

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
20 February 2003 (20.02.2003)

PCT

(10) International Publication Number
WO 03/014374 A2

(51) International Patent Classification⁷: **C12Q**

(21) International Application Number: PCT/US02/25202

(22) International Filing Date: 7 August 2002 (07.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/310,496 8 August 2001 (08.08.2001) US

(71) Applicants: **UNIVERSITY OF UTAH RESEARCH FOUNDATION** [US/US]; Suite 110, 615 Arapeen Drive, Salt Lake City, UT 84108 (US). **COGNETIX, INC.** [US/US]; Suite 201, 421 Wakara Way, Salt Lake City, UT 84108 (US).

(72) Inventors: **GARRETT, James, E.**; 1584 East 3150 South, Salt Lake City, UT 84106 (US). **BANDYOPADHYAY, Pradip, K.**; 1152 East Bryan Avenue, Salt Lake City, UT 84105 (US).

(74) Agents: **IHNEN, Jeffrey, L.** et al.; Rothwell, Figg, Ernst & Manbeck, P.C., Suite 800, 1425 K Street, N.W., Washington, DC 20005 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CONUS γ -CARBOXYLASE

(57) Abstract: The present invention relates to a γ -carboxylase from *Conus* snails, a nucleic acid sequence encoding the *Conus* γ -carboxylase and to a method for using the nucleic acid or protein sequences for preparing γ -carboxylated proteins.



WO 03/014374 A2

TITLE OF THE INVENTION*CONUS* γ -CARBOXYLASE

[0001] This invention was made with Government support under Grant No. PO1 GM48677
5 awarded by the National Institute of General Medical Sciences, National Institutes of Health, Bethesda, Maryland. The United States Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[0002] The present invention relates to a γ -carboxylase from *Conus* snails, a nucleic acid
10 sequence encoding the *Conus* γ -carboxylase and to a method for using the nucleic acid or protein sequences for preparing γ -carboxylated proteins.

[0003] The publications and other materials used herein to illuminate the background of the invention, and in particular, cases to provide additional details respecting the practice, are incorporated by reference, and for convenience are referenced in the following text by number and
15 are listed numerically in the appended Bibliography.

[0004] The vitamin K-dependent γ -carboxylation of glutamate residues was originally discovered as a novel post-translational modification in the blood coagulation cascade (Stenflo et al., 1974); some of the key clotting factors such as prothrombin must be γ -carboxylated in order for proper blood clotting to occur. Somewhat later, this post-translational modification was also found
20 in certain bone proteins (Price and Williamson, 1985). In mammalian blood coagulation and bone Gla proteins, γ -carboxylation of glutamate residues is carried out by a vitamin K-dependent carboxylase. A conserved motif (Price et al., 1987) γ -carboxylation recognition sequence in the propeptide sequence binds the γ -carboxylase and is required for a polypeptide substrate to be a high affinity target for the γ -carboxylase.

[0005] This modification was restricted to these rather specialized mammalian systems until
25 a very unusual peptide, conantokin-G, was described from the venom of the predatory marine snail, *Conus geographus* (McIntosh et al., 1984). Conantokin-G is a 17-amino acid peptide that inhibits the N-methyl-D-aspartate receptor (Olivera et al., 1990). Unlike most *Conus* peptides, which are multiply disulfide-bonded, conantokin-G has no disulfide cross-links but has five residues of
30 γ -carboxyglutamate residues; this remains the highest density of γ -carboxyglutamate found in any functional gene product characterized to date. Most of the biologically active components of the *Conus* venom are multiply disulfide bonded peptides (the conotoxins). These have been shown to be initially translated as prepropeptide precursors, which are then post-translationally processed to

yield the mature disulfide-crosslinked conotoxin. Conantokin-G differs strikingly from most conotoxins not only in having γ -carboxyglutamate residues, but also because it has no disulfide crosslinks. U.S. Patent No. 6,197,535 describes the analysis of the conantokin-G precursor and sequence recognition by a γ -carboxylase for the maturation of the functional conantokin-G peptide.

5 It was found a γ -carboxylation recognition sequence is included in the -1 to -20 region of the conantokin-G prepropeptide. This sequence appears to increase the affinity of the *Conus* carboxylase by approximately two orders of magnitude.

[0006] The presence of γ -carboxyglutamate in a non-mammalian system was initially controversial because vitamin K-dependent carboxylation of glutamate residues had primarily been
10 thought to be a highly specialized mammalian innovation. However, conantokin-G is only one member of a family of peptides; a variety of other conantokins have been found including conantokin-T and conantokin-R from two other fish-hunting cone snails (Haack et al., 1990; White et al., 1997). All three peptides have a high content of γ -carboxyglutamate (4-5 residues). γ -Glutamyl carboxylase has been purified from mammalian sources (Wu et al., 1991a; Berkner et
15 al., 1992), has been expressed both in mammalian and insect cell lines (Wu et al., 1991b; Roth et al., 1993) and has been purified from *Conus* (U.S. Patent No. 6,197,535). Recently it was shown that, as is the case in the mammalian system, the carboxylation reaction in *Conus* venom ducts absolutely requires vitamin K, and the net carboxylation increases greatly in the presence of high concentrations of ammonium sulfate. In these respects, the mammalian and the *Conus* γ -
20 carboxylation venom systems are very similar (Stanley et al., 1997).

[0007] Knobloch and Suttie (1987) and Cheung et al. (1989) found that the propeptide sequences of Factors IX and X at micromolar concentrations stimulated the carboxylation of oligopeptide substrates, suggesting a probable positive allosteric effector role. In addition, the propeptide at micromolar concentrations acted as a competitive inhibitor of carboxylation of a
25 substrate whose sequences were based on residues - 18 to + 10 of prothrombin (Ulrich et al., 1988). Similarly, the *Conus* propeptide (-20 to -1) inhibits the carboxylation of propeptide-containing substrates, (e.g., -10.Pro-E.Con-G and -20.Pro-E.Con-G) (U.S. Patent No. 6,197,535).

[0008] The orientation in which a Glu presents itself to the active site of the carboxylase may determine whether it will be carboxylated. In the case of Con-G not all the Glu residues are
30 γ -carboxylated (e.g., Glu² is not carboxylated, whereas Glu³ and Glu⁴ are carboxylated). The solution structures of Con-G and Con-T as determined by CD and NMR spectroscopy (Skjaebaek

et al., 1997; Warder et al., 1997) are a mixture of α and 3_{10} helices. Rigby et al. (1997) also determined the structure of the metal-free conformer of conantokin-G by NMR spectroscopy. In all of these structures, the Glu residues are on the same side of the conantokin structure; this would allow a membrane-bound enzyme to carry out efficient carboxylation of Glu residues oriented in the same direction with optimum stereochemistry.

[0009] There is a need in the art to identify the nucleic acid sequence encoding *Conus* γ -carboxylase, to identify the sequence of *Conus* γ -carboxylase, and to use the nucleic acids or proteins in the production of γ -carboxylated proteins.

10 SUMMARY OF THE INVENTION

[0010] The present invention is relates to a γ -carboxylase from *Conus* snails, a nucleic acid sequence encoding the *Conus* γ -carboxylase and to a method for using the nucleic acid or protein sequences for preparing γ -carboxylated proteins.

[0011] Thus, one aspect of the invention is directed to the amino acid sequence of *C. textile* γ -carboxylase. The amino acid sequence of *C. textile* γ -carboxylase is set forth in SEQ ID NO:2 or SEQ ID NO:4.

[0012] A second aspect of the invention is directed to a nucleic acid encoding a *C. textile* γ -carboxylase. A preferred nucleotide sequence of the nucleic acid is set forth in SEQ ID NO:1 or SEQ ID NO:3.

[0013] A third aspect of the invention is directed to amino acid sequences and nucleic acid sequences of other *Conus* γ -carboxylases, as well as amino acid sequences and nucleic acid sequences having 95% identity with the disclosed sequences.

[0014] A fourth aspect of the invention is directed to vectors containing the γ -carboxylase encoding nucleic acid.

[0015] A fifth aspect of the invention is directed to host cells containing an expression cassette with the γ -carboxylase encoding nucleic acid.

[0016] A sixth aspect of the invention is directed to host cells containing an expression cassette with the γ -carboxylase encoding nucleic acid sequence and an expression cassette with a nucleic acid sequence encoding a protein which is γ -carboxylated. Such proteins include conantokins and other vitamin K-dependent proteins.

[0017] A seventh aspect of the invention is directed to the use of a γ -carboxylase for the preparation of γ -carboxylated proteins (the term used herein to refer to proteins which are γ -carboxylated), such as conantokins and other vitamin K-dependent proteins.

[0018] An eighth aspect of the invention is directed to the use of a γ -carboxylase nucleic acid for the preparation of γ -carboxylated proteins, such as conantokins and other vitamin K-dependent proteins.

DETAILED DESCRIPTION OF THE INVENTION

[0019] The present invention relates to a γ -carboxylase from *Conus* snails, a nucleic acid sequence encoding the *Conus* γ -carboxylase and to a method for using the nucleic acid or protein sequences for preparing γ -carboxylated proteins.

[0020] In one aspect, the present invention relates to the amino acid sequence of *C. textile* γ -carboxylase. The amino acid sequence of *C. textile* γ -carboxylase is set forth in SEQ ID NO:2 or SEQ ID NO:4. In a further embodiment, the present invention relates to a γ -carboxylase which has at least 95% identity with the amino acid sequence set forth in SEQ ID NO:2 or SEQ ID NO:4 and which has γ -carboxylation activity. The γ -carboxylation activity can be assayed as described herein to identify those proteins having the proper biological activity.

[0021] In a second aspect, the present invention relates to a nucleic acid encoding a *C. textile* γ -carboxylase. A preferred nucleic acid sequence is set forth in SEQ ID NO:1 or SEQ ID NO:3. In a further embodiment, the present invention relates to a γ -carboxylase encoding nucleic acid which has at least 95% identity with the nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID NO:3. The encoded γ -carboxylase has γ -carboxylation activity which can be assayed as described herein to identify those nucleic acids which encode proteins having the proper biological activity.

[0022] In a third aspect, the present invention relates to vectors containing the nucleic acid encoding a γ -carboxylase of the present invention. In one embodiment, the vector is an expression vector.

[0023] In a fourth aspect, the present invention relates to host cells containing an expression cassette or expression vector with the γ -carboxylase encoding nucleic acid of the present invention. The host cells produce the γ -carboxylase when grown under suitable growth conditions.

[0024] In a fifth aspect, the present invention relates to host cells containing an expression cassette or expression vector with the γ -carboxylase encoding nucleic acid of the present invention and an expression cassette with a nucleic acid sequence encoding a protein which is γ -carboxylated. Such proteins include conantokins and other vitamin K-dependent proteins.

5 [0025] In a sixth aspect, the present invention relates to the use of a γ -carboxylase of the present invention for the preparation of γ -carboxylated proteins, such as conantokins and other vitamin K-dependent proteins.

[0026] In a seventh aspect, the present invention relates to the use of a γ -carboxylase encoding nucleic acid of the present invention for the preparation of γ -carboxylated proteins, such
10 as conantokins and other vitamin K-dependent proteins.

[0027] A nucleic acid or fragment thereof has substantial identity with another if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least
15 about 90%, and more preferably at least about 95-98% of the nucleotide bases. A protein or fragment thereof has substantial identity with another if, optimally aligned, there is an amino acid sequence identity of at least about 30% identity with an entire naturally-occurring protein or a portion thereof, usually at least about 70% identity, more usually at least about 80% identity, preferably at least about 90% identity, and more preferably at least about 95-98% identity.

20 [0028] Identity means the degree of sequence relatedness between two polypeptide or two polynucleotides sequences as determined by the identity of the match between two strings of such sequences, such as the full and complete sequence. Identity can be readily calculated. While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (*Computational Molecular Biology*,
25 Lesk, A. M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D. W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991).
30 Methods commonly employed to determine identity between two sequences include, but are not limited to those disclosed in *Guide to Huge Computers*, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H., and Lipman, D., *SIAM J. Applied Math.* 48: 1073 (1988). Preferred

methods to determine identity are designed to give the largest match between the two sequences tested. Such methods are codified in computer programs. Preferred computer program methods to determine identity between two sequences include, but are not limited to, GCG (Genetics Computer Group, Madison Wis.) program package (Devereux, J., et al., *Nucleic Acids Research* 12(1). 387 (1984)), BLASTP, BLASTN, FASTA (Altschul et al. (1990); Altschul et al. (1997)). The well-known Smith Waterman algorithm may also be used to determine identity.

[0029] As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence of is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5 or 3 terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

[0030] Alternatively, substantial homology or (similarity) exists when a nucleic acid or fragment thereof will hybridize to another nucleic acid (or a complementary strand thereof) under selective hybridization conditions, to a strand, or to its complement. Selectivity of hybridization exists when hybridization which is substantially more selective than total lack of specificity occurs. Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 14 nucleotides, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90%. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will often be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides.

[0031] Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as

will be readily appreciated by those skilled in the art. Stringent temperature conditions will generally include temperatures in excess of 30°C, typically in excess of 37°C, and preferably in excess of 45°C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. However, the combination of parameters is much more
5 important than the measure of any single parameter. The stringency conditions are dependent on the length of the nucleic acid and the base composition of the nucleic acid, and can be determined by techniques well known in the art. See, e.g., Asubel, 1992; Wetmur and Davidson, 1968.

[0032] Thus, as herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. Such
10 hybridization techniques are well known to those of skill in the art. Stringent hybridization conditions are as defined above or, alternatively, conditions under overnight incubation at 42° C in a solution comprising: 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65° C.

[0033] The terms "isolated", "substantially pure", and "substantially homogeneous" are used interchangeably to describe a protein or polypeptide which has been separated from components which accompany it in its natural state. A monomeric protein is substantially pure when at least about 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure protein will typically comprise about 60 to 90% W/W of a protein sample, more usually about 95%, and
20 preferably will be over about 99% pure. Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art which are utilized for purification.

[0034] Large amounts of the nucleic acids of the present invention may be produced by (a) replication in a suitable host or transgenic animal or (b) chemical synthesis using techniques well known in the art. Nucleic acids made by either of these techniques are also referred to as synthetic nucleic acids herein. Constructs prepared for introduction into a prokaryotic or eukaryotic host may comprise a replication system recognized by the host, including the intended polynucleotide
30 fragment encoding the desired polypeptide, and will preferably also include transcription and translational initiation regulatory sequences operably linked to the polypeptide encoding segment.

Such vectors may be prepared by means of standard recombinant techniques well known in the art. See for example, see Ausbel (1992); Sambrook and Russell (2001); and U.S. Patent No. 5,837,492.

[0035] Large amounts of the protein of the present invention may be produced by (a) expression in a suitable host or transgenic animal or (b) chemical synthesis using techniques well known in the art. Proteins acids made by either of these techniques are also referred to as synthetic proteins herein. Expression vectors may include, for example, an origin of replication or autonomously replicating sequence (ARS) and expression control sequences, a promoter, an enhancer and necessary processing information sites, such as ribosome-binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, and mRNA stabilizing sequences.

Secretion signals may also be included where appropriate which allow the protein to cross and/or lodge in cell membranes, and thus attain its functional topology, or be secreted from the cell. Such vectors may be prepared by means of standard recombinant techniques well known in the art. See for example, see Ausbel (1992); Sambrook and Russell (2001); and U.S. Patent No. 5,837,492.

[0036] The γ -carboxylase of the present invention is isolated following expression in a suitable host or chemical synthesis using techniques well known in the art. The isolated γ -carboxylase of the present invention is used to γ -carboxylate γ -carboxylated proteins, such as conantokins and other vitamin K-dependent proteins. The γ -carboxylase is contacted with the pro-protein which contains the γ -carboxylation recognition sequence and allowed to γ -carboxylate the protein. The γ -carboxylated protein is isolated and purified using techniques well known in the art.

[0037] The nucleic acid encoding γ -carboxylase of the present invention is used to γ -carboxylate γ -carboxylated proteins, such as conantokins and other vitamin K-dependent proteins, *in vivo* using techniques well known in the art. In one embodiment, a suitable host is prepared which contains an expression vector containing a γ -carboxylase encoding nucleic acid of the present invention and an expression vector containing a nucleic acid encoding a γ -carboxylated protein, such as conantokin and other vitamin K-dependent protein. Nucleic acids encoding conantokins are well known in the art. See U.S. Patent No. 6,172,041. Nucleic acids encoding other vitamin K-dependent proteins are also well known in the art. In a second embodiment, a suitable host is prepared which contains an expression vector containing a γ -carboxylase encoding nucleic acid and a nucleic acid encoding a γ -carboxylated protein, such as conantokin and other vitamin K-dependent protein. In either embodiment, the host cells are grown under conditions suitable for growth and expression of

the γ -carboxylase and the γ -carboxylated protein. The γ -carboxylase acts on the γ -carboxylated protein *in vivo* to properly γ -carboxylate the Glu residues in the protein.

EXAMPLES

[0038] The present invention is described by reference to the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described below were utilized.

EXAMPLE 1

In vitro γ -Carboxylation of ConG with γ -Carboxylase

[0039] In order to ascertain the fidelity of γ -carboxylation, it is essential to determine if the appropriate Glu residues are being modified and if the modification goes to completion. ConG, -20pro-ConG and pro-ConG are used as substrates to determine fidelity of carboxylation *in vitro* with the *Conus* γ -carboxylase produced in accordance with the present invention as described in U.S. Patent No. 6,197,535. The identification of Gla by routine amino acid sequencing is not efficient due to poor recovery of Gla residues. In the present modified method, Gla-containing peptides are decarboxylated by heating under vacuum. The Gla residues are converted to Glu which can then be sequenced. If $^{14}\text{CO}_2$ is incorporated in the γ -carboxylation reaction, half of the molecules in the decarboxylated product will contain $^{14}\text{CO}_2$ covalently linked to the γ -C of modified Glu residues. Thus, by monitoring the radioactivity recovered at each step of the sequencing reaction, the position of Gla residues can be determined.

[0040] Products of the γ -carboxylation reaction are purified using a Waters Oasis™ HLB Extraction Cartridge followed by reversed phase HPLC using Vydac C₁₈ column. The [^{14}C]-containing fractions are dried and sequenced chemically with concomitant determination of radioactivity at each position in the sequence. When -20.pro-ConG and pro-ConG are used as the substrate, the radioactivity-containing fraction from the reversed phase HPLC column are dried and digested with endoproteinase Lys C. This is done to reduce the length of the peptide for sequencing without interfering with the identification of the γ -carboxylated residues. The Lys C digest is purified using a Waters Oasis™ HLB Extraction Cartridge followed by reversed phase HPLC using

Vydac C₁₈ column. The radioactivity eluted as a single peak coincidental with the A₂₂₀ peak. The chemical sequence of the material had the expected sequence.

[0041] The Gla determinations are carried out on a mixture of unmodified and variously modified substrate molecules. On the basis of these experiments, it is not possible to assign Gla residues to individual post-translationally modified substrate molecules. However, an average picture emerged. As in the case of the native product, Glu² is not carboxylated. The rest of the Glu residues are carboxylated.

EXAMPLE 2

10 Cloning of *Conus textile* γ -Carboxylase cDNA

[0042] The full-length *C. textile* γ -carboxylase cDNA was isolated by reverse transcription-PCR (RT-PCR) of venom duct RNA, using primers designed from conserved regions of mammalian γ -carboxylase proteins. A number of different internal primer sets were utilized to generate overlapping segments of the *C. textile* γ -carboxylase sequence. Most of the internal region of the *C. textile* cDNA was obtained in this manner, generating sequence to within ~100 amino acids of the putative N- and C-termini of the protein. To obtain the ends of the cDNA sequence, nested PCR primers based on the *C. textile* cDNA sequence were used in 5' and 3' RACE to identify the transcription start site and poly A termination site, respectively. Merging of these overlapping segments generated a ~2460 bp cDNA sequence with a single long open reading frame encoding a protein of 799 amino acids. This *C. textile* protein has substantial homology to the mammalian γ -carboxylase. Overall homology of the *Conus* sequence to the mammalian enzymes is ~50%, although distinct regions within the protein show substantially higher and lower levels of sequence conservation. The RT-PCR with degenerate primers consistently identified the sequence presented here, with no evidence for any other related γ -carboxylase isoform being expressed in the *C. textile* venom duct. Much of the degenerate PCR used to initially clone the *C. textile* sequence employed non-proofreading thermostable polymerases and many cycles of PCR, both conditions that could introduce minor sequence errors. To obtain an accurate sequence, primers designed to the very most 5' and 3' ends of the *C. textile* cDNA sequence were used to amplify the full-length 2460 bp cDNA using a proofreading thermostable polymerase mixture and a minimal number of PCR cycles. This Long, Accurate-PCR generated the predicted 2460 bp product in good yield, with no evidence for alternative-splice isoforms of the *C. textile* γ -carboxylase mRNA being expressed in the venom

duct. The full-length cDNA product was cloned and completely sequenced on both strands, yielding the sequence presented here.

[0043] The *C. textile* γ -carboxylase cDNA has a short 5' untranslated region of ~50 bp, and the first ATG start codon encountered initiates the long open reading frame encoding the γ -carboxylase protein. The cDNA sequence obtained by 3' RACE terminates in a poly A tail, and this is preceded by a typical poly A addition signal (AATAA). An unusual feature is that the open reading frame lacks a typical stop codon (TAG, TAG, TGA) and instead continues into the poly A tail. We considered the possibility that the 3' RACE technique had not identified the true 3' end of the mRNA, or that a sequencing error obscured a stop codon that is present, but various experiments tend to rule this out. A variety of 3' RACE experiments have been performed, using different PCR primers and cDNA preparations with different 3' adapters, yet all of these experiments consistently identify the same 3' poly A site. Numerous different PCR product clones representing this 3' region have been thoroughly sequenced, and there are no ambiguities on the sequencing chromatograms that would suggest a sequencing anomaly that could change the open reading frame.

[0044] Finally, 3' RACE was used to isolate the corresponding region of γ -carboxylase cDNA from the venom duct RNA of two other *Conus* species, *omaria* and *episcopatus*, that are snail-hunting species related to *textile*. The 3' RACE identified poly A sites at essentially the same location in all three different species. The DNA sequences (and corresponding protein sequence) were highly homologous between the three species, but there is sequence variation as expected between species. Even though the sequence varies between the species, in all three the open reading frame lacks a typical stop codon and extends into the poly A tail. This suggests that our initial finding was not just a cloning or sequence artifact restricted to *C. textile*, but is a conserved feature of the γ -carboxylase mRNA across related *Conus* species. It is possible that *Conus* recognizes an atypical triplet as a termination codon, or that post-translation processing generates the true C-terminus of the *Conus* γ -carboxylase enzyme. Assuming that it terminates at the poly A tail, the size of the *Conus* γ -carboxylase protein that we have identified (799 amino acids) is very similar to that of the mammalian γ -carboxylase enzymes, so the overall size of the proteins appears to be conserved.

[0045] The nucleic acid sequence (SEQ ID NO:1 or 3) and amino acid sequence (SEQ ID NO:2 or 4) for *C. textile* γ -carboxylase are set forth in Tables 1 and 2, respectively. The 3' nucleic

acid sequence (SEQ ID NO:5 or 7) and C-terminal amino acid sequence (SEQ ID NO:6 or 8) for *C. omaria* are set forth in Tables 3 and 4, respectively. The 3' nucleic acid sequence (SEQ ID NO:9 or 11) and C-terminal amino acid sequence (SEQ ID NO:10 or 12) for *C. episcopatus* are set forth in Tables 5 and 6, respectively.

5 [0046] The sequences for *C. omaria* and *C. episcopatus* represent about 200-220 amino acids at the C-terminus, starting at a position corresponding to approximately 590 of *C. textile*. Both of these sequences have several base pair and amino acid changes compared to the *C. textile* sequence as is expected for species homologs, although the sequence identity is quite high. All three species have the poly A tail in roughly the same location (the *C. textile* sequence is about 30 bp
10 longer) and none of the three species has a typical stop codon.

TABLE 1

Nucleic Acid Sequence of *C. textile* γ -Carboxylase

ATCTTTGTGAGCGTGATCCATCGCACAAACCATGCAAAGGCCAGGCAAGAAAGTGGCTGCTGATTCAGAG
15 GAATCAAATGACATCAGCCAACAAGCAGAAAACAGAGACCAGCTCCTCCCCCAGGAAGCCAGTCCCAAAG
CGTGTGAGGAAGAGGACACAGAGGATGAAGAGGAAGAAGAGGACAAGTTCTACAAACTCTTTGGTTTCAG
CTTGAGCGACCTCAAGTCATGGGACAGCTTTGTTTCGTCTGTTGTGCGCGCCCGCTGACCCTGCCGGTCTG
GCTTATATCCGTGTCACCTTATGGGTTTTTGATGATGTGGGACGTGTTTGAGGAAAGGGGCCCTGTCCCGTG
20 CAGATATGCGATGGGGTGATGATGAGGCATGACAGGTTTCCTCTCTTCGACTTCATGCAACCCCTGCCCT
GCACATGATGGTCCTGCTGTACCTGATCATGCTGATTGGAACAGGAGGAATTCATTAGGAGCCAAGTAC
CGTGTGTGCTGCGTTATGCACCTGCTGCCCTACTGGTACATAGTGCTTCTGGACGAGTGACAGTTGGAACA
ATCACTCCTATCTGTTTGGTCTCCTCTCTTCTCCTCTGCTTTGCGATGCTAACCCTACTGGTCCAT
GGACGGTCTGTTCAATGCCAAGGTTTCGAAATACGGATGTTCCCTTGTGGAACCTACACCCTCCTACGTACA
CAGGTGTTTCTGGTGTACTTTTTGGCTGGGCTGAAGAACTGGACATGGACTGGATCGCTGGTTACTCCA
25 TGGGCCGTTTGTGAGTGATCATTGGGTCTTTTACCCGTTTACGTTTCTGATGACAGAAGACCAGGTGAGTGT
GCTTGTGGTCCACCTGGGTGGACTTGCCATTGACTTGTTCTGTTGGGCTACCTGCTCTTCTTTGACAAGACA
CCACCGATCGGTGTCATTATCAGTTCGTCACTCCACCTGATGAATGCACAGATGTTTACGATAGGAATGT
TTCCGTATGCCATGTTGGGTTTGACGCCTGTGTTCTTCTATGCCAACTGGCCGAGGGCCCCGTTTCGCCG
30 CATTCACGATCCTTGAGGATTCTTACCCCTGATGATGGAGAGGATGATACGCTGCCTTCGGAGAAGTGC
TTATACACAAAAGAACAGGCCAAACCAGAACTGGCCAGCACCCCTGAGCATGAAAACACTGCAGTCCGCA
AACAGTTGACACCACCCACTCAGCCACGTTCCGGCATCATGCTGCCGCTGCCTTCACCGTTTTCTTCAT
TCTGTGGCAGATGTTTTTGCCTTTCTCTCATTTTATCACAAAGGGCAACAACAGCTGGACCCAGGGACTC
TACGGCTACTCCTGGGACATGATGGTTCACACCCGCAGCACTCAGCACACCAGGATCTCCTTCATCAACA
AGGACACAGGAGAGCGAGGGTTCTGGACCCGCAGGCATGGAGCAAGTCACATCGATGGGCGCATAACGC
35 TAAGATGATGAAGCAGTACGCCAGGTGCATCGCTCGCCGACTGAAGAAGCATGAAATCGACAATGTGGAA
ATCTATTTTGTATGTCTGGATATCTCTGAATCATCGCTTCCAGCAACGGATCGTGAACCCCAATGTGGACA
TTTTAACAGCCGAATGGAGTGTCTTTAAGTCCACTCCATGGATGATGCCCTTGCTGGTTCGACTTGTCTAA
TTGGCGAAGCAAGTTGAAAGAGATTGAGGACGACATTTTCAACTCAACCGACCTGTATGAAATAGTCTTT
CTGGCTGACTTTTCTGGTTTGTACCTGGGAACTTTGTCCACGGCAGCGTCGGGAGTCTCAACATCTCTG
40 TACTGCAGGGCCAGGTGGTGGTGGAGGTGCTTCCAGAGGAGGACAGTCTAGAAGAGCCCTACAACATCAG
CATCAGTGATGGCCAAGAGTCATTGATTCCACAGGGGTGTTCCACAAGGTGTACACAGTGTCTGAAGTG
CCCTCCTGTTACATGTACATCTACATGGTCACGGAAGAGACAGAGTTCCTTGAGAACTCAAAGAGCTGG
AACACGCCCTCAACGGCTCCCTGGATGCTCCAGTTCAGACAAGTTTGCCGAAGATCCTAAACTTGATCA
GTATATGGAGGTACTCAAACGAAGAATGCATTATATGCTATCGTGGACTGCAGTGATAAAAGCGC
45 CAGCTGTTTATGAGTTTTCTGAAAATGCATTATATGCTATCGTGGACTGCAGTGATAAAAGCGC
CCATGTGGTCCATGTACTCCGGGAATCTTACCAGAGTTCTTGAAGAACTGGAGCTACAGAAAATGCT
GGCGGAGAATGCCACCCTGGTGGCAAACGCCACCCAAGGGGTGAATAACACCCAGACGATGAACAACACC

TTGAACAACACCAAGGAGAAAGACAACACCCAAAGGGTTAACAAACCG (SEQ ID NO:1)

5 ATGCAAAGGCCAGGCAAGAAAGTGGCTGCTGATTTCAGAGGAATCAAATGACATCAGCCAACAAGCAGAAA
 ACAGAGACCAGCTCCTCCCCAGGAAGCCAGTCCCAAAGCGTGTGAGGAAGAGGACACAGAGGATGAAGA
 GGAAGAAGAGGACAAGTTCTACAAACTCTTTGGTTTCAGCTTGAGCGACCTCAAGTCATGGGACAGCTTT
 GTTCGTCTGTTGTCGCGCCCCGCTGACCCTGCCGGTCTGGCTTATATCCGTGTCACCTTATGGGTTTTTGA
 TGATGTGGGACGTGTTTGAGGAAAGGGCCTGTCCCGTGCAGATATGCGATGGGGTGATGATGAGGCATG
 10 CAGGTTTCTCTCTTCGACTTCATGCAACCCTTGCCCTGCACATGATGGTCCTGCTGTACCTGATCATG
 CTGATTGGAACAGGAGGAATTCTATTAGGAGCCAAGTACCGTGTGTGCTGCGTTATGCACCTGCTGCCCT
 ACTGGTACATAGTGCTTCTGGACGAGTGCAGTTGGAACAATCACTCCTATCTGTTTGGTCTCCTCTCTTT
 CCTCCTTCTGCTTTGCGATGCTAACCACTACTGGTCCATGGACGGTCTGTTCAATGCCAAGGTTTCGAAAT
 ACGGATGTTCCCTTGTGGAACACACCCTACTACGTACACAGGTGTTTCTGGTGTACTTTTTGGCTGGGC
 TGAAGAAACTGGACATGGACTGCTGCTGGTTACTCCATGGGCCGTTTGAGTGATCATTGGGTCTTTTA
 15 CCGTTTTACGTTCTTGATGACAGAAGACCAGGTGAGTGTGCTTGTGGTCCACCTGGGTGGACTTGCCATT
 GACTTGTTTCGTGGGCTACCTGCTCTTCTTTGACAAGACACGACCGATCGGTGTCATTATCAGTTCGTTCAT
 TCCACCTGATGAATGCACAGATGTTTCAGCATAGGAATGTTTCCGTATGCCATGTTGGGTTTGACGCCTGT
 GTTCTTCTATGCCAACTGGCCGAGGGCCCTGTTTCGCCGCATTCCACGATCCTTGAGGATTCTTACCCCT
 GATGATGGAGAGGATGATACGCTGCCTTCGGAGAAGTGCTTATACACAAAAGAACAGGCCAAACCAGAAC
 20 TGGCCAGCACCCCTGAGCATGAAAACACTGCAGTCCGCAAACAGTTGACACCACCCACTCAGCCCACGTT
 CCGGCATCATGCTGCCGCTGCCTTCACCGTTTTCTTCATTCTGTGGCAGATGTTTTTGCCTTCTCTCAT
 TTTATCACAAAGGGCAACAACAGCTGGACCCAGGGACTCTACGGCTACTCCTGGGACATGATGGTTCACA
 CCCGCAGCACTCAGCACACCAGGATCTCCTTCATCAACAAGGACACAGGAGAGCGAGGGTTCCTGGACCC
 GCAGGCATGGAGCAAGTCACATCGATGGGCGCATAACGCTAAGATGATGAAGCAGTACGCCAGGTGCATC
 25 GCTCGCCGACTGAAGAAGCATGAAATCGACAATGTGGAAATCTATTTTGATGTCTGGATATCTCTGAATC
 ATCGCTTCCAGCAACGGATCGTGAACCCCAATGTGGACATTTTAACAGCCGAATGGAGTGTCTTTAAGTC
 CACTCCATGGATGATGCCCTTGCTGGTCTGACTTGTCTAATTGGCGAAGCAAGTTGAAAGAGATTGAGGAC
 GACATTTTCAACTCAACCGACCTGTATGAAATAGTCTTTCTGGCTGACTTTCCTGGTTTTGTACCTGGAGA
 ACTTTGTCCACGGCAGCGTCGGGAGTCTCAACATCTCTGTACTGCAGGGCCAGGTGGTGGTGGAGGTGCT
 30 TCCAGAGGAGGACAGTCTAGAAGAGCCCTACAACATCAGCATCAGTGATGGCCAAGAGTCATTGATTCCC
 ACAGGGGTGTTCCACAAGGTGTACACAGTGTCTGAAGTGCCCTCCTGTTACATGTACATCTACATGGTCA
 CGGAAGAGACAGAGTTCCTTGAGAACTCAAAGAGCTGGAACACGCCCTCAACGGCTCCCTGGATGCTCC
 AGTTCCAGACAAGTTTGCCGAAGATCCTAACTTGATCAGTATATGGAGGTACTCAAACGAAGAATGCA
 ACTCCACCACCAACCTCTCAAGAGGAGCAAAGTTTCATACAGCTGTTTATGAGTTTTCTGAAAATGCATT
 35 ATATGTCTATGTATCGTGGACTGCAGCTGATAAAGGCGCCATGTGGTCCATGTACTCCGGGGAATCTTA
 CCGAGAGTTCTTGAAGAACTGGAGCTACAGAAAATGCTGGCGGAGAATGCCACCTGGTGGCAAACGCC
 ACCCAAGGGGTGAATAACACCCAGACGATGAACAAACACCTTGAACAACACCAAGGAGAAAGACAACACCC
 AAAGGGTTAACAAACCCAGAGAAAAGAGGCCCCCAGAAGGACAGACGCCCTAACAGCATCTCTGCAG
 AATGAGGGCGTCATGCCTCTGTTCTGATTGTAAATCTTCAATGTCAGACTCGCTGTCATGAGTCAGGAT
 40 GCCAAGGGTTGATTCTAAATGAAAAAA (SEQ ID NO:3)

TABLE 2

Protein Sequence of *C. textile* γ -Carboxylase

45 MQRPGKKVAADSEESNDISQQAENRDQLLPQEASPKACEEEDTEDEEEEDKFYKLFGLFSLSDLKSWDSF
 VRLSRPADPAGLAYIRVTYGFLLMMWDVFEERGLSRADMRWGDDEACRFPLDFMQPLPLHMMVLLYLIM
 LIGTGGILLGAKYRVCCVMHLLPYWYIVLLDECSWNNHSYLFGLLSFLLLLCDANHYWSMDGLFNAKVRN
 TDVPLWNYTLLRTQVFLVYFLAGLKKLDMDWIAGYSMGRLSHDHWVFYFPFTFLMTEDQVSVLVVHLGLAI
 DLFVGYLLFFDKTPPIGVIISSSFHLMNAQMFSIGMFPYAMLGLTPVFFYANWPRAPFRRI PRSLRILTP
 DDGEDDTLPSEKCLYTKEQAKPELASTPEHENTAVRKQLTPPTQPTFRHHAAAFTVFFILWQMFLPFSSH
 50 FITKGNSWTQGLYGYSWDMMVHTRSTQHTRISFINKDTGERGFLDPQAWSKSHRWAHNAKMMKQYARCI
 ARRLKKHEIDNVEIYFDVWISLNRHFQQRIVNPNVDILTAEWSVFKSTPWMMPLLDLSNWRSKLKEIED
 DIFNSTDLYEIVFLADFPGLYLENFVHGSVGLSLNISVLQGVVVEVLPEEDSLEOPYNISISDGQESLIP
 TGVFHKVYTVSEVPSCYMIYMVTEETEFLEKLKELEHALNGSLDAPVPDKFAEDPKLDQYMEVLKTKNA
 TTPPTSQEEQSFIQLFMSFLKMHYSMYRGLQLIKGAMWSMYSGESYREFLKKLELQKMLAENATLVANA
 55 TQGVNNTQTMNNTLNNTKEKDNTQRVNKP (SEQ ID NO:2)

5 MQRPGKKVAADSEESNDISQQAENRDQLLPQEASPKACEEEDTEDEEEEDKFKYKLFGFSLSDLKSWDSF
 VRLLSRPADPAGLAYIRVTYGFLLMMWDVFEERGLSRADMWRGDDDEACRFPLFDFMQPLPLHMMVLLYLIM
 LIGTGILLGAKYRVCCVMHLLPYWYIVLLDECSWNNHSYLFGLLSFLLLLCDANHYWSMDGLFNAKVRN
 TDVPLWNYTLLRTQVFLVYFLAGLKKLMDWIAGYSMGRSLDHWFYFPFTFLMTEDQVSVLVVHLGGLAI
 DLFVGYLFFDKTRPIGVIISSSFHLMNAQMFSIGMFPYAMLGLTPVFFYANWPRALFRRI PRSLRILTP
 DDGEDDTLPSEKCLYTKAQKPELASTPEHENTAVRKQLTPPTQPTFRHHAAAFTVFFILWQMFLPFSH
 FITKGNNSWTQGLYGYSWDMMVHTRSTQHTRISFINKDTGERGFLDPQAWSKSHRWAHNAKMMKQYARCI
 10 ARRLKKHEIDNVEIYFDVWISLNHRFQQRIVNPNVDILTAEWSVFKSTPWMMPLLVDLSNWRSKLKEIED
 DIFNSTDLYEIVFLADFPGLYLENFVHGSVGSLSNISVLQGQVVVEVLPEEDSLEEPYNISISDGQESLIP
 TGVFHKVYTVSEVPSCYMYIYMVTEETEFLEKLKELEHALNGSLDAPVPDKFAEDPKLDQYMEVLKTKNA
 TPPPTSQEEQSFIQLFMSFLKMHYMSMYRGLQLIKGAMWSMYSGESYREFLKKLELQKMLAENATLVANA
 TQGVNNTQTMNNTLNNTKEKDNTQRVNKPQEKKAPQKADSP (SEQ ID NO:4)

15

TABLE 33' Nucleic Acid Sequence of *C. omaria* γ -Carboxylase

GGGAGTCTCAACATCTCTGTACTGCAGGGCCAGGTGGTGGTGGAGGTGCTTCCAGAGGAGGACAGTCTAG
 AAAAACCCCTACAACATCAGCATCAATGATGGCCACGAGTCATTGATTCCCACAGGGGTATTCCACAAGGT
 20 GTACACAGTGTCTGAAGTGCCCTCCTGTTACATGTACATCTACATGGTCACGGAAGAGACGGAGTTCTTT
 GAGAACGTCAAAGAGCTGGAACACGCCCTCAACGGCTCCCTGGATGCTCCAGTTCCAGACAAGTTTGCCA
 AAGATCCTAAACTTGATCAATATATGGAGGTACTCAAAGTGAAGAATGCAGCTCCACCACCGGCCCTCG
 AGCGGAGAGAAGTTTCATAGAGCTGTTTATGAGTTTCTGAAAATGCATTATATGTCTATGTATCGTGGA
 CTGCAGCTGATAAAAGGCGCCGTGTGGTCCATGTACTCTGGGGAATCTTACCGAGAGTACCTGAAGGAAC
 25 TGGAATTACAGGCAATGCTGGGGGAGAATGCCACCCTGGTGGCAAACGCCACCCAAGGGGTGAATAACAC
 CCAGACGATGAACAACACCTTATTGAACAACACCAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO:5)

GGGAGTCTCAACATCTCTGTACTGCAGGGCCAGGTGGTGGTGGAGGTGCTTCCAGAGGAGGACAGTCTAG
 30 AAAAACCCCTACAACATCAGCATCAATGATGGCCACGAGTCATTGATTCCCACAGGGGTATTCCACAAGGT
 GTACACAGTGTCTGAAGTGCCCTCCTGTTACATGTACATCTACATGGTCACGGAAGAGACGGAGTTCTTT
 GAGAACGTCAAAGAGCTGGAACACGCCCTCAACGGCTCCCTGGATGCTCCAGTTCCAGACAAGTTTGCCA
 AAGATCCTAAACTTGATCAATATATGGAGGTACTCAAAGTGAAGAATGCAGCTCCACCACCGGCCCTCG
 AGCGGAGAGAAGTTTCATAGAGCTGTTTATGAGTTTCTGAAAATGCATTATATGTCTATGTATCGTGGA
 35 CTGCAGCTGATAAAAGGCGCCGTGTGGTCCATGTACTCTGGGGAATCTTACCGAGAGTACCTGAAGGAAC
 TGGAATTACAGGCAATGCTGGGGGAGAATGCCACCCTGGTGGCAAACGCCACCCAAGGGGTGAATAACAC
 CCAGACGATGAACAACACCTTATTGAACAACACCAAGGAGAAAAACAACCCCAAAGGGTTAACAAGCCG
 CAGGAAAAGAAGGCCCCCAGAAGGCAGACAGCCCCTAACAGCATCTCTGCAGAATGAGGGCATCATGCC
 TCTGTTCTGCATTTTAAATCTTCGATGTCAGACACGCTGTCATGAGTCAGGATGCCAAGGGTTGATTCTA
 40 AATGAAAAAA (SEQ ID NO:7)

TABLE 4C-Terminal Protein Sequence of *C. omaria* γ -Carboxylase

45 GSLNISVLQGQVVVEVLPEEDSLEKPYNISINDGHESLIPTGVFHKVYTVSEVPSCYMYIYMVTEETEFF
 ENVKELEHALNGSLDAPVPDKFAKDPKLDQYMEVLKVNAAPPPAPRAERSFIELFMSFLKMHYMSMYRG
 LQLIKGAVWSMYSGESYREYLKELELQAMLGENATLVANATQGVNNTQTMNNTLLNNTKKKKKK (SEQ
 ID NO:6)

50 GSLNISVLQGQVVVEVLPEEDSLEKPYNISINDGHESLIPTGVFHKVYTVSEVPSCYMYIYMVTEETEFF
 ENVKELEHALNGSLDAPVPDKFAKDPKLDQYMEVLKVNAAPPPAPRAERSFIELFMSFLKMHYMSMYRG
 LQLIKGAVWSMYSGESYREYLKELELQAMLGENATLVANATQGVNNTQTMNNTLLNNTKEKNNTQRVNKP
 QEKKAPQKADSP (SEQ ID NO:8)

TABLE 53' Nucleic Acid Sequence of *C. episcopatus* γ -Carboxylase

5 GGGAGTCTCAACATCTCTGTACTGCAGGGCCAGGTGGTGGTGGAGGTGCTTCCAGAGGAGGACAGTCTAG
 AACAGCCCTACAACATCAGCATCAGTGATGGCCACGAGTCATTGATTCCCACAGGGGTGTTCCACAAGGT
 GTACACAGTGTCTGAAGTGCCCTCCTGTTACATGTACATCTACATGGTCACGGAAGAGACGGAGTTCTTT
 GAGAACCTCAAAGAGCTGGAACACGCCCTCAACGGCTCCCTGGATGCTCCAGTACCAGACAAGTTTGCCA
 AAGATCCTAAACTTGATCAATATATGGAGGTACTCAAGGTGAAGAATGCAGCTCCACCACCGGCCCTCC
 10 AGCGGACAGAAGTTTCATACAGCTGTTTATGAGTTTCTGAAAATGCATTATATGTCTATGTATCGTGGA
 CTGCAGCTGATAAAAGGCGCCGTGTGGTCCATGTACTCTGGGGAATCTTACCGAGAGTACCTGAAGGAAC
 TGGAGCTACAGGCAATGCTGGGGGAGAATGTCAACCTGGTGGCAAATGCCACCCAAGGGGTGAATAAAAC
 CCAGATGATGAACAACACCTTGAACAACACCAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO:9)

15 GGGAGTCTCAACATCTCTGTACTGCAGGGCCAGGTGGTGGTGGAGGTGCTTCCAGAGGAGGACAGTCTAG
 AACAGCCCTACAACATCAGCATCAGTGATGGCCACGAGTCATTGATTCCCACAGGGGTGTTCCACAAGGT
 GTACACAGTGTCTGAAGTGCCCTCCTGTTACATGTACATCTACATGGTCACGGAAGAGACGGAGTTCTTT
 GAGAACCTCAAAGAGCTGGAACACGCCCTCAACGGCTCCCTGGATGCTCCAGTACCAGACAAGTTTGCCA
 AAGATCCTAAACTTGATCAATATATGGAGGTACTCAAAGTGAAGAATGCAGCTCCACCACCGGCCCTCC
 20 AGCGGACAGAAGTTTCATACAGCTGTTTATGAGTTTCTGAAAATGCATTATATGTCTATGTATCGTGGA
 CTGCAGCTGATAAAAGGCGCCGTGTGGTCCATGTACTCTGGGGAATCTTACCGAGAGTACCTGAAGGAAC
 TGGAGCTACAGGCAATGCTGGGGGAGAATGTCAACCTGGTGGCAAATGCCACCCAAGGGGTGAATAAAAC
 CCAGATGATGAACAACACCTTGAACAACACCAAGGAGAAAAACAACACCCAAAGGGTTAACAAGCCGCAG
 GAAAAGAAGGCCCCCAGAAGGCAGACAGCCCTAACAGCATCTCTGCAGAATGAGGGCATCATGCCTCT
 25 GTTCTGCATTTTAAATCTTCAATGTCAACACGCTGTCATGAGTCAGGATGCCAAGGGTTGATTCTAAAT
 GAAAAA (SEQ ID NO:11)

TABLE 6C-Terminal Protein Sequence of *C. episcopatus* γ -Carboxylase

30 GSLNISVLQGQVVVEVLPEEDSLEQPYNISISDGHESLIPTGVFHKVYTVSEVPSCYMYIYMVTEETEFTF
 ENLKELEHALNGSLDAPVPDKFAKDPKLDQYMEVLKVNAAPPPAPPADRSFIQLFMSFLKMHYMSMYRG
 LQLIKGAVWSMYSGESYREYLKELELQAMLGENVTLVANATQGVNKTQMMNNTLNNTKKKKKK (SEQ
 ID NO:10)

35 GSLNISVLQGQVVVEVLPEEDSLEQPYNISISDGHESLIPTGVFHKVYTVSEVPSCYMYIYMVTEETEFTF
 ENLKELEHALNGSLDAPVPDKFAKDPKLDQYMEVLKVNAAPPPAPPADRSFIQLFMSFLKMHYMSMYRG
 LQLIKGAVWSMYSGESYREYLKELELQAMLGENVTLVANATQGVNKTQMMNNTLNNTKEKNNTQRVNKPQ
 40 EKKAPQKADSP (SEQ ID NO:12)

EXAMPLE 3

Expression of γ -Carboxylase in Host Cells

[0047] The *C. textile* γ -carboxylase cDNA sequence is cloned and expressed as described
 by Walker et al. (2001). Briefly, the cDNA coding sequence is cloned in frame with green
 45 fluorescent protein (GFP) in the expression plasmid pRmHa-3.GFP (Walker et al., 2001).
 Expression in this plasmid is under control of the *Drosophila* inducible metallothionein promoter
 and carries the alcohol dehydrogenase poly (A) addition signal. *Drosophila* Schneider 2 (S2) cells

are transfected with the resultant plasmid containing the *Conus* γ -carboxylase coding sequence using CellFECTIN™ (Life Technologies). Twenty-four hours after transfection, cells are induced with 0.7mM CuSO₄. Forty-eight hours after transfection, cells are found to express GFP as seen by fluorescent microscopy. The plasmid containing the *Conus* γ -carboxylase coding sequence and the GFP sequence is modified to add a stop codon at the end of the *Conus* γ -carboxylase coding sequence and to delete the GFP coding sequence.

[0048] This modified expression vector and a vector DNA expressing the hygromycin gene are used to cotransfect *Drosophila* S2 cells. Hygromycin resistant cells are selected and individual clones are expanded. The expanded clones are analyzed for expression of *Conus* γ -carboxylase.

Briefly, the cells are induced with 0.7 mM CuSO₄ and harvested 48 hours after induction. Cells are washed twice with phosphate-buffered saline and resuspended in buffer containing 25 mM 4-morpholinepropanesulfonic acid, Ph7.0, 0.5 M NaCl, 0.2% 3-[(3-chloramidopropyl)dimethylammonio]-1-propane sulfonic acid/poshphatidyl choline, 2 MM EDTA, 2 MM dithiothreitol, 0.2 μ g/ml leupeptin, 0.8 μ g/ml pepstatin and 0.04 Mg/ml phenylmethylsulfonyl fluoride. The cell suspension is briefly sonicated and incubated in ice for 20 min. The lysate is assayed for *Conus* γ -carboxylase activity as described in Example 1. The isolated *Conus* γ -carboxylase is found to be biologically active and to properly γ -carboxylate ConG, i.e. Glu² is not γ -carboxylated while the remaining Glu residues are γ -carboxylated. The cells expressing the *Conus* γ -carboxylase are grown and maintained.

EXAMPLE 4

Synthesis of γ -Carboxylated ConG in Host Cells

[0049] The cDNA sequence coding for the ConG propeptide (U.S. Patent No. 6,172,041) is cloned and expressed as described by Walker et al. (2001). Briefly, the cDNA for the ConG propeptide sequence is cloned into pRmHa-3.GFP under control of the *Drosophila* metallothionenin promoter as described in Example 3. The resultant plasmid is modified to insert a stop codon and to delete the GFP coding sequence as described in Example 3. This expression vector is used to transfect cells expressing γ -carboxylase prepared in Example 3. Cells expressing γ -carboxylase and ConG propeptide are selected and expanded. ConG is isolated from these cells and analyzed for proper γ -carboxylation as described in Example 1. The Glu residues in ConG are found to be

properly γ -carboxylated, i.e. Glu² is not γ -carboxylated while the remaining Glu residues are γ -carboxylated.

[0050] It will be appreciated that the methods and compositions of the instant invention can be incorporated in the form of a variety of embodiments, only a few of which are disclosed herein. It will be apparent to the artisan that other embodiments exist and do not depart from the spirit of the invention. Thus, the described embodiments are illustrative and should not be construed as restrictive.

Bibliography

- 10 Ausubel, F.M., et al. (1992). *Current Protocols in Molecular Biology* and supplements, J. Wiley and Sons, New York.
- Berkner, K.L. et al. (1992). *Proc. Natl. Acad. Sci. USA* 89: 6242-6246.
- Cheung, A. et al. (1989). *Arch. Biochem. Biophys.* 274: 574-581.
- Haack, J.A. et al. (1990). *J. Biol. Chem.* 265: 6025-6029.
- 15 Knobloch, J.E. and Suttie, J.W. (1987). *J. Biol. Chem.* 262: 15334-15337.
- McIntosh, J.M. et al. (1984). *J. Biol. Chem.* 259: 14343-14346.
- Olivera, B.M. et al. (1990). *Science* 249: 257-263.
- Price, P.A. and Williamson, M.K. (1985). *J. Biol. Chem.* 260: 14971-14975.
- Price, P.A. et al. (1987). *Proc. Natl. Acad. Sci. USA* 84: 8335-8339.
- 20 Rigby, A.C. et al. (1997). *Biochemistry* 36: 6906-6914.
- Roth, D.A. et al. (1993). *Proc. Natl. Acad. Sci. USA* 90: 8372-8376.
- Sambrook, J. and Russell, D.W. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Skjaebaek, N. et al. (1997). *J. Biol. Chem.* 272: 2291-2299.
- 25 Stanley, T.B. et al. (1997). *FEBS Lett.* 407: 85-88.
- Stenflo, J. et al. (1974). *Proc. Natl. Acad. Sci. USA* 71: 2730-2733.
- Ulrich, M. et al. (1988). *J. Biol. Chem.* 263: 9697-9702.
- Walker, C.S. et al. (2001). *J. Biol. Chem.* 276: 7769-7774.
- Warder, S.E. et al. (1997). *FEBS Lett.* 411: 19-26.
- 30 White, H.S. et al. (1997). *J. Neurosci. Abst.*
- Wu, S.-M. et al. (1991a). *Proc. Natl. Acad. Sci. USA* 88: 2236-2240.
- Wu, S.-M. et al. (1991b). *Science* 254: 1634-1636.
- U.S. Patent No. 5,837,492

U.S. Patent No. 6,172,041

U.S. Patent No. 6,197,535

WHAT IS CLAIMED IS:

1. A synthetic nucleic acid encoding a protein selected from the group consisting of a γ -carboxylase comprising an amino acid sequence as set forth in SEQ ID NO:2 or 4 and a
5 protein having at least 95% identity to said γ -carboxylase and capable of γ -carboxylating Conantokin G.
2. The synthetic nucleic acid of claim 1, wherein said nucleic acid is selected from the group consisting of a nucleic acid which comprises a nucleotide sequence as set forth in SEQ ID
10 NO:1 or 3 and a nucleic acid which comprises a nucleotide sequence having at least 95% identity to the nucleotide sequence set forth in SEQ ID NO:1 or 3.
3. A vector comprising the nucleic acid of claim 1.
- 15 4. The vector of claim 3 which is an expression vector.
5. A vector comprising the nucleic acid of claim 2.
6. The vector of claim 5 which is an expression vector.
- 20 7. Host cells containing the vector of claim 3.
8. Host cells containing the vector of claim 4.
- 25 9. Host cells containing the vector of claim 5.
10. Host cells containing the vector of claim 6.
11. The host cells of claim 8 which further comprises an expression vector comprising a nucleic
30 acid sequence encoding a conantokin.

12. The host cells of claim 10 which further comprises an expression vector comprising a nucleic acid sequence encoding a conantokin.
13. A method for producing γ -carboxylase which comprises growing the host cells of claim 8 under conditions suitable for growth and isolating the expressed γ -carboxylase.
14. A method for producing γ -carboxylase which comprises growing the host cells of claim 10 under conditions suitable for growth and isolating the expressed γ -carboxylase.
15. A method for producing γ -carboxylated conantokin which comprises growing the host cells of claim 11 under conditions suitable for growth and isolating the γ -carboxylated conantokin.
16. A method for producing γ -carboxylated conantokin which comprises growing the host cells of claim 12 under conditions suitable for growth and isolating the γ -carboxylated conantokin.
17. An isolated γ -carboxylase selected from the group consisting of a γ -carboxylase comprising an amino acid sequence as set forth in SEQ ID NO:2 or 4 and a protein having at least 95% identity to said γ -carboxylase and capable of γ -carboxylating Cononatokin G.
18. A method for producing γ -carboxylated conantokin which comprises combining together a γ -carboxylase and a conantokin propeptide containing a *Conus* γ -carboxylase recognition sequence to produce a γ -carboxylated conantokin and isolating the γ -carboxylated conantokin, wherein said γ -carboxylase is selected from the group consisting of a γ -carboxylase comprising an amino acid sequence as set forth in SEQ ID NO:2 or 4 and a protein having at least 95% identity to said γ -carboxylase and capable of γ -carboxylating Conantokin G.

1

SEQUENCE LISTING

<110> Cognetix, Inc.
 University of Utah Research Foundation
 Garrett, James E.
 Bandyopadhyay, Pradip K.

<120> Conus Gamma-Carboxylase

<130> 2314-222

<150> US 60/310,496
 <151> 2001-08-08

<160> 12

<170> PatentIn version 3.0

<210> 1
 <211> 2428
 <212> DNA
 <213> Conus textile

<220>
 <221> misc_feature
 <222> (32)..(34)
 <223> start codon

<220>
 <221> misc_feature
 <222> (2426)..(2428)
 <223> last codon of coding sequence

<400> 1
 atctttgtga gcgtgatcca tcgcacaaac catgcaaagg ccaggcaaga aagtggctgc 60
 tgattcagag gaatcaaatg acatcagcca acaagcagaa aacagagacc agctcctccc 120
 ccaggaagcc agtcccaaag cgtgtgagga agaggacaca gaggatgaag aggaagaaga 180
 ggacaagttc tacaaactct ttggtttcag cttgagcgac ctcaagtcac gggacagctt 240
 tgttcgtctg ttgtcgcgcc ccgctgaccc tgccggctctg gcttatatcc gtgtcactta 300
 tgggtttttg atgatgtggg acgtgtttga ggaaaggggc ctgtcccgtg cagatatgcg 360
 atgggggtgat gatgaggcat gcagggttcc tctcttcgac ttcattgcaac ccttgcccct 420
 gcacatgatg gtccctgctgt acctgatcat gctgattgga acaggaggaa ttctattagg 480
 agccaagtac cgtgtgtgct gcgttatgca cctgctgccc tactggtaca tagtgcttct 540
 ggacgagtgc agttggaaca atcaactcta tctgtttggt ctctctctct tctccttct 600
 gctttgcgat gctaaccact actgggtccat ggacgggtctg ttcaatgcca aggttcgaaa 660
 tacggatgtt cccttggtga actacaccct cctacgtaca caggtgtttc tgggtgactt 720
 tttggctggg ctgaagaaac tggacatgga ctggatcgct gggtactcca tgggccgttt 780
 gagtgatcat tgggtctttt acccggtttac gttcctgatg acagaagacc aggtgagtgt 840
 gcttggtggtc cacctgggtg gacttgccat tgacttggtc gtgggctacc tgctcttctt 900

2

```

tgacaagaca ccaccgatcg gtgtcattat cagttcgtca ttccacctga tgaatgcaca      960
gatgttcagc ataggaatgt ttccgtatgc catgttgggt ttgacgcctg tgttcttcta    1020
tgccaaactgg ccgagggccc cgtttcgccg cattccacga tccttgagga ttcttaccct    1080
tgatgatgga gaggatgata cgctgccttc ggagaagtgc ttatacacia aagaacaggc    1140
caaaccagaa ctggccagca cccctgagca tgaaaacact gcagtccgca aacagttgac    1200
accaccact cagcccacgt tccggcatca tgctgccgct gccttcaccg ttttcttcat    1260
tctgtggcag atgtttttgc ctttctctca ttttatcaca aagggcaaca acagctggac    1320
ccagggactc tacggctact cctgggacat gatggttcac acccgagca ctcagcacac    1380
caggatctcc ttcatcaaca aggacacagg agagcgaggg ttcttgacc cgcaggcatg    1440
gagcaagtca catcgatggg cgcataacgc taagatgatg aagcagtacg ccaggtgcat    1500
cgctcgccga ctgaagaagc atgaaatcga caatgtggaa atctattttg atgtctggat    1560
atctctgaat catcgcttcc agcaacggat cgtgaacccc aatgtggaca ttttaacagc    1620
cgaatggagt gtctttaagt ccactccatg gatgatgcc ttgctggctg acttgtctaa    1680
ttggcgaagc aagttgaaag agattgagga cgacattttc aactcaaccg acctgtatga    1740
aatagtcttt ctggctgact ttcttggttt gtacctggag aactttgtcc acggcagcgt    1800
cgggagtctc aacatctctg tactgcaggg ccaggtggtg gtggaggtgc ttccagagga    1860
ggacagtcta gaagagccct acaacatcag catcagtgat ggccaagagt cattgattcc    1920
cacaggggtg ttccacaagg tgtacacagt gtctgaagtg ccctcctgtt acatgtacat    1980
ctacatggtc acggaagaga cagagttcct tgagaaactc aaagagctgg aacacgccct    2040
caacggctcc ctggatgctc cagttccaga caagtttgcc gaagatccta aacttgatca    2100
gtatatggag gtactcaaaa cgaagaatgc aactccacca ccaacctctc aagaggagca    2160
aagtttcata cagctgttta tgagttttct gaaaatgcat tatatgtcta tgtatcgtgg    2220
actgcagctg ataaaaggcg ccatgtggtc catgtactcc ggggaatctt accgagagtt    2280
cttgaagaaa ctggagctac agaaaatgct ggcggagaat gccaccctgg tggcaaacgc    2340
cacccaaggg gtgaataaca ccagacgat gaacaacacc ttgaacaaca ccaaggagaa    2400
agacaacacc caaagggtta acaaaccg

```

<210> 2

<211> 799

<212> PRT

<213> Conus textile

<400> 2

```

Met Gln Arg Pro Gly Lys Lys Val Ala Ala Asp Ser Glu Glu Ser Asn
1          5          10          15
Asp Ile Ser Gln Gln Ala Glu Asn Arg Asp Gln Leu Leu Pro Gln Glu
20          25          30

```

3

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

Val	Arg	Lys	Gln	Leu	Thr	Pro	Pro	Thr	Gln	Pro	Thr	Phe	Arg	His	His
385					390					395					400
Ala	Ala	Ala	Ala	Phe	Thr	Val	Phe	Phe	Ile	Leu	Trp	Gln	Met	Phe	Leu
				405					410					415	
Pro	Phe	Ser	His	Phe	Ile	Thr	Lys	Gly	Asn	Asn	Ser	Trp	Thr	Gln	Gly
			420					425					430		
Leu	Tyr	Gly	Tyr	Ser	Trp	Asp	Met	Met	Val	His	Thr	Arg	Ser	Thr	Gln
		435					440					445			
His	Thr	Arg	Ile	Ser	Phe	Ile	Asn	Lys	Asp	Thr	Gly	Glu	Arg	Gly	Phe
	450					455					460				
Leu	Asp	Pro	Gln	Ala	Trp	Ser	Lys	Ser	His	Arg	Trp	Ala	His	Asn	Ala
465					470					475					480
Lys	Met	Met	Lys	Gln	Tyr	Ala	Arg	Cys	Ile	Ala	Arg	Arg	Leu	Lys	Lys
				485					490					495	
His	Glu	Ile	Asp	Asn	Val	Glu	Ile	Tyr	Phe	Asp	Val	Trp	Ile	Ser	Leu
			500					505					510		
Asn	His	Arg	Phe	Gln	Gln	Arg	Ile	Val	Asn	Pro	Asn	Val	Asp	Ile	Leu
		515					520					525			
Thr	Ala	Glu	Trp	Ser	Val	Phe	Lys	Ser	Thr	Pro	Trp	Met	Met	Pro	Leu
	530					535					540				
Leu	Val	Asp	Leu	Ser	Asn	Trp	Arg	Ser	Lys	Leu	Lys	Glu	Ile	Glu	Asp
545					550					555					560
Asp	Ile	Phe	Asn	Ser	Thr	Asp	Leu	Tyr	Glu	Ile	Val	Phe	Leu	Ala	Asp
			565						570					575	
Phe	Pro	Gly	Leu	Tyr	Leu	Glu	Asn	Phe	Val	His	Gly	Ser	Val	Gly	Ser
			580					585					590		
Leu	Asn	Ile	Ser	Val	Leu	Gln	Gly	Gln	Val	Val	Val	Glu	Val	Leu	Pro
		595					600					605			
Glu	Glu	Asp	Ser	Leu	Glu	Glu	Pro	Tyr	Asn	Ile	Ser	Ile	Ser	Asp	Gly
	610					615					620				
Gln	Glu	Ser	Leu	Ile	Pro	Thr	Gly	Val	Phe	His	Lys	Val	Tyr	Thr	Val
625					630					635					640
Ser	Glu	Val	Pro	Ser	Cys	Tyr	Met	Tyr	Ile	Tyr	Met	Val	Thr	Glu	Glu
				645					650					655	
Thr	Glu	Phe	Leu	Glu	Lys	Leu	Lys	Glu	Leu	Glu	His	Ala	Leu	Asn	Gly
			660					665					670		
Ser	Leu	Asp	Ala	Pro	Val	Pro	Asp	Lys	Phe	Ala	Glu	Asp	Pro	Lys	Leu
		675					680					685			
Asp	Gln	Tyr	Met	Glu	Val	Leu	Lys	Thr	Lys	Asn	Ala	Thr	Pro	Pro	Pro
	690					695					700				
Thr	Ser	Gln	Glu	Glu	Gln	Ser	Phe	Ile	Gln	Leu	Phe	Met	Ser	Phe	Leu
705					710					715					720

5

Lys Met His Tyr Met Ser Met Tyr Arg Gly Leu Gln Leu Ile Lys Gly
725 730 735

Ala Met Trp Ser Met Tyr Ser Gly Glu Ser Tyr Arg Glu Phe Leu Lys
740 745 750

Lys Leu Glu Leu Gln Lys Met Leu Ala Glu Asn Ala Thr Leu Val Ala
755 760 765

Asn Ala Thr Gln Gly Val Asn Asn Thr Gln Thr Met Asn Asn Thr Leu
770 775 780

Asn Asn Thr Lys Glu Lys Asp Asn Thr Gln Arg Val Asn Lys Pro
785 790 795

<210> 3

<211> 2547

<212> DNA

<213> Conus textile

<220>

<221> CDS

<222> (1)..(2433)

<400> 3

atg caa agg cca ggc aag aaa gtg gct gct gat tca gag gaa tca aat 48
Met Gln Arg Pro Gly Lys Lys Val Ala Ala Asp Ser Glu Glu Ser Asn
1 5 10 15

gac atc agc caa caa gca gaa aac aga gac cag ctc ctc ccc cag gaa 96
Asp Ile Ser Gln Gln Ala Glu Asn Arg Asp Gln Leu Leu Pro Gln Glu
20 25 30

gcc agt ccc aaa gcg tgt gag gaa gag gac aca gag gat gaa gag gaa 144
Ala Ser Pro Lys Ala Cys Glu Glu Glu Asp Thr Glu Asp Glu Glu Glu
35 40 45

gaa gag gac aag ttc tac aaa ctc ttt ggt ttc agc ttg agc gac ctc 192
Glu Glu Asp Lys Phe Tyr Lys Leu Phe Gly Phe Ser Leu Ser Asp Leu
50 55 60

aag tca tgg gac agc ttt gtt cgt ctg ttg tcg cgc ccc gct gac cct 240
Lys Ser Trp Asp Ser Phe Val Arg Leu Leu Ser Arg Pro Ala Asp Pro
65 70 75 80

gcc ggt ctg gct tat atc cgt gtc act tat ggg ttt ttg atg atg tgg 288
Ala Gly Leu Ala Tyr Ile Arg Val Thr Tyr Gly Phe Leu Met Met Trp
85 90 95

gac gtg ttt gag gaa agg ggc ctg tcc cgt gca gat atg cga tgg ggt 336
Asp Val Phe Glu Glu Arg Gly Leu Ser Arg Ala Asp Met Arg Trp Gly
100 105 110

gat gat gag gca tgc agg ttt cct ctc ttc gac ttc atg caa ccc ttg 384
Asp Asp Glu Ala Cys Arg Phe Pro Leu Phe Asp Phe Met Gln Pro Leu
115 120 125

ccc ctg cac atg atg gtc ctg ctg tac ctg atc atg ctg att gga aca 432
Pro Leu His Met Met Val Leu Leu Tyr Leu Ile Met Leu Ile Gly Thr
130 135 140

gga gga att cta tta gga gcc aag tac cgt gtg tgc tgc gtt atg cac 480
Gly Gly Ile Leu Leu Gly Ala Lys Tyr Arg Val Cys Cys Val Met His
145 150 155 160

ctg	ctg	ccc	tac	tgg	tac	ata	gtg	ctt	ctg	gac	gag	tgc	agt	tgg	aac	528
Leu	Leu	Pro	Tyr	Trp	Tyr	Ile	Val	Leu	Leu	Asp	Glu	Cys	Ser	Trp	Asn	
				165					170					175		
aat	cac	tcc	tat	ctg	ttt	ggt	ctc	ctc	tct	ttc	ctc	ctt	ctg	ctt	tgc	576
Asn	His	Ser	Tyr	Leu	Phe	Gly	Leu	Leu	Ser	Phe	Leu	Leu	Leu	Leu	Cys	
			180					185					190			
gat	gct	aac	cac	tac	tgg	tcc	atg	gac	ggt	ctg	ttc	aat	gcc	aag	gtt	624
Asp	Ala	Asn	His	Tyr	Trp	Ser	Met	Asp	Gly	Leu	Phe	Asn	Ala	Lys	Val	
		195					200					205				
cga	aat	acg	gat	gtt	ccc	ttg	tgg	aac	tac	acc	ctc	cta	cgt	aca	cag	672
Arg	Asn	Thr	Asp	Val	Pro	Leu	Trp	Asn	Tyr	Thr	Leu	Leu	Arg	Thr	Gln	
	210					215					220					
gtg	ttt	ctg	gtg	tac	ttt	ttg	gct	ggg	ctg	aag	aaa	ctg	gac	atg	gac	720
Val	Phe	Leu	Val	Tyr	Phe	Leu	Ala	Gly	Leu	Lys	Lys	Leu	Asp	Met	Asp	
225					230					235					240	
tgg	atc	gct	ggt	tac	tcc	atg	ggc	cgt	ttg	agt	gat	cat	tgg	gtc	ttt	768
Trp	Ile	Ala	Gly	Tyr	Ser	Met	Gly	Arg	Leu	Ser	Asp	His	Trp	Val	Phe	
				245					250					255		
tac	ccg	ttt	acg	ttc	ctg	atg	aca	gaa	gac	cag	gtg	agt	gtg	ctt	gtg	816
Tyr	Pro	Phe	Thr	Phe	Leu	Met	Thr	Glu	Asp	Gln	Val	Ser	Val	Leu	Val	
			260					265					270			
gtc	cac	ctg	ggt	gga	ctt	gcc	att	gac	ttg	ttc	gtg	ggc	tac	ctg	ctc	864
Val	His	Leu	Gly	Gly	Leu	Ala	Ile	Asp	Leu	Phe	Val	Gly	Tyr	Leu	Leu	
		275					280					285				
ttc	ttt	gac	aag	aca	cga	ccg	atc	ggt	gtc	att	atc	agt	tcg	tca	ttc	912
Phe	Phe	Asp	Lys	Thr	Arg	Pro	Ile	Gly	Val	Ile	Ile	Ser	Ser	Ser	Phe	
	290					295					300					
cac	ctg	atg	aat	gca	cag	atg	ttc	agc	ata	gga	atg	ttt	ccg	tat	gcc	960
His	Leu	Met	Asn	Ala	Gln	Met	Phe	Ser	Ile	Gly	Met	Phe	Pro	Tyr	Ala	
305					310					315					320	
atg	ttg	ggt	ttg	acg	cct	gtg	ttc	ttc	tat	gcc	aac	tgg	ccg	agg	gcc	1008
Met	Leu	Gly	Leu	Thr	Pro	Val	Phe	Phe	Tyr	Ala	Asn	Trp	Pro	Arg	Ala	
				325					330					335		
ctg	ttt	cgc	cgc	att	cca	cga	tcc	ttg	agg	att	ctt	acc	cct	gat	gat	1056
Leu	Phe	Arg	Arg	Ile	Pro	Arg	Ser	Leu	Arg	Ile	Leu	Thr	Pro	Asp	Asp	
			340					345					350			
gga	gag	gat	gat	acg	ctg	cct	tcg	gag	aag	tgc	tta	tac	aca	aaa	gaa	1104
Gly	Glu	Asp	Asp	Thr	Leu	Pro	Ser	Glu	Lys	Cys	Leu	Tyr	Thr	Lys	Glu	
		355					360					365				
cag	gcc	aaa	cca	gaa	ctg	gcc	agc	acc	cct	gag	cat	gaa	aac	act	gca	1152
Gln	Ala	Lys	Pro	Glu	Leu	Ala	Ser	Thr	Pro	Glu	His	Glu	Asn	Thr	Ala	
	370					375					380					
gtc	cgc	aaa	cag	ttg	aca	cca	ccc	act	cag	ccc	acg	ttc	cgg	cat	cat	1200
Val	Arg	Lys	Gln	Leu	Thr	Pro	Pro	Thr	Gln	Pro	Thr	Phe	Arg	His	His	
385					390					395					400	
gct	gcc	gct	gcc	ttc	acc	gtt	ttc	ttc	att	ctg	tgg	cag	atg	ttt	ttg	1248
Ala	Ala	Ala	Ala	Phe	Thr	Val	Phe	Phe	Ile	Leu	Trp	Gln	Met	Phe	Leu	
				405					410					415		

cct ttc tct cat ttt atc aca aag ggc aac aac agc tgg acc cag gga	1296
Pro Phe Ser His Phe Ile Thr Lys Gly Asn Asn Ser Trp Thr Gln Gly	
420 425 430	
ctc tac ggc tac tcc tgg gac atg atg gtt cac acc cgc agc act cag	1344
Leu Tyr Gly Tyr Ser Trp Asp Met Met Val His Thr Arg Ser Thr Gln	
435 440 445	
cac acc agg atc tcc ttc atc aac aag gac aca gga gag cga ggg ttc	1392
His Thr Arg Ile Ser Phe Ile Asn Lys Asp Thr Gly Glu Arg Gly Phe	
450 455 460	
ctg gac ccg cag gca tgg agc aag tca cat cga tgg gcg cat aac gct	1440
Leu Asp Pro Gln Ala Trp Ser Lys Ser His Arg Trp Ala His Asn Ala	
465 470 475 480	
aag atg atg aag cag tac gcc agg tgc atc gct cgc cga ctg aag aag	1488
Lys Met Met Lys Gln Tyr Ala Arg Cys Ile Ala Arg Arg Leu Lys Lys	
485 490 495	
cat gaa atc gac aat gtg gaa atc tat ttt gat gtc tgg ata tct ctg	1536
His Glu Ile Asp Asn Val Glu Ile Tyr Phe Asp Val Trp Ile Ser Leu	
500 505 510	
aat cat cgc ttc cag caa cgg atc gtg aac ccc aat gtg gac att tta	1584
Asn His Arg Phe Gln Gln Arg Ile Val Asn Pro Asn Val Asp Ile Leu	
515 520 525	
aca gcc gaa tgg agt gtc ttt aag tcc act cca tgg atg atg ccc ttg	1632
Thr Ala Glu Trp Ser Val Phe Lys Ser Thr Pro Trp Met Met Pro Leu	
530 535 540	
ctg gtc gac ttg tct aat tgg cga agc aag ttg aaa gag att gag gac	1680
Leu Val Asp Leu Ser Asn Trp Arg Ser Lys Leu Lys Glu Ile Glu Asp	
545 550 555 560	
gac att ttc aac tca acc gac ctg tat gaa ata gtc ttt ctg gct gac	1728
Asp Ile Phe Asn Ser Thr Asp Leu Tyr Glu Ile Val Phe Leu Ala Asp	
565 570 575	
ttt cct ggt ttg tac ctg gag aac ttt gtc cac ggc agc gtc ggg agt	1776
Phe Pro Gly Leu Tyr Leu Glu Asn Phe Val His Gly Ser Val Gly Ser	
580 585 590	
ctc aac atc tct gta ctg cag ggc cag gtg gtg gtg gag gtg ctt cca	1824
Leu Asn Ile Ser Val Leu Gln Gly Gln Val Val Val Glu Val Leu Pro	
595 600 605	
gag gag gac agt cta gaa gag ccc tac aac atc agc atc agt gat ggc	1872
Glu Glu Asp Ser Leu Glu Glu Pro Tyr Asn Ile Ser Ile Ser Asp Gly	
610 615 620	
caa gag tca ttg att ccc aca ggg gtg ttc cac aag gtg tac aca gtg	1920
Gln Glu Ser Leu Ile Pro Thr Gly Val Phe His Lys Val Tyr Thr Val	
625 630 635 640	
tct gaa gtg ccc tcc tgt tac atg tac atc tac atg gtc acg gaa gag	1968
Ser Glu Val Pro Ser Cys Tyr Met Tyr Ile Tyr Met Val Thr Glu Glu	
645 650 655	
aca gag ttc ctt gag aaa ctc aaa gag ctg gaa cac gcc ctc aac ggc	2016
Thr Glu Phe Leu Glu Lys Leu Lys Glu Leu Glu His Ala Leu Asn Gly	
660 665 670	

8

tcc ctg gat gct cca gtt cca gac aag ttt gcc gaa gat cct aaa ctt 2064
 Ser Leu Asp Ala Pro Val Pro Asp Lys Phe Ala Glu Asp Pro Lys Leu
 675 680 685

gat cag tat atg gag gta ctc aaa acg aag aat gca act cca cca cca 2112
 Asp Gln Tyr Met Glu Val Leu Lys Thr Lys Asn Ala Thr Pro Pro Pro
 690 695 700

acc tct caa gag gag caa agt ttc ata cag ctg ttt atg agt ttt ctg 2160
 Thr Ser Gln Glu Glu Gln Ser Phe Ile Gln Leu Phe Met Ser Phe Leu
 705 710 715 720

aaa atg cat tat atg tct atg tat cgt gga ctg cag ctg ata aaa ggc 2208
 Lys Met His Tyr Met Ser Met Tyr Arg Gly Leu Gln Leu Ile Lys Gly
 725 730 735

gcc atg tgg tcc atg tac tcc ggg gaa tct tac cga gag ttc ttg aag 2256
 Ala Met Trp Ser Met Tyr Ser Gly Glu Ser Tyr Arg Glu Phe Leu Lys
 740 745 750

aaa ctg gag cta cag aaa atg ctg gcg gag aat gcc acc ctg gtg gca 2304
 Lys Leu Glu Leu Gln Lys Met Leu Ala Glu Asn Ala Thr Leu Val Ala
 755 760 765

aac gcc acc caa ggg gtg aat aac acc cag acg atg aac aac acc ttg 2352
 Asn Ala Thr Gln Gly Val Asn Asn Thr Gln Thr Met Asn Asn Thr Leu
 770 775 780

aac aac acc aag gag aaa gac aac acc caa agg gtt aac aaa ccg cag 2400
 Asn Asn Thr Lys Glu Lys Asp Asn Thr Gln Arg Val Asn Lys Pro Gln
 785 790 795 800

gaa aag aag gcc ccc cag aag gca gac agc ccc taacagcatc tctgcagaat 2453
 Glu Lys Lys Ala Pro Gln Lys Ala Asp Ser Pro
 805 810

gagggcgctca tgcctctgtt ctgcattgta aatcttcaat gtcagactcg ctgtcatgag 2513

tcaggatgcc aagggttgat tctaaatgaa aaaa . 2547

<210> 4
 <211> 811
 <212> PRT
 <213> Conus textile

<400> 4
 Met Gln Arg Pro Gly Lys Lys Val Ala Ala Asp Ser Glu Glu Ser Asn
 1 5 10 15
 Asp Ile Ser Gln Gln Ala Glu Asn Arg Asp Gln Leu Leu Pro Gln Glu
 20 25 30
 Ala Ser Pro Lys Ala Cys Glu Glu Glu Asp Thr Glu Asp Glu Glu Glu
 35 40 45
 Glu Glu Asp Lys Phe Tyr Lys Leu Phe Gly Phe Ser Leu Ser Asp Leu
 50 55 60
 Lys Ser Trp Asp Ser Phe Val Arg Leu Leu Ser Arg Pro Ala Asp Pro
 65 70 75 80
 Ala Gly Leu Ala Tyr Ile Arg Val Thr Tyr Gly Phe Leu Met Met Trp
 85 90 95

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

10

435		440		445
His Thr Arg Ile Ser Phe Ile Asn Lys Asp Thr Gly Glu Arg Gly Phe	450	455	460	
Leu Asp Pro Gln Ala Trp Ser Lys Ser His Arg Trp Ala His Asn Ala	465	470	475	480
Lys Met Met Lys Gln Tyr Ala Arg Cys Ile Ala Arg Arg Leu Lys Lys	485	490	495	
His Glu Ile Asp Asn Val Glu Ile Tyr Phe Asp Val Trp Ile Ser Leu	500	505	510	
Asn His Arg Phe Gln Gln Arg Ile Val Asn Pro Asn Val Asp Ile Leu	515	520	525	
Thr Ala Glu Trp Ser Val Phe Lys Ser Thr Pro Trp Met Met Pro Leu	530	535	540	
Leu Val Asp Leu Ser Asn Trp Arg Ser Lys Leu Lys Glu Ile Glu Asp	545	550	555	560
Asp Ile Phe Asn Ser Thr Asp Leu Tyr Glu Ile Val Phe Leu Ala Asp	565	570	575	
Phe Pro Gly Leu Tyr Leu Glu Asn Phe Val His Gly Ser Val Gly Ser	580	585	590	
Leu Asn Ile Ser Val Leu Gln Gly Gln Val Val Val Glu Val Leu Pro	595	600	605	
Glu Glu Asp Ser Leu Glu Glu Pro Tyr Asn Ile Ser Ile Ser Asp Gly	610	615	620	
Gln Glu Ser Leu Ile Pro Thr Gly Val Phe His Lys Val Tyr Thr Val	625	630	635	640
Ser Glu Val Pro Ser Cys Tyr Met Tyr Ile Tyr Met Val Thr Glu Glu	645	650	655	
Thr Glu Phe Leu Glu Lys Leu Lys Glu Leu Glu His Ala Leu Asn Gly	660	665	670	
Ser Leu Asp Ala Pro Val Pro Asp Lys Phe Ala Glu Asp Pro Lys Leu	675	680	685	
Asp Gln Tyr Met Glu Val Leu Lys Thr Lys Asn Ala Thr Pro Pro Pro	690	695	700	
Thr Ser Gln Glu Glu Gln Ser Phe Ile Gln Leu Phe Met Ser Phe Leu	705	710	715	720
Lys Met His Tyr Met Ser Met Tyr Arg Gly Leu Gln Leu Ile Lys Gly	725	730	735	
Ala Met Trp Ser Met Tyr Ser Gly Glu Ser Tyr Arg Glu Phe Leu Lys	740	745	750	
Lys Leu Glu Leu Gln Lys Met Leu Ala Glu Asn Ala Thr Leu Val Ala	755	760	765	
Asn Ala Thr Gln Gly Val Asn Asn Thr Gln Thr Met Asn Asn Thr Leu	770	775	780	

11

Asn Asn Thr Lys Glu Lys Asp Asn Thr Gln Arg Val Asn Lys Pro Gln
 785 790 795 800

Glu Lys Lys Ala Pro Gln Lys Ala Asp Ser Pro
 805 810

<210> 5
 <211> 612
 <212> DNA
 <213> Conus omaria

<400> 5
 gggagtctca acatctctgt actgcagggc caggtggtgg tggaggtgct tccagaggag 60
 gacagtctag aaaaacccta caacatcagc atcaatgatg gccacgagtc attgattccc 120
 acaggggtat tccacaaggt gtacacagtg tctgaagtgc cctcctgtta catgtacatc 180
 tacatgggtca cggaagagac ggagttcttt gagaacgtca aagagctgga acacgccctc 240
 aacggctccc tggatgctcc agttccagac aagtttgcca aagatcctaa acttgatcaa 300
 tatatggagg tactcaaagt gaagaatgca gctccaccac cggcccctcg agcggagaga 360
 agtttcatag agctgtttat gagttttctg aaaatgcatt atatgtctat gtatcgtgga 420
 ctgcagctga taaaaggcgc cgtgtggtcc atgtactctg gggaatctta ccgagagtac 480
 ctgaaggaac tggaattaca ggcaatgctg ggggagaatg ccaccctggt ggcaaacgcc 540
 acccaagggg tgaataacac ccagacgatg aacaacacct tattgaacaa caccaaaaaa 600
 aaaaaaaaaa aa 612

<210> 6
 <211> 204
 <212> PRT
 <213> Conus omaria

<400> 6
 Gly Ser Leu Asn Ile Ser Val Leu Gln Gly Gln Val Val Val Glu Val
 1 5 10 15
 Leu Pro Glu Glu Asp Ser Leu Glu Lys Pro Tyr Asn Ile Ser Ile Asn
 20 25 30
 Asp Gly His Glu Ser Leu Ile Pro Thr Gly Val Phe His Lys Val Tyr
 35 40 45
 Thr Val Ser Glu Val Pro Ser Cys Tyr Met Tyr Ile Tyr Met Val Thr
 50 55 60
 Glu Glu Thr Glu Phe Phe Glu Asn Val Lys Glu Leu Glu His Ala Leu
 65 70 75 80
 Asn Gly Ser Leu Asp Ala Pro Val Pro Asp Lys Phe Ala Lys Asp Pro
 85 90 95
 Lys Leu Asp Gln Tyr Met Glu Val Leu Lys Val Lys Asn Ala Ala Pro
 100 105 110
 Pro Pro Ala Pro Arg Ala Glu Arg Ser Phe Ile Glu Leu Phe Met Ser
 115 120 125

Phe Leu Lys Met His Tyr Met Ser Met Tyr Arg Gly Leu Gln Leu Ile
 130 135 140
 Lys Gly Ala Val Trp Ser Met Tyr Ser Gly Glu Ser Tyr Arg Glu Tyr
 145 150 155 160
 Leu Lys Glu Leu Glu Leu Gln Ala Met Leu Gly Glu Asn Ala Thr Leu
 165 170 175
 Val Ala Asn Ala Thr Gln Gly Val Asn Asn Thr Gln Thr Met Asn Asn
 180 185 190
 Thr Leu Leu Asn Asn Thr Lys Lys Lys Lys Lys Lys
 195 200

<210> 7
 <211> 780
 <212> DNA
 <213> Conus omaria

<220>
 <221> CDS
 <222> (1)..(666)

<400> 7
 ggg agt ctc aac atc tct gta ctg cag ggc cag gtg gtg gtg gag gtg 48
 Gly Ser Leu Asn Ile Ser Val Leu Gln Gly Gln Val Val Val Glu Val
 1 5 10 15
 ctt cca gag gag gac agt cta gaa aaa ccc tac aac atc agc atc aat 96
 Leu Pro Glu Glu Asp Ser Leu Glu Lys Pro Tyr Asn Ile Ser Ile Asn
 20 25 30
 gat ggc cac gag tca ttg att ccc aca ggg gta ttc cac aag gtg tac 144
 Asp Gly His Glu Ser Leu Ile Pro Thr Gly Val Phe His Lys Val Tyr
 35 40 45
 aca gtg tct gaa gtg ccc tcc tgt tac atg tac atc tac atg gtc acg 192
 Thr Val Ser Glu Val Pro Ser Cys Tyr Met Tyr Ile Tyr Met Val Thr
 50 55 60
 gaa gag acg gag ttc ttt gag aac gtc aaa gag ctg gaa cac gcc ctc 240
 Glu Glu Thr Glu Phe Phe Glu Asn Val Lys Glu Leu Glu His Ala Leu
 65 70 75 80
 aac ggc tcc ctg gat gct cca gtt cca gac aag ttt gcc aaa gat cct 288
 Asn Gly Ser Leu Asp Ala Pro Val Pro Asp Lys Phe Ala Lys Asp Pro
 85 90 95
 aaa ctt gat caa tat atg gag gta ctc aaa gtg aag aat gca gct cca 336
 Lys Leu Asp Gln Tyr Met Glu Val Leu Lys Val Lys Asn Ala Ala Pro
 100 105 110
 cca ccg gcc cct cga gcg gag aga agt ttc ata gag ctg ttt atg agt 384
 Pro Pro Ala Pro Arg Ala Glu Arg Ser Phe Ile Glu Leu Phe Met Ser
 115 120 125
 ttt ctg aaa atg cat tat atg tct atg tat cgt gga ctg cag ctg ata 432
 Phe Leu Lys Met His Tyr Met Ser Met Tyr Arg Gly Leu Gln Leu Ile
 130 135 140
 aaa ggc gcc gtg tgg tcc atg tac tct ggg gaa tct tac cga gag tac 480
 Lys Gly Ala Val Trp Ser Met Tyr Ser Gly Glu Ser Tyr Arg Glu Tyr

13

145	150	155	160	
ctg aag gaa ctg gaa tta cag gca atg ctg ggg gag aat gcc acc ctg				528
Leu Lys Glu Leu Glu Leu Gln Ala Met Leu Gly Glu Asn Ala Thr Leu	165	170	175	
gtg gca aac gcc acc caa ggg gtg aat aac acc cag acg atg aac aac				576
Val Ala Asn Ala Thr Gln Gly Val Asn Asn Thr Gln Thr Met Asn Asn	180	185	190	
acc tta ttg aac aac acc aag gag aaa aac aac acc caa agg gtt aac				624
Thr Leu Leu Asn Asn Thr Lys Glu Lys Asn Asn Thr Gln Arg Val Asn	195	200	205	
aag ccg cag gaa aag aag gcc ccc cag aag gca gac agc ccc				666
Lys Pro Gln Glu Lys Lys Ala Pro Gln Lys Ala Asp Ser Pro	210	215	220	
taacagcatc tctgcagaat gagggcatca tgcctctgtt ctgcatttta aatcttcgat				726
gtcagacacg ctgtcatgag tcaggatgcc aagggttgat tctaaatgaa aaaa				780
<210> 8				
<211> 222				
<212> PRT				
<213> Conus omaria				
<400> 8				
Gly Ser Leu Asn Ile Ser Val Leu Gln Gly Gln Val Val Val Glu Val				
1 5 10 15				
Leu Pro Glu Glu Asp Ser Leu Glu Lys Pro Tyr Asn Ile Ser Ile Asn	20	25	30	
Asp Gly His Glu Ser Leu Ile Pro Thr Gly Val Phe His Lys Val Tyr	35	40	45	
Thr Val Ser Glu Val Pro Ser Cys Tyr Met Tyr Ile Tyr Met Val Thr	50	55	60	
Glu Glu Thr Glu Phe Phe Glu Asn Val Lys Glu Leu Glu His Ala Leu	65	70	75	80
Asn Gly Ser Leu Asp Ala Pro Val Pro Asp Lys Phe Ala Lys Asp Pro	85	90	95	
Lys Leu Asp Gln Tyr Met Glu Val Leu Lys Val Lys Asn Ala Ala Pro	100	105	110	
Pro Pro Ala Pro Arg Ala Glu Arg Ser Phe Ile Glu Leu Phe Met Ser	115	120	125	
Phe Leu Lys Met His Tyr Met Ser Met Tyr Arg Gly Leu Gln Leu Ile	130	135	140	
Lys Gly Ala Val Trp Ser Met Tyr Ser Gly Glu Ser Tyr Arg Glu Tyr	145	150	155	160
Leu Lys Glu Leu Glu Leu Gln Ala Met Leu Gly Glu Asn Ala Thr Leu	165	170	175	
Val Ala Asn Ala Thr Gln Gly Val Asn Asn Thr Gln Thr Met Asn Asn	180	185	190	

14

Thr Leu Leu Asn Asn Thr Lys Glu Lys Asn Asn Thr Gln Arg Val Asn
 195 200 205

Lys Pro Gln Glu Lys Lys Ala Pro Gln Lys Ala Asp Ser Pro
 210 215 220

<210> 9
 <211> 609
 <212> DNA
 <213> Conus episcopatus

<400> 9
 gggaggtctca acatctctgt actgcagggc caggtggtgg tggaggtgct tccagaggag 60
 gacagtctag aacagcccta caacatcagc atcagtgatg gccacgagtc attgattccc 120
 acaggggtgt tccacaaggt gtacacagtg tctgaagtgc cctcctgtta catgtacatc 180
 tacatggtca cggaagagac ggagttcttt gagaacctca aagagctgga acacgccctc 240
 aacggctccc tggatgctcc agtaccagac aagtttgcca aagatcctaa acttgatcaa 300
 tatatggagg tactcaaggt gaagaatgca gctccaccac cggcccctcc agcggacaga 360
 agtttcatac agctgtttat gagttttctg aaaatgcatt atatgtctat gtatcgtgga 420
 ctgcagctga taaaaggcgc cgtgtggtcc atgtactctg gggaaatctta ccgagagtac 480
 ctgaaggaac tggagctaca ggcaatgctg ggggagaatg tcaccctggt ggcaaatgcc 540
 acccaagggg tgaataaaac ccagatgatg aacaacacct tgaacaacac caaaaaaaaaa 600
 aaaaaaaaaa 609

<210> 10
 <211> 203
 <212> PRT
 <213> Conus episcopatus

<400> 10
 Gly Ser Leu Asn Ile Ser Val Leu Gln Gly Gln Val Val Val Glu Val
 1 5 10 15
 Leu Pro Glu Glu Asp Ser Leu Glu Gln Pro Tyr Asn Ile Ser Ile Ser
 20 25 30
 Asp Gly His Glu Ser Leu Ile Pro Thr Gly Val Phe His Lys Val Tyr
 35 40 45
 Thr Val Ser Glu Val Pro Ser Cys Tyr Met Tyr Ile Tyr Met Val Thr
 50 55 60
 Glu Glu Thr Glu Phe Phe Glu Asn Leu Lys Glu Leu Glu His Ala Leu
 65 70 75 80
 Asn Gly Ser Leu Asp Ala Pro Val Pro Asp Lys Phe Ala Lys Asp Pro
 85 90 95
 Lys Leu Asp Gln Tyr Met Glu Val Leu Lys Val Lys Asn Ala Ala Pro
 100 105 110
 Pro Pro Ala Pro Pro Ala Asp Arg Ser Phe Ile Gln Leu Phe Met Ser
 115 120 125

15

Phe Leu Lys Met His Tyr Met Ser Met Tyr Arg Gly Leu Gln Leu Ile
 130 135 140

Lys Gly Ala Val Trp Ser Met Tyr Ser Gly Glu Ser Tyr Arg Glu Tyr
 145 150 155 160

Leu Lys Glu Leu Glu Leu Gln Ala Met Leu Gly Glu Asn Val Thr Leu
 165 170 175

Val Ala Asn Ala Thr Gln Gly Val Asn Lys Thr Gln Met Met Asn Asn
 180 185 190

Thr Leu Asn Asn Thr Lys Lys Lys Lys Lys Lys
 195 200

<210> 11
 <211> 777
 <212> DNA
 <213> Conus episcopatus

<220>
 <221> CDS
 <222> (1)..(663)

<400> 11
 ggg agt ctc aac atc tct gta ctg cag ggc cag gtg gtg gtg gag gtg 48
 Gly Ser Leu Asn Ile Ser Val Leu Gln Gly Gln Val Val Val Glu Val
 1 5 10 15

ctt cca gag gag gac agt cta gaa cag ccc tac aac atc agc atc agt 96
 Leu Pro Glu Glu Asp Ser Leu Glu Gln Pro Tyr Asn Ile Ser Ile Ser
 20 25 30

gat ggc cac gag tca ttg att ccc aca ggg gtg ttc cac aag gtg tac 144
 Asp Gly His Glu Ser Leu Ile Pro Thr Gly Val Phe His Lys Val Tyr
 35 40 45

aca gtg tct gaa gtg ccc tcc tgt tac atg tac atc tac atg gtc acg 192
 Thr Val Ser Glu Val Pro Ser Cys Tyr Met Tyr Ile Tyr Met Val Thr
 50 55 60

gaa gag acg gag ttc ttt gag aac ctc aaa gag ctg gaa cac gcc ctc 240
 Glu Glu Thr Glu Phe Phe Glu Asn Leu Lys Glu Leu Glu His Ala Leu
 65 70 75 80

aac ggc tcc ctg gat gct cca gta cca gac aag ttt gcc aaa gat cct 288
 Asn Gly Ser Leu Asp Ala Pro Val Pro Asp Lys Phe Ala Lys Asp Pro
 85 90 95

aaa ctt gat caa tat atg gag gta ctc aaa gtg aag aat gca gct cca 336
 Lys Leu Asp Gln Tyr Met Glu Val Leu Lys Val Lys Asn Ala Ala Pro
 100 105 110

cca ccg gcc cct cca gcg gac aga agt ttc ata cag ctg ttt atg agt 384
 Pro Pro Ala Pro Pro Ala Asp Arg Ser Phe Ile Gln Leu Phe Met Ser
 115 120 125

ttt ctg aaa atg cat tat atg tct atg tat cgt gga ctg cag ctg ata 432
 Phe Leu Lys Met His Tyr Met Ser Met Tyr Arg Gly Leu Gln Leu Ile
 130 135 140

aaa ggc gcc gtg tgg tcc atg tac tct ggg gaa tct tac cga gag tac 480
 Lys Gly Ala Val Trp Ser Met Tyr Ser Gly Glu Ser Tyr Arg Glu Tyr
 145 150 155 160

16

```

ctg aag gaa ctg gag cta cag gca atg ctg ggg gag aat gtc acc ctg      528
Leu Lys Glu Leu Glu Leu Gln Ala Met Leu Gly Glu Asn Val Thr Leu
                               165                               170                               175

gtg gca aat gcc acc caa ggg gtg aat aaa acc cag atg atg aac aac      576
Val Ala Asn Ala Thr Gln Gly Val Asn Lys Thr Gln Met Met Asn Asn
                               180                               185                               190

acc ttg aac aac acc aag gag aaa aac aac acc caa agg gtt aac aag      624
Thr Leu Asn Asn Thr Lys Glu Lys Asn Asn Thr Gln Arg Val Asn Lys
                               195                               200                               205

ccg cag gaa aag aag gcc ccc cag aag gca gac agc ccc taacagcatc      673
Pro Gln Glu Lys Lys Ala Pro Gln Lys Ala Asp Ser Pro
                               210                               215                               220

tctgcagaat gagggcatca tgcctctgtt ctgcatttta aatcttcaat gtcagacacg      733
ctgtcatgag tcaggatgcc aagggttgat tctaaatgaa aaaa      777

<210> 12
<211> 221
<212> PRT
<213> Conus episcopatus

<400> 12
Gly Ser Leu Asn Ile Ser Val Leu Gln Gly Gln Val Val Val Glu Val
1                               5                               10                               15

Leu Pro Glu Glu Asp Ser Leu Glu Gln Pro Tyr Asn Ile Ser Ile Ser
20                               25                               30

Asp Gly His Glu Ser Leu Ile Pro Thr Gly Val Phe His Lys Val Tyr
35                               40                               45

Thr Val Ser Glu Val Pro Ser Cys Tyr Met Tyr Ile Tyr Met Val Thr
50                               55                               60

Glu Glu Thr Glu Phe Phe Glu Asn Leu Lys Glu Leu Glu His Ala Leu
65                               70                               75                               80

Asn Gly Ser Leu Asp Ala Pro Val Pro Asp Lys Phe Ala Lys Asp Pro
85                               90                               95

Lys Leu Asp Gln Tyr Met Glu Val Leu Lys Val Lys Asn Ala Ala Pro
100                              105                              110

Pro Pro Ala Pro Pro Ala Asp Arg Ser Phe Ile Gln Leu Phe Met Ser
115                              120                              125

Phe Leu Lys Met His Tyr Met Ser Met Tyr Arg Gly Leu Gln Leu Ile
130                              135                              140

Lys Gly Ala Val Trp Ser Met Tyr Ser Gly Glu Ser Tyr Arg Glu Tyr
145                              150                              155                              160

Leu Lys Glu Leu Glu Leu Gln Ala Met Leu Gly Glu Asn Val Thr Leu
165                              170                              175

Val Ala Asn Ala Thr Gln Gly Val Asn Lys Thr Gln Met Met Asn Asn
180                              185                              190

Thr Leu Asn Asn Thr Lys Glu Lys Asn Asn Thr Gln Arg Val Asn Lys
195                              200                              205

```

Pro	Gln	Glu	Lys	Lys	Ala	Pro	Gln	Lys	Ala	Asp	Ser	Pro
210						215					220	