 34
cccaagcttt cagtggagaa ccgttacagc a 31

<210> 35
<211> 49
<212> DNA
<213> Artificial

<220>
<223> Marinasin4-F

<400> 35
ggatgcgaac caacttcaga aacgtatccc ctgttggact ccgacatgt 49

<210> 36
<211> 30
<212> DNA
<213> Artificial

<220>
<223> Marinasin4-R

<400> 36
cccaagctta gcccagtatg ctctgcacgt 30

<210> 37
<211> 49
<212> DNA
<213> Artificial

<220>
<223> Marinasin5-F

<400> 37
ggatgcgaac caacttcaga aacgtggggg ccggccctgt cataggcat 49

<210> 38
<211> 32
<212> DNA
<213> Artificial

<220>
<223> Marinasin5-R

<400> 38
cccaagctta gttgttcgct ccattagcac ct 32

<210> 39
<211> 40
<212> DNA
<213> Artificial

<220>
<223> Marinasin6-F

<400> 39
accaacttca gaaacgtggc aggtcgtgta atttctggtt 40

<210> 40
<211> 32
<212> DNA
<213> Artificial

<220>
<223> Marinasin6-R

<400> 40
cccaagcttt gggcgattct atctgcctct ta 32

<210> 41
<211> 37
<212> DNA
<213> Artificial

<220>
<223> Marinasin7-F

<400> 41
accaacttca gaaacgtggt gggatgtgtg gtgacga 37

<210> 42
<211> 29
<212> DNA
<213> Artificial

<220>
<223> Marinasin7-R

<400> 42
cccaagcttg cacatcgttt ccaccgcaa 29

<210> 43
<211> 36
<212> DNA
<213> Artificial

<220>
<223> PlecBHI-F

<400> 43
cgcggatccc accatgcaat ttaccaccat cctctc 36

<210> 44
<211> 49

<212> DNA
<213> Artificial

<220>
<223> Plec-Mar1B-R

<400> 44
gcccaaccag ttacatagcc agctacgttt ctgaagttgg ttcgcatcc 49

<210> 45
<211> 49
<212> DNA
<213> Artificial

<220>
<223> Plec-Mar2A-R

<400> 45
acatgtccac tgccagcacc aaccacgttt ctgaagttgg ttcgcatcc 49

<210> 46
<211> 49
<212> DNA
<213> Artificial

<220>
<223> Plec-Mar3B-R

<400> 46
acatgaccac tcgaaacacc aatgacgttt ctgaagttgg ttcgcatcc 49

<210> 47
<211> 38
<212> DNA
<213> Artificial

<220>
<223> Plec-Mar4-R

<400> 47
tccaacaggg gatacgtttc tgaagttggt tcgcatcc 38

<210> 48
<211> 49
<212> DNA
<213> Artificial

<220>
<223> Plec-Mar5-R

<400> 48
atgcctatga cagggccggc ccccacgttt ctgaagttgg ttcgcatcc 49

<210> 49
<211> 42
<212> DNA
<213> Artificial

<220>
<223> Plec-Mar6-R

<400> 49

aaattacacg acctgccacg tttctgaagt tggttcgcat cc 42

<210> 50
<211> 38
<212> DNA
<213> Artificial

<220>
<223> Plec-Mar7-R

<400> 50
cacacatccc accacgtttc tgaagttggt tcgcatcc 38