



Office de la Propriété
Intellectuelle
du Canada

Un organisme
d'Industrie Canada

Canadian
Intellectual Property
Office

An agency of
Industry Canada

CA 2451374 A1 2003/01/16

(21) **2 451 374**

(12) **DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION**

(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2002/07/05

(87) Date publication PCT/PCT Publication Date: 2003/01/16

(85) Entrée phase nationale/National Entry: 2003/12/19

(86) N° demande PCT/PCT Application No.: US 2002/021056

(87) N° publication PCT/PCT Publication No.: 2003/004607

(30) Priorités/Priorities: 2001/07/05 (60/303,051) US;
2002/01/16 (60/349,133) US

(51) Cl.Int.⁷/Int.Cl.⁷ C12N 9/48

(71) Demandeur/Applicant:
WYETH, US

(72) Inventeurs/Inventors:
AGOSTINO, MICHAEL J., US;
DI BLASIO, ELIZABETH, US;
LAVALLIE, EDWARD R., US;
RACIE, LISA A., US

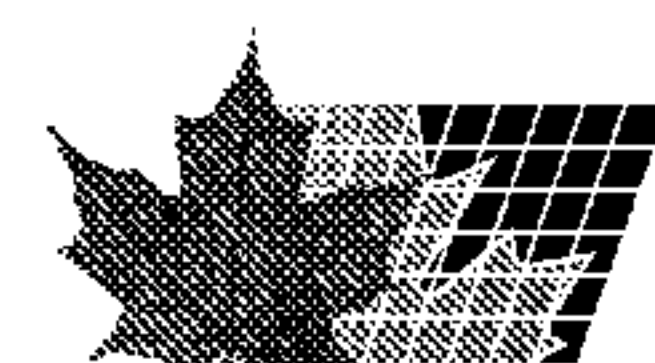
(74) Agent: TORYS LLP

(54) Titre : MOLECULES D'AGGRECANASE

(54) Title: AGGRECANASE MOLECULES

(57) Abrégé/Abstract:

Novel aggrecanase proteins and the nucleotide sequences encoding them as well as processes for producing them are disclosed. Methods for developing inhibitors of the aggrecanase enzymes and antibodies to the enzymes for treatment of conditions characterized by the degradation of aggrecan are also disclosed.



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
16 January 2003 (16.01.2003)

PCT

(10) International Publication Number
WO 03/004607 A3

(51) International Patent Classification⁷: **C12N 9/48**

(21) International Application Number: PCT/US02/21056

(22) International Filing Date: 5 July 2002 (05.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/303,051 5 July 2001 (05.07.2001) US
60/349,133 16 January 2002 (16.01.2002) US

(71) Applicant (for all designated States except US): **WYETH**
[US/US]; 5 Giralda Farms, Madison, NJ 07940 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **AGOSTINO, Michael, J.** [US/US]; 26 Wolcott Avenue, Andover, MA 01810 (US). **DI BLASIO, Elizabeth** [US/US]; 17 Chestnut Road, Tyngsboro, MA 01879 (US). **LAVALLE, Edward, R.** [US/US]; 113 Ann Lee Road, Harvard, MA 01452 (US). **RACIE, Lisa, A.** [US/US]; 124 School Street, Acton, MA 01720 (US).

(74) Agent: **GARRETT, Arthur, S.**; Finnegan, Henderson, Farabow Garrett & Dunner, L. L.P., 1300 I Street, NW, Washington, DC 20005-3315 (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, US, UZ, 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:

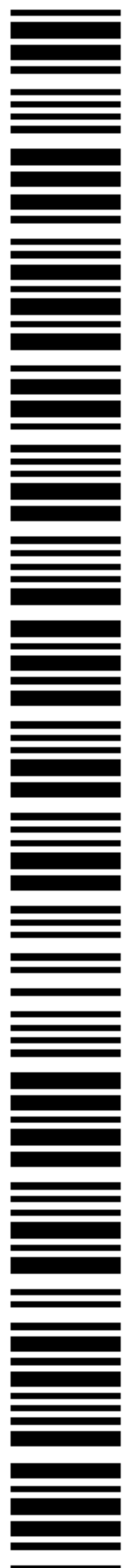
— with international search report

(88) Date of publication of the international search report:
31 July 2003

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: AGGRECANASE MOLECULES

(57) Abstract: Novel aggrecanase proteins and the nucleotide sequences encoding them as well as processes for producing them are disclosed. Methods for developing inhibitors of the aggrecanase enzymes and antibodies to the enzymes for treatment of conditions characterized by the degradation of aggrecan are also disclosed.



WO 03/004607 A3

AGGRECANASE MOLECULES

RELATED APPLICATION

[001] This application relies on the benefit of priority of U.S. provisional patent application Nos. 60/303,051, filed on July 5, 2001, and 60/349,133, filed January 16, 2002.

FIELD OF THE INVENTION

[002] The present invention relates to the discovery of nucleotide sequences encoding novel aggrecanase molecules, the aggrecanase proteins and processes for producing them. The invention further relates to the development of inhibitors of, as well as antibodies to the aggrecanase enzymes. These inhibitors and antibodies may be useful for the treatment of various aggrecanase-associated conditions including osteoarthritis.

BACKGROUND OF THE INVENTION

[003] Aggrecan is a major extracellular component of articular cartilage. It is a proteoglycan responsible for providing cartilage with its mechanical properties of compressibility and elasticity. The loss of aggrecan has been implicated in the degradation of articular cartilage in arthritic diseases. Osteoarthritis is a debilitating disease which affects at least 30 million Americans (MacLean *et al.*, *J Rheumatol* 25:2213-8 (1998)). Osteoarthritis can severely reduce quality of life due to degradation of articular cartilage and the resulting chronic pain. An early and important characteristic of the osteoarthritic process is loss of aggrecan from the extracellular matrix (Brandt and Mankin, *Pathogenesis of Osteoarthritis*, in Textbook of Rheumatology, WB Saunders Company, Philadelphia, PA, at 1355-1373

(1993)). The large, sugar-containing portion of aggrecan is thereby lost from the extra-cellular matrix, resulting in deficiencies in the biomechanical characteristics of the cartilage.

[004] A proteolytic activity termed "aggrecanase" is thought to be responsible for the cleavage of aggrecan thereby having a role in cartilage degradation associated with osteoarthritis and inflammatory joint disease. Work has been conducted to identify the enzyme responsible for the degradation of aggrecan in human osteoarthritic cartilage. Two enzymatic cleavage sites have been identified within the interglobular domain of aggrecan. One (Asn³⁴¹-Phe³⁴²) is observed to be cleaved by several known metalloproteases. Flannery *et al.*, *J Biol Chem* 267:1008-14 (1992); Fosang *et al.*, *Biochemical J.* 304:347-351 (1994). The aggrecan fragment found in human synovial fluid, and generated by IL-1 induced cartilage aggrecan cleavage is at the Glu³⁷³-Ala³⁷⁴ bond (Sandy *et al.*, *J Clin Invest* 69:1512-1516 (1992); Lohmander *et al.*, *Arthritis Rheum* 36: 1214-1222 (1993); Sandy *et al.*, *J Biol Chem* 266: 8683-8685 (1991)), indicating that none of the known enzymes are responsible for aggrecan cleavage *in vivo*.

[005] Recently, identification of two enzymes, aggrecanase-1 (ADAMTS 4) and aggrecanase-2 (ADAMTS-11) within the "Disintegrin-like and Metalloprotease with Thrombospondin type 1 motif" (ADAM-TS) family have been identified which are synthesized by IL-1 stimulated cartilage and cleave aggrecan at the appropriate site (Tortorella *et al.*, *Science* 284:1664-6 (1999); Abbaszade *et al.*, *J Biol Chem* 274: 23443-23450 (1999)). It is

possible that these enzymes could be synthesized by osteoarthritic human articular cartilage. It is also contemplated that there are other, related enzymes in the ADAM-TS family which are capable of cleaving aggrecan at the Glu³⁷³-Ala³⁷⁴ bond and could contribute to aggrecan cleavage in osteoarthritis. There is a need to identify other aggrecanase enzymes and determine ways to block their activity.

SUMMARY OF THE INVENTION

[006] The present invention is directed to the identification of novel aggrecanase protein molecules capable of cleaving aggrecan, the nucleotide sequences which encode the aggrecanase enzymes, and processes for the production of aggrecanases. These enzymes are contemplated to be characterized as having proteolytic aggrecanase activity. The invention further includes compositions comprising these enzymes.

[007] The invention also includes antibodies to these enzymes, in one embodiment, for example, antibodies that block aggrecanase activity. In addition, the invention includes methods for developing inhibitors of aggrecanase which block the enzyme's proteolytic activity. These inhibitors and antibodies may be used in various assays and therapies for treatment of conditions characterized by the degradation of articular cartilage.

[008] The invention provides an isolated DNA molecule comprising a DNA sequence chosen from: the sequence of SEQ ID NO. 5 from nucleotide #1-#2270; SEQ ID NO. 7 from nucleotide #1-#2339; SEQ ID NO. 3 from nucleotide #1 to #3899; SEQ ID NO. 9 from nucleotide #1 to #5004; SEQ. ID

NO. 11 from nucleotide #1 to #3369; and naturally occurring human allelic sequences and equivalent degenerative codon sequences.

[009] The invention also comprises a purified aggrecanase protein comprising an amino acid sequence chosen from: the amino acid sequence set forth in SEQ ID NO. 6 from amino acid #1-#756; SEQ ID NO. 8 from amino acid #1-#779; Figure 2 (SEQ ID NO. 10) from amino acid #1-#1057; Figure 5 (SEQ ID NO. 13) from amino acid #1-#1122; and homologous aggrecanase proteins consisting of addition, substitution, and deletion mutants of the sequences.

[010] The invention also provides a method for producing a purified aggrecanase protein produced by the steps of culturing a host cell transformed with a DNA molecule according to the invention, and recovering and purifying from said culture medium a protein comprising the amino acid sequence set forth in one of SEQ. ID NOs. 6, 8, 10, and 13.

[011] The invention also provides an antibody that binds to a purified aggrecanase protein of the invention. It also provides a method for developing inhibitors of aggrecanase comprising the use of aggrecanase protein chosen from SEQ ID NOs. 6, 8, 10, 13, and a fragment thereof.

[012] Additionally, it provides a pharmaceutical composition for inhibiting the proteolytic activity of aggrecanase, wherein the composition comprises at least one antibody according to the invention and at least one pharmaceutical carrier. It also provides a method for inhibiting aggrecanase in a mammal comprising administering to said mammal an effective amount of

the pharmaceutical composition and allowing the composition to inhibit aggrecanase activity.

BRIEF DESCRIPTION OF THE DRAWINGS

[013] Figure 1 is the nucleotide sequence of an aggrecanase protein as set forth in SEQ ID NO. 9.

[014] Figure 2 is the amino acid sequence (SEQ ID NO. 10) of an aggrecanase protein encoded from the nucleotide sequence as set forth in SEQ ID NO. 9.

[015] Figure 3 is an extended nucleotide sequence (SEQ ID NO. 11) of EST14.

[016] Figure 4 is an exon insert of 69 bases (SEQ ID NO. 12) from nucleotide #2138(7) through #2206(7) for SEQ ID NO. 11.

[017] Figure 5 is the predicted protein translation (SEQ ID NO. 13) of SEQ ID NO. 11.

[018] Figure 6 is an amino acid sequence (SEQ ID NO. 14) containing SEQ ID NO. 5 and 24 extra in frame amino acids as a result of an additional exon.

BRIEF DESCRIPTION OF THE SEQUENCES

a.a. = amino acid

SEQUENCES	FIGURES	DESCRIPTION
1		EST 14
2		a.a. seq. of EST 14
3		aggrecanase DNA
4		a.a. seq. of SEQ ID NO. 3
5		aggrecanase DNA
6		a.a. seq. of SEQ ID NO. 5
7		aggrecanase DNA
8		a.a. seq. of SEQ ID NO. 7
9	Figure 1	aggrecanase DNA
10	Figure 2	a.a. seq. of SEQ ID NO. 9
11	Figure 3	aggrecanase DNA
12	Figure 4	exon nucleotide insert
13	Figure 5	a.a. seq. of SEQ ID NO. 11
14	Figure 6	exon a.a. insert
15		zinc binding signature region of aggrecanase-1
16		nucleotide insert
17		nucleotide sequence containing an insert with an Xho1 site
18		a 68 bp adapter nucleotide sequence
19		exon nucleotide insert
20		exon a.a. insert
21		primer
22		primer
23		primer
24		primer
25		primer
26		primer
27		primer
28		primer
29		primer
30		primer
31		synthesized nucleotides

32		synthesized nucleotides
33		synthesized nucleotides
34		synthesized nucleotides

DETAILED DESCRIPTION OF THE INVENTION

I. Novel Aggrecanase Proteins

[019] In one embodiment, the nucleotide sequence of an aggrecanase molecule of the present invention is set forth in SEQ ID NO. 3, as nucleotides #1 to #3899. It is contemplated that nucleotides #80-134 represent the pro domain. The metalloprotease domain comprises nucleotides #135-#254; intron nucleotides #255-#317, nucleotides #318-#560, intron nucleotides #561-#1264, nucleotides #1265-#1372, intron nucleotides #1373-#1801, and nucleotides #1802-#1976. The disintegrin domain comprises nucleotides #1977-#2236. The thrombospondin type I domain comprises amino acids #2237-#2492. The spacer region comprises amino acids #2493-#2636, intron nucleotides #2637-#2759, and nucleotides #2760-#3233. The thrombospondin type I sub motif comprises nucleotides #3234-#3416. The invention further includes equivalent degenerative codon sequences of the sequence set forth in SEQ ID NO. 3, as well as fragments thereof which exhibit aggrecanase activity. The full length sequence of the aggrecanase of the present invention may be obtained using the sequences of SEQ ID NO. 3 to design probes for screening for the full sequence using standard techniques.

[020] The amino acid sequence of the isolated aggrecanase-like molecule is set forth in SEQ ID. NO. 4, as nucleotides #1 to #807. The partial

Pro domain comprises amino acids #1-#18. A probable PACE processing site comprises amino acids #15-#18. The proposed metalloprotease domain comprises amino acids #19-#209. A partial catalytic Zn binding domain comprises amino acids #145-#155. The Met turn is amino acid #168. The proposed disintegrin domain comprises amino acids #210-#298. The proposed thrombospondin type I domain comprises amino acids #299-#377. The proposed cysteine rich and cysteine poor spacer domain comprises amino acids #378-#586. The proposed thrombospondin type I sub motif comprises amino acids #587-#644. Amino acids #648-#807 are an intron

[020] sequence. The invention further includes fragments of the amino acid sequence which encode molecules exhibiting aggrecanase activity.

[021] In another embodiment, the nucleotide sequence of an aggrecanase molecule of the present invention derived from thymus DNA is set forth in SEQ ID NO. 5 from nucleotide #1-#2270. The invention includes longer aggrecanase sequences obtained using the sequences of SEQ ID NO. 5 to design probes for screening. The invention further includes equivalent degenerative codon sequences of the sequence set forth in SEQ ID NO. 5, as well as fragments thereof which exhibit aggrecanase activity.

[022] The nucleotide sequence of the thymus clones set forth in SEQ ID NO. 5 encodes the amino acid sequence set forth in SEQ ID NO. 6 from amino acid #1-#756. With respect to SEQ ID NO. 6 the domains are contemplated as follows: The pro-domain comprises amino acid #1-#88. The probable PACE site is represented by amino acids RERR, amino acids #85-

#88. The metalloprotease domain comprises amino acids #89-#317 with catalytic Zn binding domain at #264-265, and a Met turn at #278. The disintegrin domain comprises amino acids #318-#408. The thrombospondin type I domain comprises amino acids #409-#487. The cysteine rich and cysteine poor spacer domain comprises amino acids #488-#695. The proposed thrombospondin type I sub motif comprises amino acids #696-#752. The invention further includes fragments of the amino acid sequence set forth in SEQ ID NO. 6 which encode molecules exhibiting aggrecanase activity.

[023] In a further embodiment, the nucleotide sequence of an aggrecanase molecule of the present invention derived from liver DNA is set forth in SEQ ID NO. 7 from nucleotide #1-#2339. The invention includes longer aggrecanase sequences obtained using the sequences of SEQ ID NO. 7 to design probes for screening. The invention further includes equivalent degenerative codon sequences of the sequence set forth in SEQ ID NO. 7, as well as fragments thereof which exhibit aggrecanase activity. The invention further includes fragments of the amino acid sequence set forth in SEQ ID NO. 8 which encode molecules exhibiting aggrecanase activity.

[024] The nucleotide sequence set forth in SEQ ID NO. 7 encodes the amino acid sequence set forth in SEQ ID NO. 8 from amino acid #1-#779. This sequence contains a 69 base insertion encoding from amino acid #578-#601 found in the spacer domain. The domains are contemplated as follows: The pro-domain comprises amino acid #1-#88. The probable PACE site is represented by amino acids RERR , amino acids #85-#88. The

metalloprotease domain comprises amino acids #89-#317 with catalytic Zn binding domain at #264-265, and a Met turn at #278. The disintegrin domain comprises amino acids #318-#408. The thrombospondin type I domain comprises amino acids #409-#487. The cysteine rich and cysteine poor spacer domain comprises amino acids #488-#577 and #602-#718. The proposed thrombospondin type I sub motif comprises amino acids #719-#776.

[025] In a further embodiment, the nucleotide sequence of an aggrecanase molecule of the present invention is set forth in SEQ ID NO. 9 from nucleotide #1-#5004. The invention further includes equivalent degenerative codon sequences of the sequence set forth in SEQ ID NO. 9, as well as fragments thereof which exhibit aggrecanase activity.

[026] The nucleotide sequence set forth in SEQ ID NO. 9 encodes the amino acid sequence set forth in SEQ ID NO. 10 from amino acid #1-#1057. The Pro domain is contemplated to comprise amino acids #1(R) through #158(R) (probable PACE processing site is underlined in Figure 2). The proposed metalloprotease domain comprises amino acids 159 (N) through 378 (K) with catalytic Zn binding domain at #324-335, Met turn at #347. The proposed disintegrin domain comprises amino acid #379 (V) through #478 (D). The proposed thrombospondin type I domain comprises amino acid #479 (G) through #557 (L). The proposed cysteine rich and cysteine poor spacer domain comprises amino acids #558 (L) through #760 (Q). The proposed thrombospondin type I sub motifs (4) comprise amino acids #761 (D) through #990 (C). The proposed PLAC domain comprises amino acids #991(N)

through #1057 (S) (found in C terminus of papilin, lacunin, PACE4 and PC5/6 proteases as well as ADAMTS2, ADAMTS3, ADAMTS10, ADAMTS12 and EST16). The invention further includes fragments of the amino acid sequence set forth in SEQ ID NO. 10 which encode molecules exhibiting aggrecanase activity.

[027] In a further embodiment, the nucleotide sequence of an aggrecanase molecule of the present invention is set forth in SEQ ID NO. 11 from nucleotide #1-#3369. The invention further includes equivalent degenerative codon sequences of the sequence set forth in SEQ ID NO. 11, as well as fragments thereof which exhibit aggrecanase activity.

[028] The nucleotide sequence set forth in SEQ ID NO. 11 encodes the amino acid sequence set forth in SEQ ID NO. 13 from amino acid #1-#1122. The proposed leader sequence comprises amino acids #1(M) through #21 (G). The proposed Pro domain comprises amino acids #22 (L) through #223 (R) (probable PACE processing site is underlined in Figure 5). Amino acid #244 (M) is the proposed first met of N-terminal alternate splice variant. The proposed metalloprotease domain comprises amino acids #224 (N) through #443 (K) with catalytic Zn binding domain at #389-400, and a Met turn at #413. The proposed disintegrin domain comprises amino acids #444(V) through #543(D). The proposed thrombospondin type I domain comprises amino acids #544(G) through #522. The proposed cysteine rich and cysteine poor spacer domain comprises amino acids #523(L) to #830(I). The proposed thrombospondin type I sub motifs (4) comprises amino acids #831(W) to

#1055(C). The proposed PLAC domain comprises amino acids #1056 (N) through #1022(S). NxS/Tx proposed N-linked glycosylation comprise amino acids #167-169 (NNS), #812-814 (NRT) #817-819 (NQS), amino acids #859-861 (NKT), amino acids #866-868 (NDS) and amino acids #921-923 (NGT). The invention further includes fragments of the amino acid sequence set forth in SEQ ID NO. 13 which encode molecules exhibiting aggrecanase activity.

[029] The invention includes methods for obtaining the full length aggrecanase molecule, the DNA sequence obtained by this method and the protein encoded thereby. The method for isolation of the full length sequence involves utilizing the aggrecanase sequence set forth in SEQ ID NOs. 3, 5, 7, 9, and 11 to design probes for screening, or otherwise screen, using standard procedures known to those skilled in the art. The preferred sequence for designing probes is the longer sequence of SEQ ID NOs. 5 or 7.

[030] The human aggrecanase protein or a fragment thereof may be produced by culturing a cell transformed with a DNA sequence chosen from SEQ ID NOs. 3, 5, 7, 9, and 11 and recovering and purifying from the culture medium a protein characterized by an amino acid sequence set forth in at least one of SEQ ID NOs. 4, 6, 8, 10, and 13 substantially free from other proteinaceous materials with which it is co-produced. For production in mammalian cells, the DNA sequence further comprises a DNA sequence encoding a suitable propeptide 5' to and linked in frame to the nucleotide sequence encoding the aggrecanase enzyme.

[031] The human aggrecanase proteins produced by the method discussed above are characterized by having the ability to cleave aggrecan and having an amino acid sequence chosen from SEQ ID NOs. 4, 6, 8, 10, or 13 variants of the amino acid sequence of SEQ ID NOs. 4, 6, 8, 10, or 13 including naturally occurring allelic variants, and other variants in which the proteins retain the ability to cleave aggrecan characteristic of aggrecanase proteins. Preferred proteins include a protein which is at least about 80% homologous, and more preferably at least about 90% homologous, to the amino acid sequence shown in SEQ ID NOs. 4, 6, 8, 10, or 13. Finally, allelic or other variations of the sequences of SEQ ID NOs. 4, 6, 8, 10, or 13 whether such amino acid changes are induced by mutagenesis, chemical alteration, or by alteration of DNA sequence used to produce the protein, where the peptide sequence still has aggrecanase activity, are also included in the present invention. The present invention also includes fragments of the amino acid sequence of SEQ ID NOs. 4, 6, 8, 10, or 13 which retain the activity of aggrecanase protein.

II. Identification of Homologous Aggrecanase Proteins and DNA Encoding Them

[032] It is expected that additional human sequences and other species have DNA sequences homologous to human aggrecanase enzymes. The invention, therefore, includes methods for obtaining the DNA sequences encoding other aggrecanase proteins, the DNA sequences obtained by those methods, and the protein encoded by those DNA sequences. This method entails utilizing the nucleotide sequence of the invention or portions thereof to

design probes to screen libraries for the corresponding gene from other species or coding sequences or fragments thereof from using standard techniques. Thus, the present invention may include DNA sequences from other species, which are homologous to the human aggrecanase protein and can be obtained using the human sequence. The present invention may also include functional fragments of the aggrecanase protein, and DNA sequences encoding such functional fragments, as well as functional fragments of other related proteins. The ability of such a fragment to function is determinable by assay of the protein in the biological assays described for the assay of the aggrecanase protein.

[033] For example, the amino acid translation of SEQ ID NO. 20 was used in a query against the databases TREMBL, swissprot, NCBI NR, PIR, and geneseqp in a BLASTP 2.2.2 search. Several sequences were identified as similar to SEQ ID NO. 20, differing only by splicing or incomplete sequence. These sequences were identified by the following accession numbers: AAE10350, AAE10347, AAU72894, AAE10349, AAE10348. It is believed that these sequences are all part of the same family of ADAMTS. One member of this family has already been published as ADAMTS17, which appears to have as its nearest family member ADAMTS19. The cloning of ADAMTS17 has been described in Cal, S., *et al.*, *Gene*, 283 (1-2), 49-62 (2002).

[034] SEQ ID NO. 11 was used as a query against the genesqn database using BLASTN 2.2.2. SEQ ID NO. 11 was determined to have

identity (with variable splicing or incomplete sequence) to several published sequences. For example, the published sequences were cited in EP-A2-1134286 (AAD17498, AAD17499, AAD17500, AAD17501, and AAD17502) and WO 20/0183782 (AAS97177).

[035] Some examples of homologous, non-human sequences include a mouse sequence 20834206 (found in the NCBI NR database), a rat sequence 13242316 (found in the NCBI NR database), a worm sequence AAY53898 (found in the geneseqp1 database), and a cow sequence 11131272 (found in the NCBI NR database). It is expected that these sequences, from non-human species, are homologous to human aggrecanase enzymes.

[036] The aggrecanase proteins provided herein also include factors encoded by the sequences similar to those of SEQ ID NOs. 3, 5, 7, 9 or 11, but into which modifications or deletions are naturally provided (e.g. allelic variations in the nucleotide sequence which may result in amino acid changes in the protein) or deliberately engineered. For example, synthetic proteins may wholly or partially duplicate continuous sequences of the amino acid residues of SEQ ID NOs. 4, 6, 8, 10, or 13. These sequences, by virtue of sharing primary, secondary, or tertiary structural and conformational characteristics with aggrecanase proteins may possess biological properties in common therewith. It is known, for example that numerous conservative amino acid substitutions are possible without significantly modifying the structure and conformation of a protein, thus maintaining the biological

properties as well. For example, it is recognized that conservative amino acid substitutions may be made among amino acids with basic side chains, such as lysine (Lys or K), arginine (Arg or R) and histidine (His or H); amino acids with acidic side chains, such as aspartic acid (Asp or D) and glutamic acid (Glu or E); amino acids with uncharged polar side chains, such as asparagine (Asn or N), glutamine (Gln or Q), serine (Ser or S), threonine (Thr or T), and tyrosine (Tyr or Y); and amino acids with nonpolar side chains, such as alanine (Ala or A), glycine (Gly or G), valine (Val or V), leucine (Leu or L), isoleucine (Ile or I), proline (Pro or P), phenylalanine (Phe or F), methionine (Met or M), tryptophan (Trp or W) and cysteine (Cys or C). Thus, these modifications and deletions of the native aggrecanase may be employed as biologically active substitutes for naturally-occurring aggrecanase and in the development of inhibitors or other proteins in therapeutic processes. It can be readily determined whether a given variant of aggrecanase maintains the biological activity of aggrecanase by subjecting both aggrecanase and the variant of aggrecanase, as well as inhibitors thereof, to the assays described in the examples.

[037] Other specific mutations of the sequences of aggrecanase proteins described herein involve modifications of glycosylation sites. These modifications may involve O-linked or N-linked glycosylation sites. For instance, the absence of glycosylation or only partial glycosylation results from amino acid substitution or deletion at asparagine-linked glycosylation recognition sites. The asparagine-linked glycosylation recognition sites

comprise tripeptide sequences which are specifically recognized by appropriate cellular glycosylation enzymes. These tripeptide sequences are either asparagine-X-threonine or asparagine-X-serine, where X is usually any amino acid. A variety of amino acid substitutions or deletions at one or both of the first or third amino acid positions of a glycosylation recognition site (and/or amino acid deletion at the second position) results in non-glycosylation at the modified tripeptide sequence. Additionally, bacterial expression of aggrecanase-related protein will also result in production of a non-glycosylated protein, even if the glycosylation sites are left unmodified.

III. Novel Aggrecanase Nucleotide Sequences

[038] Still a further aspect of the invention are DNA sequences coding for expression of an aggrecanase protein having aggrecanase proteolytic activity or other disclosed activities of aggrecanase. Such sequences include the sequence of nucleotides in a 5' to 3' direction illustrated in SEQ ID NOs. 3, 5, 7, 9 and 11 and DNA sequences which, but for the degeneracy of the genetic code, are identical to the DNA sequence of SEQ ID NOs. 3, 5, 7, 9 and 11 and encode an aggrecanase protein.

[039] Further included in the present invention are DNA sequences which hybridize under stringent conditions with the DNA sequence of SEQ ID NOs. 1, 3, 5, 7, 9 and 11 and encode a protein having the ability to cleave aggrecan. Preferred DNA sequences include those which hybridize under stringent conditions (see Maniatis *et al*, *Molecular Cloning (A Laboratory Manual)*, Cold Spring Harbor Laboratory, at 387-389 (1982)). Such stringent conditions comprise, for example, 0.1X SSC, 0.1% SDS, at 65°C. It is

generally preferred that such DNA sequences encode a protein which is at least about 80% homologous, and more preferably at least about 90% homologous, to the sequence of set forth in SEQ ID NOs. 3, 5, 7, 9 or 11. Finally, allelic or other variations of the sequences of SEQ ID NOs. 1, 3, 5, 7, 9 or 11 whether such nucleotide changes result in changes in the peptide sequence or not, but where the peptide sequence still has aggrecanase activity, are also included in the present invention. The present invention also includes fragments of the DNA sequence shown in SEQ ID NOs 1, 3, 5, 7, 9 or 11 which encode a protein which retains the activity of aggrecanase.

[040] Similarly, DNA sequences which code for aggrecanase proteins coded for by the sequences of SEQ ID NO. 3, 5, 7, 9 or 11 or aggrecanase proteins which comprise the amino acid sequence of SEQ ID NOs. 4, 6, 8, 10, or 13 but which differ in codon sequence due to the degeneracies of the genetic code or allelic variations (naturally-occurring base changes in the species population which may or may not result in an amino acid change) also encode the novel factors described herein. Variations in the DNA sequences of SEQ ID NOs. 3, 5, 7, 9 or 11 which are caused by point mutations or by induced modifications (including insertion, deletion, and substitution) to enhance the activity, half-life or production of the proteins encoded are also encompassed in the invention.

[041] The DNA sequences of the present invention are useful, for example, as probes for the detection of mRNA encoding aggrecanase in a given cell population. Thus, the present invention includes methods of

detecting or diagnosing genetic disorders involving the aggrecanase, or disorders involving cellular, organ or tissue disorders in which aggrecanase is irregularly transcribed or expressed. Antisense DNA sequences may also be useful for preparing vectors for gene therapy applications. Antisense DNA sequences are also useful for *in vivo* methods, such as to introduce the antisense DNA into the cell, to study the interaction of the antisense DNA with the native sequences, and to test the capacity of a promoter operatively linked to the antisense DNA in a vector by studying the interaction of antisense DNA in the cell as a measure of how much antisense DNA was produced.

[042] A further aspect of the invention includes vectors comprising a DNA sequence as described above in operative association with an expression control sequence therefor. These vectors may be employed in a novel process for producing an aggrecanase protein of the invention in which a cell line transformed with a DNA sequence encoding an aggrecanase protein in operative association with an expression control sequence therefor, is cultured in a suitable culture medium and an aggrecanase protein is recovered and purified therefrom. This process may employ a number of known cells both prokaryotic and eukaryotic as host cells for expression of the protein. The vectors may be used in gene therapy applications. In such use, the vectors may be transfected into the cells of a patient *ex vivo*, and the cells may be reintroduced into a patient. Alternatively, the vectors may be introduced into a patient *in vivo* through targeted transfection.

IV. Production of Aggrecanase Proteins

[043] Another aspect of the present invention provides a method for producing novel aggrecanase proteins. The method of the present invention involves culturing a suitable cell line, which has been transformed with a DNA sequence encoding an aggrecanase protein of the invention, under the control of known regulatory sequences. The transformed host cells are cultured and the aggrecanase proteins recovered and purified from the culture medium. The purified proteins are substantially free from other proteins with which they are co-produced as well as from other contaminants. The recovered purified protein is contemplated to exhibit proteolytic aggrecanase activity cleaving aggrecan. Thus, the proteins of the invention may be further characterized by the ability to demonstrate aggrecanase proteolytic activity in an assay which determines the presence of an aggrecan-degrading molecule. These assays or the development thereof is within the knowledge of one skilled in the art. Such assays may involve contacting an aggrecan substrate with the aggrecanase molecule and monitoring the production of aggrecan fragments (see for example, Hughes *et al.*, *Biochem J* 305: 799-804 (1995); Mercuri *et al.*, *J Bio Chem* 274:32387-32395 (1999)).

[044] Suitable cells or cell lines may be mammalian cells, such as Chinese hamster ovary cells (CHO). The selection of suitable mammalian host cells and methods for transformation, culture, amplification, screening, product production and purification are known in the art. (See, e.g., Gething and Sambrook, *Nature*, 293:620-625 (1981); Kaufman *et al.*, *Mol Cell Biol*, 5(7):1750-1759 (1985); Howley *et al.*, U.S. Patent 4,419,446.) Another

suitable mammalian cell line, which is described in the accompanying examples, is the monkey COS-1 cell line. The mammalian cell CV-1 may also be suitable.

[045] Bacterial cells may also be suitable hosts. For example, the various strains of *E. coli* (e.g., HB101, MC1061) are well-known as host cells in the field of biotechnology. Various strains of *B. subtilis*, *Pseudomonas*, other bacilli and the like may also be employed in this method. For expression of the protein in bacterial cells, DNA encoding the propeptide of aggrecanase is generally not necessary.

[046] Many strains of yeast cells known to those skilled in the art may also be available as host cells for expression of the proteins of the present invention. Additionally, where desired, insect cells may be utilized as host cells in the method of the present invention. See, e.g., Miller *et al.*, *Genetic Engineering*, 8:277-298 (Plenum Press 1986).

[047] Another aspect of the present invention provides vectors for use in the method of expression of these novel aggrecanase proteins. Preferably the vectors contain the full novel DNA sequences described above which encode the novel factors of the invention. Additionally, the vectors contain appropriate expression control sequences permitting expression of the aggrecanase protein sequences. Alternatively, vectors incorporating modified sequences as described above are also embodiments of the present invention. Additionally, the sequence of SEQ ID NOs. 3, 5, 7, 9 or 11 or other sequences encoding aggrecanase proteins could be manipulated to express

composite aggrecanase proteins. Thus, the present invention includes chimeric DNA molecules encoding an aggrecanase protein comprising a fragment from SEQ ID NOs. 3, 5, 7, 9 or 11 linked in correct reading frame to a DNA sequence encoding another aggrecanase protein.

[048] The vectors may be employed in the method of transforming cell lines and contain selected regulatory sequences in operative association with the DNA coding sequences of the invention which are capable of directing the replication and expression thereof in selected host cells. Regulatory sequences for such vectors are known to those skilled in the art and may be selected depending upon the host cells. Such selection is routine and does not form part of the present invention.

V. Generation of Antibodies

[049] The purified proteins of the present inventions may be used to generate antibodies, either monoclonal or polyclonal, to aggrecanase and/or other aggrecanase-related proteins, using methods that are known in the art of antibody production. Thus, the present invention also includes antibodies to aggrecanase or other related proteins. The antibodies include both those that block aggrecanase activity and those that do not. The antibodies may be useful for detection and/or purification of aggrecanase or related proteins, or for inhibiting or preventing the effects of aggrecanase. The aggrecanase of the invention or portions thereof may be utilized to prepare antibodies that specifically bind to aggrecanase.

[050] The term "antibody" as used herein, refers to an immunoglobulin or a part thereof, and encompasses any protein comprising an antigen binding

site regardless of the source, method of production, and characteristics. The term includes but is not limited to polyclonal, monoclonal, monospecific, polyspecific, non-specific, humanized, single-chain, chimeric, synthetic, recombinant, hybrid, mutated, DCR-grafted antibodies. It also includes, unless otherwise stated, antibody fragments such as Fab, F(ab')₂, Fv, scFv, Fd, dAb, and other antibody fragments which retain the antigen binding function.

[051] Antibodies can be made, for example, via traditional hybridoma techniques (Kohler and Milstein, *Nature* 256:495-499 (1975)), recombinant DNA methods (U.S. Pat. No. 4,816,567), or phage display techniques using antibody libraries (Clackson et al., *Nature* 352: 624-628 (1991); Marks et al, *J. Mol. Biol.* 222:581-597 (1991)). For various other antibody production techniques, see *Antibodies: A Laboratory Manual*, eds. Harlow et al., Cold Spring Harbor Laboratory (1988).

[052] An antibody "specifically" binds to at least one novel aggrecanase molecule of the present invention when the antibody will not show any significant binding to molecules other than at least one novel aggrecanase molecule. The term is also applicable where, e.g., an antigen binding domain is specific for a particular epitope, which is carried by a number of antigens, in which case the specific binding member (the antibody) carrying the antigen binding domain will be able to bind to the various antigens carrying the epitope. In this fashion it is possible that an antibody of the invention will bind to multiple novel aggrecanase proteins. Typically, the

binding is considered specific when the affinity constant K_a is higher than 10^8 M^{-1} . An antibody is said to “specifically bind” or “specifically react” to an antigen if, under appropriately selected conditions, such binding is not substantially inhibited, while at the same time non-specific binding is inhibited. Such conditions are well known in the art, and a skilled artisan using routine techniques can select appropriate conditions. The conditions are usually defined in terms of concentration of antibodies, ionic strength of the solution, temperature, time allowed for binding, concentration of non-related molecules (e.g., serum albumin, milk casein), etc.

[053] Proteins are known to have certain biochemical properties including sections which are hydrophobic and sections which are hydrophilic. The hydrophobic sections would most likely be located in the interior of the structure of the protein while the hydrophilic sections would most likely be located in the exterior of the structure of the protein. It is believed that the hydrophilic regions of a protein would then correspond to antigenic regions on the protein. The hydrophobicity of SEQ ID NO. 11 was determined using GCG PepPlot. The results indicated that the n-terminus was hydrophobic presumably because of a signal sequence.

VI. Development of Inhibitors

[054] Various conditions such as osteoarthritis are known to be characterized by degradation of aggrecan. Therefore, an aggrecanase protein of the present invention which cleaves aggrecan may be useful for the development of inhibitors of aggrecanase. The invention therefore provides compositions comprising an aggrecanase inhibitor. The inhibitors may be

developed using the aggrecanase in screening assays involving a mixture of aggrecan substrate with the inhibitor followed by exposure to aggrecan. Inhibitors can be screened using high throughput processes, such as by screening a library of inhibitors. Inhibitors can also be made using three-dimensional structural analysis and/or computer aided drug design. The compositions may be used in the treatment of osteoarthritis and other conditions exhibiting degradation of aggrecan.

[055] The method may entail the determination of binding sites based on the three dimensional structure of aggrecanase and aggrecan and developing a molecule reactive with the binding site. Candidate molecules are assayed for inhibitory activity. Additional standard methods for developing inhibitors of the aggrecanase molecule are known to those skilled in the art. Assays for the inhibitors involve contacting a mixture of aggrecan and the inhibitor with an aggrecanase molecule followed by measurement of the aggrecanase inhibition, for instance by detection and measurement of aggrecan fragments produced by cleavage at an aggrecanase susceptible site. Inhibitors may be proteins or small molecules.

VII. Administration

[056] Another aspect of the invention therefore provides pharmaceutical compositions containing a therapeutically effective amount of aggrecanase antibodies and/or inhibitors, in a pharmaceutically acceptable vehicle. Aggrecanase-mediated degradation of aggrecan in cartilage has been implicated in osteoarthritis and other inflammatory diseases. Therefore, these compositions of the invention may be used in the treatment of diseases

characterized by the degradation of aggrecan and/or an up regulation of aggrecanase. The compositions may be used in the treatment of these conditions or in the prevention thereof.

[057] The invention includes methods for treating patients suffering from conditions characterized by a degradation of aggrecan or preventing such conditions. These methods, according to the invention, entail administering to a patient needing such treatment, an effective amount of a composition comprising an aggrecanase antibody or inhibitor which inhibits the proteolytic activity of aggrecanase enzymes.

[058] The antibodies and inhibitors of the present invention are useful to prevent, diagnose, or treat various medical disorders in humans or animals. In one embodiment, the antibodies can be used to inhibit or reduce one or more activities associated with the aggrecanase protein, relative to an aggrecanase protein not bound by the same antibody. Most preferably, the antibodies and inhibitors inhibit or reduce one or more of the activities of aggrecanase relative to the aggrecanase that is not bound by an antibody. In certain embodiments, the activity of aggrecanase, when bound by one or more of the presently disclosed antibodies, is inhibited at least 50%, preferably at least 60, 62, 64, 66, 68, 70, 72, 72, 76, 78, 80, 82, 84, 86, or 88%, more preferably at least 90, 91, 92, 93, or 94%, and even more preferably at least 95% to 100% relative to an aggrecanase protein that is not bound by one or more of the presently disclosed antibodies.

[059] Generally, the compositions are administered so that antibodies/their binding fragments are given at a dose between 1 μg /kg and 20 mg/kg, 1 μg /kg and 10 mg/kg, 1 μg /kg and 1 mg/kg, 10 μg /kg and 1 mg/kg, 10 μg /kg and 100 μg /kg, 100 μg and 1 mg/kg, and 500 μg /kg and 1 mg/kg. Preferably, the antibodies are given as a bolus dose, to maximize the circulating levels of antibodies for the greatest length of time after the dose. Continuous infusion may also be used after the bolus dose.

[060] In another embodiment and for administration of inhibitors, such as proteins and small molecules, an effective amount of the inhibitor is a dosage which is useful to reduce the activity of aggrecanase to achieve a desired biological outcome. Generally, appropriate therapeutic dosages for administering an inhibitor may range from 5 mg to 100 mg, from 15 mg to 85 mg, from 30 mg to 70 mg, or from 40 mg to 60 mg. Inhibitors can be administered in one dose, or at intervals such as once daily, once weekly, and once monthly. Dosage schedules can be adjusted depending on the affinity for the inhibitor to the aggrecanase target, the half-life of the inhibitor, and the severity of the patient's condition. Generally, inhibitors are administered as a bolus dose, to maximize the circulating levels of inhibitor. Continuous infusions may also be used after the bolus dose.

[061] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the

population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50} . Antibodies and inhibitors, which exhibit large therapeutic indices, are preferred.

[062] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any antibody and inhibitor used in the present invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (*i.e.*, the concentration of the test antibody which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Levels in plasma may be measured, for example, by high performance liquid chromatography. The effects of any particular dosage can be monitored by a suitable bioassay. Examples of suitable bioassays include DNA replication assays, transcription-based assays, GDF protein/receptor binding assays, creatine kinase assays, assays based on the differentiation of pre-adipocytes, assays based on glucose uptake in adipocytes, and immunological assays.

[063] The therapeutic methods of the invention include administering the aggrecanase inhibitor compositions topically, systemically, or locally as an implant or device. The dosage regimen will be determined by the attending

physician considering various factors which modify the action of the aggrecanase protein, the site of pathology, the severity of disease, the patient's age, sex, and diet, the severity of any inflammation, time of administration and other clinical factors. Generally, systemic or injectable administration will be initiated at a dose which is minimally effective, and the dose will be increased over a preselected time course until a positive effect is observed. Subsequently, incremental increases in dosage will be made limiting such incremental increases to such levels that produce a corresponding increase in effect, while taking into account any adverse affects that may appear. The addition of other known factors, to the final composition, may also affect the dosage.

[064] Progress can be monitored by periodic assessment of disease progression. The progress can be monitored, for example, by x-rays, MRI or other imaging modalities, synovial fluid analysis, patient perception, and/or clinical examination.

VIII. Assays and Methods of Detection

[065] The inhibitors and antibodies of the invention can be used in assays and methods of detection to determine the presence or absence of, or quantify aggrecanase in a sample. The inhibitors and antibodies of the present invention may be used to detect aggrecanase proteins, *in vivo* or *in vitro*. By correlating the presence or level of these proteins with a medical condition, one of skill in the art can diagnose the associated medical condition or determine its severity. The medical conditions that may be

diagnosed by the presently disclosed inhibitors and antibodies are set forth above.

[066] Such detection methods for use with antibodies are well known in the art and include ELISA, radioimmunoassay, immunoblot, western blot, immunofluorescence, immuno-precipitation, and other comparable techniques. The antibodies may further be provided in a diagnostic kit that incorporates one or more of these techniques to detect a protein (e.g., an aggrecanase protein). Such a kit may contain other components, packaging, instructions, or other material to aid the detection of the protein and use of the kit. When protein inhibitors are used in such assays, protein-protein interaction assays can be used.

[067] Where the antibodies and inhibitors are intended for diagnostic purposes, it may be desirable to modify them, for example, with a ligand group (such as biotin) or a detectable marker group (such as a fluorescent group, a radioisotope or an enzyme). If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms, electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase can be detected by its ability to convert tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. Other suitable binding partners include biotin and avidin

or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art.

EXAMPLES

EXAMPLE 1: Isolation of DNA

[068] Potential novel aggrecanase family members were identified using a database screening approach. Aggrecanase-1 (*Science* 284:1664-1666 (1999)) has at least six domains: signal, propeptide, catalytic domain, disintegrin, tsp and c-terminal. The catalytic domain contains a zinc binding signature region, TAAHELGHVKF (SEQ. ID NO. 15) and a "MET turn" which are responsible for protease activity. Substitutions within the zinc binding region in the number of the positions still allow protease activity, but the histidine (H) and glutamic acid (E) residues must be present. The thrombospondin domain of Aggrecanase-1 is also a critical domain for substrate recognition and cleavage. It is these two domains that determine our classification of a novel aggrecanase family member. The protein sequence of the Aggrecanase-1 DNA sequence was used to query against the GeneBank ESTs focusing on human ESTs using TBLASTN. The resulting sequences were the starting point in the effort to identify full length sequence for potential family members. The nucleotide sequence of the aggrecanase of the present invention is comprised of an EST that contains homology over the catalytic domain and zinc binding motif of Aggrecanase-1. EST14 (SEQ ID NO. 1), a compilation of three ESTs (GenBank accession AW575922, AW501874, AW341169) was used to predict a peptide, SEQ ID NO. 2, having similarity to a portion of the Pro and Catalytic domains of

ADAMTS4. In SEQ ID NO. 1, bases #20-#581 are most homologous to ADAMTS 7 with a 37% identity. The predicted translation of nucleotides #21-#581 encodes part of the Pro domain (bases #21-#317); PACE processing site; and partial metalloprotease domain (bases #318-#581). EST14 was located on the human genome (Celera Discovery System (Rockville, MD, USA) and Celera's associated databases) and precomputed gene predictions (FgenesH) were used to extend EST14 sequence as shown in SEQ ID NO. 3. It is contemplated to be truncated by 600-700 bases and the C terminus is expected to be truncated.

[069] The gene for EST14 was isolated using a PCR strategy with tissue sources initially determined by preliminary PCR. Using 5' primer sequence CCGGCTCCCTCGTCTCGCTCAG (SEQ ID NO. 21) and 3' primer sequence AGCAGAAGGGCTGGGGGTCAAGGAC (SEQ ID NO. 22) on nine different Marathon-Ready cDNAs from Clontech (Palo Alto, CA, USA), a 172 bp fragment corresponding to nucleotide # 52-224 of SEQ ID NO. 1 was generated using the Advantage-GC2 PCR kit from Clontech. Reaction conditions were those recommended in the user manual and included 0.5ng cDNA and 20pmole of each primer per 50 μ l reaction. Cycling conditions were as follows: 94°C for 1 min, one cycle; followed by 35 cycles consisting of 94°C for 30 sec/68°C for 3 min; followed by one cycle of 68°C for 3 min.

[070] To initiate cloning of EST14, a 2270 bp fragment (SEQ ID NO. 5) or a 2339 bp fragment (SEQ ID NO. 7) encoding the middle portion of EST14 beginning at nucleotide #52 of the EST compilation in SEQ ID NO. 3

to nucleotide # 3416 of EST14 FgenesH prediction in SEQ ID NO. 3 were generated using 5' primer sequence CCGGCTCCCTCGTCTCGCTCAG (SEQ ID NO. 21) and 3' primer sequence ACGTGACTGGCAGGGGTGCAAGTT (SEQ ID NO. 23) from human thymus (pooled from 4 male and 1 female Caucasians) (SEQ ID NO. 5) or human liver (1 male Caucasian) (SEQ ID NO. 7) Marathon-Ready (from Clontech) cDNA substrates. The MasterAmp High Fidelity Extra-Long PCR kit from Epicentre Technologies (Madison, WI, USA) was used for the PCR reactions. Premix 4 or 8 were used as described in the user manual with 0.5ng cDNA and 20pmole of each primer per 50 µl reaction. Cycling conditions were as follows: 94°C for 3 min, one cycle; followed by 35 cycles consisting of 94°C for 30 sec/68°C for 4 min and; followed by cycle of 68°C for 6 min. The PCR products resulting from these amplifications were ligated into the pT-Adv vector using the AdvanTAge PCR Cloning Kit per manufacturer's instructions (Clontech). Ligated products were transformed into ElectroMAX DH5α-E cells from Invitrogen (Carlsbad, CA, USA). Clones originating from both libraries were sequenced to determine fidelity. This fragment's location in the full-length clone (SEQ ID NO. 11) is between nucleotides # 404 and 2674. The 69 base insertion in SEQ ID NO. 7 (from liver tissue) is also present in pancreas, kidney, and liver, but not thymus, testis, or leukemia MOLT 4 cDNA.

[071] A full determination of EST14 tissue distribution was achieved by probing a Clontech Human Multiple Tissue Expression Array (MTE). A probe for the MTE was generated from a PCR product amplifying the C-

terminal end of EST14 using 5' primer sequence CGGAGCATGTGGACGGAGACTGGA (SEQ ID NO. 24) and 3' primer sequence ACGTGACTGGCAGGGGTGCAAGTT (SEQ ID NO. 23) (nucleotide #2236 to #3416 of EST14 FgenesH prediction in SEQ ID NO. 3) on human thymus Marathon-Ready cDNA. The MasterAmp High Fidelity Extra-Long PCR kit from Epicentre Technologies was used for the PCR reactions using premix 4 and standard conditions as described above.

[072] The PCR product resulting from this amplification was ligated into the pT-Adv vector using the AdvanTAge PCR Cloning Kit (from Clontech) and sequenced. A probe encoding only the spacer domain was obtained after digestion of the plasmid containing the PCR product with the restriction endonucleases Bln I and EcoR I (NEB)(nucleotide #1842 to #2410 of Figure 3) using conditions recommended by New England Biolabs (Beverly, MA, USA). The 568bp fragment was isolated using a 5% nondenaturing polyacrylamide gel using standard molecular biology techniques found in Maniatis's Molecular Cloning A Laboratory Manual. The fragment was electroeluted out of the gel slice using Sample Concentration Cups from Isco (Little Blue Tank). The purified spacer domain probe was radiolabelled using the Ready-To-Go DNA Labelling Beads (dCTP) from Amersham Pharmacia Biotech (Piscataway, NJ, USA) per the manufacturer's instructions. The radiolabelled fragment was purified away from primers and unincorporated radionucleotides using a Nick column from Amersham Pharmacia Biotech per the manufacturer's instructions and then used to probe the MTE.

Manufacturer's conditions for hybridization of the MTE using a radiolabelled cDNA probe were followed. EST14 was found to be expressed in the following tissues and cell lines: thymus, leukemia MOLT4 cell line, pancreas, kidney, fetal thymus, and liver. For cloning the remaining portions of EST14 Clontech Marathon-Ready cDNAs of the following cell lines or tissues were used: human thymus pooled from 4 male and 1 female Caucasians, human pancreas pooled from 6 male Caucasians and human leukemia, lymphoblastic MOLT-4 cell line ATCC#CRL1582.

[073] The C-terminal sequence of EST14 was determined by 3' RACE using the Clontech Marathon cDNA Amplification Kit and human thymus and leukemia, lymphoblastic MOLT-4 cell line Marathon-ready cDNAs as substrates. 3' RACE primers used were: GSP1- TCTGGCTCTCAAAGACTCGGGTAA (SEQ ID NO. 25) (nucleotide #1811 to 1834 in SEQ ID NO. 5) and GSP2- GCAGGCACAACACTGTTCGCTATGT (SEQ ID NO. 26) (nucleotide #1887 to 1909 in SEQ ID NO. 5). The Advantage-GC2 PCR Kit from Clontech was used to set up nested RACE reactions following instructions in the user manual for the Marathon cDNA Amplification Kit: the amount of GC melt used was 5 µl/50µl reaction, and the amount of GSP oligos used was 0.2pmole/µl. GSP1 primer was used for the first round of PCR and GSP2 primer was used for the nested reactions. Information from the 3' RACE is found between nucleotide #2095 and 5004 in SEQ ID NO. 9 / Figure 1 and includes an frame termination codon (TGA) at nucleotide # 3172 to 3174.

[074] A C-terminal 1079 bp fragment of EST14 including the stop codon was generated using 5' primer sequence GCAGGCACAACCTGTTTCGCTATGT (SEQ ID NO. 26) (nucleotide #2095 to 2117 of SEQ ID NO. 9) and 3' primer sequence TCACGAGCTCGGCGGTGGC (SEQ ID NO. 27) (nucleotide #3156 to 3174, complement, of SEQ ID NO. 9) on human thymus, pancreas and leukemia, lymphoblastic MOLT-4 cell line Marathon-Ready cDNAs used in the RACE reactions. The MasterAmp High Fidelity Extra-Long PCR kit from Epicentre Technologies was used for the PCR reactions using Premix 4 and standard conditions described above. The PCR products resulting from these amplifications were ligated into the pT-Adv vector using the AdvanTAge PCR Cloning Kit per manufacturer's instructions (Clontech). Ligated products were transformed into ElectroMAX DH5 α -E cells from Invitrogen. Clones originating from all three libraries were sequenced to determine fidelity. This fragment's location in the full-length clone (Figure 3) is between nucleotides # 2290 and 3369.

[075] The N-terminal sequence of EST14 was determined by 5' RACE using the Clontech Marathon cDNA Amplification Kit and human thymus and leukemia, lymphoblastic MOLT-4 cell line Marathon-ready cDNAs as substrates. 5' RACE primers used were; GSP1-TCGGCCACCACCAGGGTCTCCAC (SEQ ID NO. 28) (nucleotide # 297 to 319, complement, in SEQ ID NO. 5) and GSP2-GTTCCTCCGCTCCCGCCAGTCCC (SEQ ID NO. 29) (nucleotide #247 to

269, complement, in SEQ ID NO. 5). The Advantage-GC2 PCR Kit from Clontech was used to set up nested RACE reactions following instructions in the user manual for the Marathon cDNA Amplification Kit: the amount of GC melt used was 5 μ l/50 μ l reaction, and the amount of GSP oligos used was 0.2pmole/ μ l. GSP1 primer was used for the first round of PCR and GSP2 primer was used for the nested reactions. Information from the 5' RACE including the initiator Methionine (ATG) is found between nucleotide # 1 and 672 in Figure. 3.

[076] A N-terminal 685 bp fragment of EST14 including the initiator Methionine was generated using 5' primer sequence GGTCCCGGGTACCATGTGTGAC (SEQ ID NO. 30) (nucleotide #1 to 9 of Figure. 3) and 3' primer sequence GTTCCTCCGCTCCCGCCAGTCCC (SEQ ID NO. 29) (nucleotide # 650 to 672, complement, of Figure. 3) on human thymus Marathon-Ready cDNA used in the RACE reactions. The Advantage-GC2 PCR kit from Clontech was used for the PCR reactions. Reaction conditions were those recommended in the user manual and included 0.5ng cDNA and 20pmole of each primer per 50 μ l reaction. Cycling conditions were as follows; 94°C for 2 min, one cycle: followed by 35 cycles consisting of 94°C for 20 sec/68°C for 3 min; followed by one cycle of 68°C for 3 min.

[077] The PCR products resulting from these amplifications were ligated into the pPCR-Script AMP vector using the PCR-Script AMP Cloning Kit per manufacturer's instructions (Stratagene, La Jolla, CA, USA). Ligated

products were transformed into ElectroMAX DH5 α -E cells from Invitrogen.

Clones were sequenced to determine fidelity.

[078] Cloned PCR fragments of EST14 were sequenced to determine fidelity. The full-length sequence for EST14 was the consensus derived from the EST14 FgenesH sequence (SEQ ID NO. 3) and the PCR products generated for EST14 from the three Clontech Marathon cDNAs (SEQ ID NO. 5, 9, and Figure 3). A full-length version of EST14 was constructed by moving the PCR products of the three fragments with correct sequences from pT-Adv or pPCR-Script AMP vectors into Cos expression vector pEDasc1 as follows. Two duplexes encoding a vector XbaI site (TCTAGA) at the 5' end, optimized Kozac sequence (GCCGCCACC) upstream of the initiator Met (ATG), to the EST14 N-terminal ApaI site (GTGCAC) were synthesized in the following oligonucleotides; 5'-

CTAGAGCCGCCACCATGTGTGACGGCGCCCTGCTGCCTCCGCTCGTCC
TGCCCGTGCTGCTGCTGCTGGT (SEQ ID NO. 31) and complementary
oligo 5'-

GTCCCCAAACCAGCAGCAGCAGCACGGGCAGGACGAGCGGAGGCAGC
AGGGCGCCGTCACACATGGTGGCGGCT (SEQ ID NO. 32), 5'-
TTGGGGACTGGACCCGGGCACAGCTGTCGGCGACGCGGGCGGCCGACG
TGGAGGTGGTGCTCCCGTGGCGGGGTGCGCCCCGACGACG (SEQ ID NO.
33) complementary oligo 5'-

TGCACGTCGTCGGGGCGCACCCGCCACGGGAGCACCTCCACGTCCG
GCCGCCGCGTCGCCGACAGCTGTGCCCGGGTCCA (SEQ ID NO. 34).

These duplexes were joined with the ApaL 1 - SgrA 1 fragment of the N-terminus of EST14, SgrA 1 -Bgl II fragment of the middle portion of EST14 and a Bgl2 - Spe I fragment containing the C-terminus and stop codon (TGA) of EST14.

[079] The aggrecanase nucleotide sequence of the invention can be used to design probes for further screening for full length clones containing the isolated sequence. For example, EST14 may be used to locate smaller ESTs isolated from a variety of cDNA libraries. Examples of such ESTs, including the genbank accession number and their library origins are as follows: AA884550-- Soares_testis_NHT; AI808729-- Soares_NFL_T_GBC_S1 (pooled from fetal lung NbHL19W, testis NHT, and B-cell NCI_CGAP_GCB1); AI871510-- NCI_CGAP_Brn25 (anaplastic oligodendroglioma from brain); AI937739-- NCI_CGAP_Brn25 (anaplastic oligodendroglioma from brain); AW293573-- NCI_CGAP_Sub4 (colon); AW341169-- NCI_CGAP_Lu24 (carcinoid lung); AW501874-- NIH_MGC_52 (lymph germinal center B cells); AW575922-- NIH_MGC_52 (lymph germinal center B cells); BF529318--NCI_CGAP_Brn67(anaplastic oligodendroglioma with 1p/19q loss); BI828046-- NIH_MGC_119 (medulla brain); and BQ053458-- NIH_MGC_106 (natural killer cells, cell line).

[080] The final nucleotide sequence of EST14 from the Met to stop codon is set forth in SEQ ID NO. 11. In alternate splice variants exon 2 is missing 371 nucleotides from nucleotide #79 to #449 set forth in SEQ ID NO. 11 (counting the exon with the initiator Met as exon 1) which throws the frame

off at the N-terminus so the initiator Met is not in frame with the remainder of the protein. M is the first met found in sequence of this alternate splice variant. As seen above, the leader sequence and pro domain are missing from this truncated form. An additional exon can be found in certain cDNAs (liver, pancreas, kidney) that encodes for 24 extra in frame amino acids set forth in SEQ ID NO. 14 from amino acid #113(V) to #136(C) following the cysteine rich spacer domain in liver but not thymus cDNA including 4 extra cysteines. These extra cysteines are not found in any of the ADAMTS family members.

[081] The expression profile from Human Multiple Tissue Expression Array and Multiple Tissue Northern from Clontech is as follows: moderate expression is found in lymphoblastic leukemia molt4 cell line and thymus. Lower expression is found in pancreas, kidney, and fetal thymus. Weak but detectable expression is found in liver, salivary gland, fetal brain, lymph node, colorectal adenocarcinoma SW480 cell line, fetal lung, trachea, fetal spleen, and testis.

EXAMPLE 2: Expression of Aggrecanase

[082] In order to produce murine, human or other mammalian aggrecanase-related proteins, the DNA encoding it is transferred into an appropriate expression vector and introduced into mammalian cells or other preferred eukaryotic or prokaryotic hosts including insect host cell culture systems by conventional genetic engineering techniques. Expression systems for biologically active recombinant human aggrecanase are

contemplated to be stably transformed mammalian cells, insect, yeast or bacterial cells.

[083] One skilled in the art can construct mammalian expression vectors by employing a sequence comprising SEQ ID NOs. 3, 5, 7, 9, 11 or other DNA sequences encoding aggrecanase-related proteins or other modified sequences and known vectors, such as pCD (Okayama *et al.*, *Mol Cell Biol*, 2:161-170 (1982)), pJL3, pJL4 (Gough *et al.*, *EMBO J*, 4:645-653 (1985)) and pMT2 CXM.

[084] The mammalian expression vector pMT2 CXM is a derivative of p91023(b) (Wong *et al.*, *Science* 228:810-815 (1985)) differing from the latter in that it contains the ampicillin resistance gene in place of the tetracycline resistance gene and further contains a XhoI site for insertion of cDNA clones. The functional elements of pMT2 CXM have been described (Kaufman, *Proc. Natl. Acad. Sci. USA* 82:689-693 (1985)) and include the adenovirus VA genes, the SV40 origin of replication including the 72 bp enhancer, the adenovirus major late promoter including a 5' splice site and the majority of the adenovirus tripartite leader sequence present on adenovirus late mRNAs, a 3' splice acceptor site, a DHFR insert, the SV40 early polyadenylation site (SV40), and pBR322 sequences needed for propagation in *E. coli*.

[085] Plasmid pMT2 CXM is obtained by EcoRI digestion of pMT2-VWF, which has been deposited with the American Type Culture Collection (ATCC), Rockville, MD (USA) under accession number ATCC 67122. EcoRI digestion excises the cDNA insert present in pMT2-VWF, yielding pMT2 in

linear form which can be ligated and used to transform *E. coli* HB 101 or DH-5 to ampicillin resistance. Plasmid pMT2 DNA can be prepared by conventional methods. pMT2 CXM is then constructed using loopout/in mutagenesis (Morinaga, *et al.*, *Biotechnology* 84: 636 (1984)). This removes bases 1075 to 1145 relative to the Hind III site near the SV40 origin of replication and enhancer sequences of pMT2. In addition it inserts the following sequence: 5' PO-CATGGGCAGCTCGAG-3' (SEQ. ID NO. 16) at nucleotide 1145. This sequence contains the recognition site for the restriction endonuclease Xho I. A derivative of pMT2CXM, termed pMT23, contains recognition sites for the restriction endonucleases PstI, Eco RI, Sall and XhoI. Plasmid pMT2 CXM and pMT23 DNA may be prepared by conventional methods.

[086] pEMC2 β 1 derived from pMT21 may also be suitable in practice of the invention. pMT21 is derived from pMT2 which is derived from pMT2-VWF. As described above EcoRI digestion excises the cDNA insert present in pMT-VWF, yielding pMT2 in linear form which can be ligated and used to transform *E. Coli* HB 101 or DH-5 to ampicillin resistance. Plasmid pMT2 DNA can be prepared by conventional methods.

[087] pMT21 is derived from pMT2 through the following two modifications. First, 76 bp of the 5' untranslated region of the DHFR cDNA including a stretch of 19 G residues from G/C tailing for cDNA cloning is deleted. In this process, a XhoI site is inserted to obtain the following sequence immediately upstream from DHFR:

5' -CTGCAGGCGAGCCTGAATTCCTCGAGCCATCATG-3' (SEQ. ID NO. 17)

PstI

Eco RI XhoI

[088] Second, a unique ClaI site is introduced by digestion with EcoRV and XbaI, treatment with Klenow fragment of DNA polymerase I, and ligation to a ClaI linker (CATCGATG). This deletes a 250 bp segment from the adenovirus associated RNA (VAI) region but does not interfere with VAI RNA gene expression or function. pMT21 is digested with EcoRI and XhoI, and used to derive the vector pEMC2B1.

[089] A portion of the EMCV leader is obtained from pMT2-ECAT1 (S.K. Jung, et al, *J. Virol* 63:1651-1660 (1989)) by digestion with Eco RI and PstI, resulting in a 2752 bp fragment. This fragment is digested with TaqI yielding an Eco RI-TaqI fragment of 508 bp which is purified by electrophoresis on low melting agarose gel. A 68 bp adapter and its complementary strand are synthesized with a 5' TaqI protruding end and a 3' XhoI protruding end which has the following sequence:

5'-

CGAGGTTAAAAACGTCTAGGCCCCCGAACCACGGGGACGTGGTTTTTCCTTT

TaqI

GAAAAACACGATTGC-3' (SEQ. ID NO. 18)

XhoI

[090] This sequence matches the EMC virus leader sequence from nucleotide 763 to 827. It also changes the ATG at position 10 within the EMC virus leader to an ATT and is followed by a XhoI site. A three way ligation of the pMT21 Eco RI-XhoI fragment, the EMC virus EcoRI-TaqI fragment, and

the 68 bp oligonucleotide adapter TaqI-XhoI adapter resulting in the vector pEMC2 β 1.

[091] This vector contains the SV40 origin of replication and enhancer, the adenovirus major late promoter, a cDNA copy of the majority of the adenovirus tripartite leader sequence, a small hybrid intervening sequence, an SV40 polyadenylation signal and the adenovirus VA I gene, DHFR and β -lactamase markers and an EMC sequence, in appropriate relationships to direct the high level expression of the desired cDNA in mammalian cells.

[092] The construction of vectors may involve modification of the aggrecanase-related DNA sequences. For instance, aggrecanase cDNA can be modified by removing the non-coding nucleotides on the 5' and 3' ends of the coding region. The deleted non-coding nucleotides may or may not be replaced by other sequences known to be beneficial for expression. These vectors are transformed into appropriate host cells for expression of aggrecanase-related proteins. Additionally, the sequence of SEQ ID NOs. 3, 5, 7, 9, 11 or other sequences encoding aggrecanase-related proteins can be manipulated to express a mature aggrecanase-related protein by deleting aggrecanase encoding propeptide sequences and replacing them with sequences encoding the complete propeptides of other aggrecanase proteins.

[093] One skilled in the art can manipulate the sequences of SEQ ID NOs. 3, 5, 7, 9, or 11 by eliminating or replacing the mammalian regulatory sequences flanking the coding sequence with bacterial sequences to create

bacterial vectors for intracellular or extracellular expression by bacterial cells. For example, the coding sequences could be further manipulated (e.g., ligated to other known linkers or modified by deleting non-coding sequences therefrom or altering nucleotides therein by other known techniques). The modified aggrecanase-related coding sequence could then be inserted into a known bacterial vector using procedures such as described in *Taniguchi et al., Proc Natl Acad Sci USA*, 77:5230-5233 (1980). This exemplary bacterial vector could then be transformed into bacterial host cells and an aggrecanase-related protein expressed thereby. For a strategy for producing extracellular expression of aggrecanase-related proteins in bacterial cells, see, e.g., European patent application EPA 177,343.

[094] Similar manipulations can be performed for the construction of an insect vector (see, e.g. procedures described in published European patent application EPA 155,476) for expression in insect cells. A yeast vector could also be constructed employing yeast regulatory sequences for intracellular or extracellular expression of the factors of the present invention by yeast cells. (See, e.g., procedures described in published PCT application WO86/00639 and European patent application EPA 123,289).

[095] A method for producing high levels of a aggrecanase-related protein of the invention in mammalian, bacterial, yeast or insect host cell systems may involve the construction of cells containing multiple copies of the heterologous aggrecanase-related gene. The heterologous gene is linked to an amplifiable marker, e.g., the dihydrofolate reductase (DHFR) gene for

which cells containing increased gene copies can be selected for propagation in increasing concentrations of methotrexate (MTX) according to the procedures of Kaufman and Sharp, *J Mol Biol*, 159:601-629 (1982). This approach can be employed with a number of different cell types.

[096] For example, a plasmid containing a DNA sequence for an aggrecanase-related protein of the invention in operative association with other plasmid sequences enabling expression thereof and the DHFR expression plasmid pAdA26SV(A)3 (Kaufman and Sharp, *Mol Cell Biol* 2:1304 (1982)) can be co-introduced into DHFR-deficient CHO cells, DUKX-BII, by various methods including calcium phosphate coprecipitation and transfection, electroporation or protoplast fusion. DHFR expressing transformants are selected for growth in alpha media with dialyzed fetal calf serum, and subsequently selected for amplification by growth in increasing concentrations of MTX (e.g. sequential steps in 0.02, 0.2, 1.0 and 5uM MTX) as described in Kaufman *et al.*, *Mol Cell Biol.*, 5:1750 (1983). Transformants are cloned, and biologically active aggrecanase expression is monitored by the assays described above. Aggrecanase protein expression should increase with increasing levels of MTX resistance. Aggrecanase proteins are characterized using standard techniques known in the art such as pulse labeling with ³⁵S methionine or cysteine and polyacrylamide gel electrophoresis. Similar procedures can be followed to produce other related aggrecanase-related proteins.

[097] In one example the aggrecanase gene of the present invention set forth in SEQ ID NO. 11 may be cloned into the expression vector pED6 (Kaufman *et al.*, *Nucleic Acid Res* 19:44885-4490 (1991)). COS and CHO DUKX B11 cells are transiently transfected with the aggrecanase sequence of the invention (+/- co-transfection of PACE on a separate pED6 plasmid) by lipofection (LF2000, Invitrogen). Duplicate transfections are performed for each gene of interest: (a) one for harvesting conditioned media for activity assay and (b) one for 35-S-methionine/cysteine metabolic labeling.

[098] On day one media is changed to DME(COS) or alpha(CHO) media + 1% heat-inactivated fetal calf serum +/- 100µg/ml heparin on wells(a) to be harvested for activity assay. After 48h (day 4), conditioned media is harvested for activity assay.

[099] On day 3, the duplicate wells (b) are changed to MEM (methionine-free/cysteine free) media + 1% heat-inactivated fetal calf serum +100µg/ml heparin + 100µCi/ml 35S-methioine/cysteine (Redivue Pro mix, Amersham). Following 6h incubation at 37°C, conditioned media is harvested and run on SDS-PAGE gels under reducing conditions. Proteins are visualized by autoradiography.

EXAMPLE 3: Biological Activity of Expressed Aggrecanase

[0100] To measure the biological activity of the expressed aggrecanase-related proteins obtained in Example 2 above, the proteins are recovered from the cell culture and purified by isolating the aggrecanase-related proteins from other proteinaceous materials with which they are co-produced as well as from other contaminants. Purification is carried out using

standard techniques known to those skilled in the art. The purified protein may be assayed in accordance with the following assays:

[0101] Assays specifically to determine if the protein is an enzyme capable of cleaving aggrecan at the aggrecanase cleavage site:

[0102] 1. Fluorescent peptide assay: Expressed protein is incubated with a synthetic peptide which encompasses amino acids at the aggrecanase cleavage site of aggrecan. One side of the synthetic peptide has a fluorophore and the other a quencher. Cleavage of the peptide separates the fluorophore and quencher and elicits fluorescence. From this assay it can be determined that the expressed protein can cleave aggrecan at the aggrecanase site, and relative fluorescence tells the relative activity of the expressed protein.

[0103] 2. Neoepitope western: Expressed protein is incubated with intact aggrecan. After several biochemical manipulations of the resulting sample (dialysis, chondroitinase treatment, lyophilization and reconstitution) the sample is run on an SDS PAGE gel. The gel is incubated with an antibody that only recognizes a site on aggrecan exposed after aggrecanase cleavage. The gel is transferred to nitrocellulose and developed with a secondary antibody (called a western assay) to result in bands running at a molecular weight consistent with aggrecanase generated cleavage products of aggrecan. This assay tells the expressed protein cleaved native aggrecan at the aggrecanase cleavage site, and also tells the molecular weight of the

cleavage products. Relative density of the bands can give some idea of relative aggrecanase activity.

[0104] Assay to determine if an expressed protein can cleave aggrecan anywhere in the protein (not specific to the aggrecanase site):

[0105] 3. Aggrecan ELISA: Expressed protein is incubated with intact aggrecan which had been previously adhered to plastic wells. The wells are washed and then incubated with an antibody that detects aggrecan. The wells are developed with a secondary antibody. If there is the original amount of aggrecan remaining in the well, the antibody will densely stain the well. If aggrecan was digested off the plate by the expressed protein, the antibody will demonstrate reduced staining due to reduced aggrecan concentration. This assay tells whether an expressed protein is capable of cleaving aggrecan (anywhere in the protein, not only at the aggrecanase site) and can determine relative aggrecan cleaving.

[0106] Protein analysis of the purified proteins is conducted using standard techniques such as SDS-PAGE acrylamide (Laemmli, *Nature* 227:680 (1970)) stained with silver (Oakley, *et al.*, *Anal Biochem.* 105:361 (1980)) and by immunoblot (Towbin, *et al.*, *Proc. Natl. Acad. Sci. USA* 76:4350 (1979)). Using the above described assays, expressed aggrecanase-related proteins are evaluated for their activity and useful aggrecanase-related molecules are identified.

EXAMPLE 4: Preparation of Antibodies

[0107] An antibody against a novel aggrecanase molecule is prepared. To develop an antibody capable of inhibiting aggrecanase activity, a group of

mice are immunized every two weeks with a novel aggrecanase protein mixed in Freund's complete adjuvant for the first two immunizations, and incomplete Freund's adjuvant thereafter. Throughout the immunization period, blood is sampled and tested for the presence of circulating antibodies. At week 9, an animal with circulating antibodies is selected, immunized for three consecutive days, and sacrificed. The spleen is removed and homogenized into cells. The spleen cells are fused to a myeloma fusion partner (line P3-x63-Ag8.653) using 50% PEG 1500 by an established procedure (Oi & Herzenberg, *Selected Methods in Cellular Immunology*, W. J. Freeman Co., San Francisco, CA, at 351 (1980)). The fused cells are plated into 96-well microtiter plates at a density of 2×10^5 cells/well. After 24 hours, the cells are subjected to HAT selection (Littlefield, *Science*, 145: 709 (1964)) effectively killing any unfused and unproductively fused myeloma cells.

[0108] Successfully fused hybridoma cells secreting anti-aggrecanase antibodies are identified by solid and solution phase ELISAs. Novel aggrecanase protein is prepared from CHO cells as described above and coated on polystyrene (for solid phase assays) or biotinylated (for a solution based assay). Neutralizing assays are also employed where aggrecan is coated on a polystyrene plate and biotin aggrecanase activity is inhibited by the addition of hybridoma supernatant. Results identify hybridomas expressing aggrecanase antibodies. These positive clones are cultured and expanded for further study. These cultures remain stable when expanded and cell lines are cloned by limiting dilution and cryopreserved.

[0109] From these cell cultures, a panel of antibodies is developed that specifically recognize aggrecanase proteins. Isotype of the antibodies is determined using a mouse immunoglobulin isotyping kit (Zymed™ Laboratories, Inc., San Francisco, CA).

EXAMPLE 5: Method of Detecting Level of Aggrecanase

[0110] The anti-aggrecanase antibody prepared according to Example 4 can be used to detect the level of aggrecanase in a sample. The antibody can be used in an ELISA, for example, to identify the presence or absence, or quantify the amount of, aggrecanase in a sample. The antibody is labeled with a fluorescent tag. In general, the level of aggrecanase in a sample can be determined using any of the assays disclosed in Example 3.

EXAMPLE 6: Method of Treating a Patient

[0111] The antibody developed according to Example 4 can be administered to patients suffering from a disease or disorder related to the loss of aggrecan, or excess aggrecanase activity. Patients take the composition one time or at intervals, such as once daily, and the symptoms and signs of their disease or disorder improve. For example, loss of aggrecan would decrease or cease and degradation of articular cartilage would decrease or cease. Symptoms of osteoarthritis would be reduced or eliminated. This shows that the composition of the invention is useful for the treatment of diseases or disorders related to the loss of aggrecan, or excess aggrecanase activity. The antibodies can also be used with patients susceptible to osteoarthritis, such as those who have a family history or markers of the disease, but have not yet begun to suffer its effects.

Patient's Condition	Route of Administration	Dosage	Frequency	Predicted Results
Osteoarthritis	Subcutaneous	500 µg/kg	Daily	Decrease in symptoms
"	"	1 mg/kg	Weekly	"
"	Intramuscular	500 µg/kg	Daily	"
"	"	1 mg/kg	Weekly	"
"	Intravenous	500 µg/kg	Daily	"
"	"	1 mg/kg	Weekly	"
Family History of Osteoarthritis	Subcutaneous	500 µg/kg	Daily	Prevention of condition
"	Intramuscular	500 µg/kg	Daily	"
"	Intravenous	500 µg/kg	Daily	"

[0112] The foregoing descriptions detail presently preferred embodiments of the present invention. Numerous modifications and variations in practice thereof are expected to occur to those skilled in the art upon consideration of these descriptions. Those modifications and variations are believed to be encompassed within the claims appended hereto. All of the documents cited in this application are incorporated by reference in their entirety. Additionally, all sequences cited in databases and all references disclosed are incorporated by reference in their entirety.

WO 03/004607

PCT/US02/21056

<110> WYETH

<120> AGGRECANASE MOLECULES

<130> 08702.0092-00304

<140>

<141>

<150> 60/349,133

<151> 2002-01-16

<150> 60/303,051

<151> 2001-06-05

<160> 34

<170> PatentIn Ver. 2.1

<210> 1

<211> 601

<212> DNA

<213> Homo sapiens

<400> 1

```

gcggccgccc cgccgagctg tgcttctact cgggccgtgt gctcggccac cccggctccc 60
tcgtctcgct cagcgctgc ggcgccgcg gcggcctggg tggcctcatt cagcttgggc 120
aggagcaggt gctaattccag cccctcaaca actcccaggg cccattcagt ggacgagAAC 180
atctgatcag gcgcaaattg tccttgaccc ccagcccttc tgctgaggcc cagagacctg 240
agcagctctg caagggttcta acagaaaaga agaagccgac gtggggcagg ccttcgcggg 300
actggcggga gcggaggAAC gctatccggc tcaccagcga gcacacggg gagaccctgg 360
tggtggccga cgccgacatg gtgcagtacc acggggccga ggccgcccag aggttcattc 420
tgaccgtcat gaacatggta tacaatatgt ttcagcacca gagcctgggg attaaaatta 480
acattcaagt gaccaagctt gtcctgctac gacaacgtcc cgctaagttg tccattgggc 540
accatgggtga gcggtccctg gagagcttct gtcactggca gaacgaggag tatgcctcgt 600
g
601

```

<210> 2

<211> 187

<212> PRT

<213> Homo sapiens

<400> 2

```

Cys Phe Tyr Ser Gly Arg Val Leu Gly His Pro Gly Ser Leu Val Ser
 1           5           10           15
Leu Ser Ala Cys Gly Ala Ala Gly Gly Leu Val Gly Leu Ile Gln Leu
          20           25           30
Gly Gln Glu Gln Val Leu Ile Gln Pro Leu Asn Asn Ser Gln Gly Pro
          35           40           45
Phe Ser Gly Arg Glu His Leu Ile Arg Arg Lys Trp Ser Leu Thr Pro
          50           55           60
Ser Pro Ser Ala Glu Ala Gln Arg Pro Glu Gln Leu Cys Lys Val Leu
          65           70           75           80

```

<400> 3						
gcttgacaga	aggcctgttc	actgcatggt	tttggaagtc	agtaagccaa	ggaccgcaca	60
aatgtttcca	tcatttttcta	gaaaagaaga	agccgacgtg	gggcaggcct	tcgcgggact	120
ggcgggagcg	gaggaacgct	atccggctca	ccagcgagca	cacggtggag	accctggtgg	180
tggccgacgc	cgacatggtg	cagtaccacg	gggccgaggc	cgcccagagg	ttcatcctga	240
ccgtcatgaa	catggaatca	gagccccgaa	gggaatccag	ggaacaggac	tgctctgggg	300
ctgcgagggc	gggcagagta	tacaatatgt	ttcagcacca	gagcctgggg	attaaaatta	360
acattcaagt	gaccaagctt	gtcctgctac	gacaacgtcc	cgctaagttg	tccattgggc	420
accatggtga	gcggtccctg	gagagcttct	gtcactggca	gaacgaggag	tatggaggag	480
cgcgatacct	cggcaataac	caggttccccg	gcgggaagga	cgacccgccc	ctggtggatg	540
ctgccgtgtt	tgtgaccagg	ctgtggtcaa	gccggacagt	gtattctcca	agacgttccc	600
tgacaaacag	gtggctaggt	ggctgccatg	gaggacatgc	ataccctctg	ggcctctctc	660
tggcagtggc	tgaagacagc	agccgcttct	ctccaagcct	ggctggcatg	gccaagtcac	720
tcctgctatt	caggaaacag	gcttggtggg	gtcacataac	ttgtccactg	acacaggaag	780
acttcagttc	tggtgacttg	gtgtcctgca	cttaccgcca	gagccctttg	tggctgcccc	840
gcgtgagacc	ttcgttgcca	tcaaaaggag	gggaagggaa	tagccgattg	ggcatctacg	900
ttccaccagc	gtgtttcatg	ttaagaagaa	ggagtgactg	ctccagccca	gggaccctcg	960
agtcatctgt	ggggactcgt	gtgattctct	taccagagac	agcatctcct	cctgaagtcc	1020
aggatcctgg	agacacctca	ggcaagttca	tgaagggagc	ccttggaag	gagcaatgtg	1080
cagctcgaca	gagggacagc	catgggggag	agcaggtgca	gctcgacaga	gggacagccc	1140
atggagagca	ggtgcagttc	gacagaggga	cagcccatgg	gggagagcag	tggctggcct	1200
gccgtcccac	caacaccacc	catccttggg	atgcggcctc	cactgccttg	cattgcgttt	1260
ctcctggaat	tgcttactta	ggaggtgtgt	gcagtgtctaa	gaggaagtgt	gtgcttgccg	1320
aagacaatgg	tctcaattttg	gccttttacca	tcgcccataga	gctggggccac	aaatcctgcc	1380
tctcctatat	catcattaac	tcccgtgtaa	ccactgagct	gaagctgtgg	attcattcga	1440
ttaacagctt	tctgattctg	tgccctgaca	aaggagcagg	ctgcagaaga	cttcccagcc	1500
ctgctgcgga	cacgagctgg	gggatggcaa	gtcctggtgc	agagcttctg	gcagcctcaa	1560
ctgagtggtt	cttgagagctg	gaagggatgt	ccagagtcga	cattttgtag	acgatatac	1620
cagcagcgac	tgaagaggag	cctcaacgat	accataaaaa	caaagcagat	tgggataaca	1680
ttgcagggcc	tctgaaaact	aaactgtcat	tggaattaaa	gccacaaaaa	ataattcgtt	1740
caataagtat	ttttacccaaa	tgctcgcttt	gcaccagttt	ctgctgccct	gagaaaatag	1800
gcttgggcat	gaaccacgac	gatgaccact	catcttgcgc	tggcaggtcc	cacatcatgt	1860

WO 03/004607

PCT/US02/21056

```

caggagagtg ggtgaaaggc cggaacccaa gtgacctctc ttggtctctc ttgagcagag 1920
atgaccttga aaacttcctc aagtcaaaag tcagcacctg cttgctagtc acggacccca 1980
gaagccagca cacagtacgc ctcccgcaca agctgccggg catgcactac agtgccaacg 2040
agcagtgccca gatcctgttt ggcatgaatg ccaccttctg cagaaacatg gagcatctaa 2100
tgtgtgctgg actgtggtgc ctggtagaag gagacacatc ctgcaagacc aagctggacc 2160
ctcccctgga tggcaccgag tgtggggcag acaagtgggt ccgcgcgggg gagtgcgtag 2220
gcaagacgcc catcccggag catgtggacg gagactggag cccgtggggc gcctggagca 2280
tgtgcagccg aacatgtggg acgggagccc gcttcgggca gaggaaatgt gacaaccccc 2340
cccctggggc tggaggcaca cactgcccgg gtgccagtgt agaacatgcg gtctgcgaga 2400
acctgccctg cccaagggt ctgcccagct tccgggacca gcagtgccag gcacacgacc 2460
ggctgagccc caagaagaaa ggcttgctga cagccgtggt ggttgacgat aagccatgtg 2520
aactctactg ctgcccctc gggaaggagt cccactgct ggtggccgac agggctcctg 2580
acggtacacc ctgcggggccc tacgagactg atctctgcgt gcacggcaag tgccaggtga 2640
cgtacttctc cttcggtcct tggggagccc accaagagct agtgacaatg gcagctcctg 2700
atgtctggag caggcagatc agtgtcagga tcaccatgcg ttgccctcac agaactgtga 2760
aaatcggctg tgacggcatc atcgggtctg cagccaaaga ggacagatgc ggggtctgca 2820
gcggggacgg caagacctgc cacttggtga agggcgactt cagccacgcc cgggggacag 2880
gttatatcga agctgcctgc attcctgctg gagctcggag gatccgtgtg gtggaggata 2940
aacctgcccc cagctttctg gctctcaaag actcgggtaa ggggtccatc aacagtgact 3000
ggaagataga gctccccgga gagttccaga ttgcaggcac aactgttcgc tatgtgagaa 3060
gggggctgtg ggagaagatc tctgccaagg gaccaaccaa actaccgctg cacttgatgg 3120
tgttggtatt tcacgaccaa gattatggaa ttcattatga atacactgtt cctgtaaacc 3180
gcactgcgga aaatcaaagc gaaccagaaa aaccgcagga ctctttgttc atctggacc 3240
acagcggctg ggaagggtgc agtgtgcagt gcggcgagg ggagcgcaga accatcgtct 3300
cgtgtacacg gattgtcaac aagaccacaa ctctggtgaa cgacagtgc tgccctcaag 3360
caagccgccc agagccccag gtccgaaggt gcaacttgca cccctgccag tcacgtgccg 3420
gcttctccca gcgcctctgt cctaagacag agaatttgcc cagtgtggtc cgttgccctt 3480
cggcaggccc tttcacagtg caccttcccc ttgctgcctc tctgcaccct ccttgccctt 3540
cccctggagg ggctttcctg caagtcatgc acccaccatg gctgccattc ccaaagactc 3600
tgacaaagaa gccctactgc ttctccctgg gccagccatc atctttgcag cctcatagaa 3660
aagccatccc gagcatcaca ttggagacac cctcccatag gctgggtggg tttggaactg 3720
agagtcaagg attttctttc cccatgttct ctgtgcttct cacttgcaag ggagcctgga 3780
cgggaccccc tatgtctctg agcagtatgt tgtacactca taacatgcag agaataacag 3840
tattctctgc atgttatattc agcaataact tggttcttgc aggatttgac attgcttaa 3899

```

<210> 4

<211> 807

<212> PRT

<213> Homo sapiens

<400> 4

```

Glu Lys Lys Lys Pro Thr Trp Gly Arg Pro Ser Arg Asp Trp Arg Glu
 1                    5                10              15

Arg Arg Asn Ala Ile Arg Leu Thr Ser Glu His Thr Val Glu Thr Leu
                20                25              30

Val Val Ala Asp Ala Asp Met Val Gln Tyr His Gly Ala Glu Ala Ala
          35                40              45

Gln Arg Phe Ile Leu Thr Val Met Asn Met Val Tyr Asn Met Phe Gln
 50                    55                60

His Gln Ser Leu Gly Ile Lys Ile Asn Ile Gln Val Thr Lys Leu Val
 65                    70                75              80

Leu Leu Arg Gln Arg Pro Ala Lys Leu Ser Ile Gly His His Gly Glu
          85                90              95

```

WO 03/004607

PCT/US02/21056

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

Glu	Thr	Asp	Leu	Cys	Val	His	Gly	Lys	Cys	Gln	Val	Lys	Leu	Gly	Cys
			420					425					430		
Asp	Gly	Ile	Ile	Gly	Ser	Ala	Ala	Lys	Glu	Asp	Arg	Cys	Gly	Val	Cys
		435					440					445			
Ser	Gly	Asp	Gly	Lys	Thr	Cys	His	Leu	Val	Lys	Gly	Asp	Phe	Ser	His
	450					455					460				
Ala	Arg	Gly	Thr	Gly	Tyr	Ile	Glu	Ala	Ala	Val	Ile	Pro	Ala	Gly	Ala
465					470					475					480
Arg	Arg	Ile	Arg	Val	Val	Glu	Asp	Lys	Pro	Ala	His	Ser	Phe	Leu	Ala
				485					490					495	
Leu	Lys	Asp	Ser	Gly	Lys	Gly	Ser	Ile	Asn	Ser	Asp	Trp	Lys	Ile	Glu
			500					505					510		
Leu	Pro	Gly	Glu	Phe	Gln	Ile	Ala	Gly	Thr	Thr	Val	Arg	Tyr	Val	Arg
		515					520					525			
Arg	Gly	Leu	Trp	Glu	Lys	Ile	Ser	Ala	Lys	Gly	Pro	Thr	Lys	Leu	Pro
	530					535					540				
Leu	His	Leu	Met	Val	Leu	Leu	Phe	His	Asp	Gln	Asp	Tyr	Gly	Ile	His
545					550					555					560
Tyr	Glu	Tyr	Thr	Val	Pro	Val	Asn	Arg	Thr	Ala	Glu	Asn	Gln	Ser	Glu
				565					570					575	
Pro	Glu	Lys	Pro	Gln	Asp	Ser	Leu	Phe	Ile	Trp	Thr	His	Ser	Gly	Trp
			580					585					590		
Glu	Gly	Cys	Ser	Val	Gln	Cys	Gly	Gly	Gly	Glu	Arg	Arg	Thr	Ile	Val
		595					600					605			
Ser	Cys	Thr	Arg	Ile	Val	Asn	Lys	Thr	Thr	Thr	Leu	Val	Asn	Asp	Ser
	610					615					620				
Asp	Cys	Pro	Gln	Ala	Ser	Arg	Pro	Glu	Pro	Gln	Val	Arg	Arg	Cys	Asn
625					630					635					640
Leu	His	Pro	Cys	Gln	Ser	Arg	Ala	Gly	Phe	Ser	Gln	Arg	Leu	Cys	Pro
				645					650					655	
Lys	Thr	Glu	Asn	Leu	Pro	Ser	Val	Val	Arg	Cys	Pro	Ser	Ala	Gly	Pro
			660					665					670		
Phe	Thr	Val	His	Leu	Pro	Leu	Ala	Ala	Ser	Leu	His	Pro	Pro	Cys	Leu
		675					680					685			
Ser	Pro	Gly	Gly	Ala	Phe	Leu	Gln	Val	Met	His	Pro	Pro	Trp	Leu	Pro
	690					695					700				
Phe	Pro	Lys	Thr	Leu	Thr	Lys	Lys	Pro	Tyr	Cys	Phe	Ser	Leu	Gly	Gln
705					710					715					720
Pro	Ser	Ser	Leu	Gln	Pro	His	Arg	Lys	Ala	Ile	Pro	Ser	Ile	Thr	Leu
				725					730					735	

WO 03/004607

PCT/US02/21056

Glu Thr Pro Ser His Arg Leu Val Gly Phe Gly Thr Glu Ser Gln Gly
 740 745 750
 Phe Ser Phe Pro Met Phe Ser Val Leu Leu Thr Cys Lys Gly Ala Trp
 755 760 765
 Thr Gly Pro Pro Met Ser Leu Ser Ser Ser Leu Tyr Thr His Asn Met
 770 775 780
 Gln Arg Ile Thr Val Phe Ser Ala Cys Tyr Phe Ser Asn Asn Leu Val
 785 790 795 800
 Leu Ala Gly Phe Asp Ile Ala
 805

<210> 5

<211> 2270

<212> DNA

<213> Homo sapiens

<400> 5

```

ccggctccct cgtctcgctc agcgccctgcg gcgccgccgg cggcctgggt ggccctcattc 60
agcttgggca ggagcaggtg ctaatccagc ccctcaacaa ctcccagggc ccattcagtg 120
gacgagaaca tctgatcagg cgcaaattggt ccttgacccc cagcccttct gctgaggccc 180
agagacctga gcagctctgc aaggttctaa cagaaaagaa gaagccgacg tggggcaggc 240
cttcgcggga ctggcgggag cggaggaacg ctatccgggt caccagcgag cacacgggtg 300
agaccctggt ggtggccgac gccgacatgg tgcagtacca cggggccgag gccgcccaga 360
ggttcatcct gaccgtcatg aacatggtat acaatatgtt tcagcaccag agcctgggga 420
ttaaatttaa cattcaagtg accaagcttg tcctgctacg acaacgtccc gctaagttgt 480
ccattgggca ccatggtgag cggctccctg agagcttctg tcactggcag aacgaggagt 540
atggaggagc gcgatacctc ggcaataacc aggttcccgg cgggaaggac gaccgcgcc 600
tggtggatgc tgctgtgttt gtgaccagga cagatttctg tgtacacaaa gatgaaccgt 660
gtgacactgt tggaattgct tacttaggag gtgtgtgcag tgctaagagg aagtgtgtgc 720
ttgccgaaga caatggtctc aatttggcct ttaccatcgc ccatgagctg ggccacaact 780
tgggcatgaa ccacgacgat gaccactcat cttgcgctgg caggtcccac atcatgtcag 840
gagagtgggt gaaaggccgg aaccaagtg acctctcttg gtcctcctgc agccgagatg 900
accttgaaaa ctccctcaag tcaaaagtca gcacctgctt gctagtacag gacccagaa 960
gccagcacac agtacgcctc ccgcacaagc tgccgggcat gcactacagt gccaacgagc 1020
agtgccagat cctgtttggc atgaatgcca ccttctgcag aaacatggag catctaattg 1080
gtgctggact gtggtgcctg gtagaaggag acacatcctg caagaccaag ctggaccctc 1140
ccctggatgg caccgagtgt ggggcagaca agtggtgccg cgcgggggag tgcgtgagca 1200
agacgcccac cccggagcat gtggacggag actggagccc gtggggcgcc tggagcatgt 1260
gcagccgaac atgtgggacg ggagcccgtt tccggcagag gaaatgtgac aaaaaaaaa 1320
ctgggcctgg aggcacacac tgcccgggtg ccagtgtaga acatgcggtc tgcgagaacc 1380
tgccctgccc caagggctctg cccagcttcc gggaccagca gtgccaggca cacgaccggc 1440
tgagcccca gaagaaaggc ctgctgacag ccgtggtggt tgacgataag ccatgtgaac 1500
tctactgctc gcccctcggg aaggagtccc cactgctggt ggccgacagg gtcctggacg 1560
gtacaccctg cgggccctac gagactgatc tctgcgtgca cggcaagtgc cagaaaatcg 1620
gctgtgacgg catcatcggg tctgcagcca aagaggacag atgcggggtc tgcagcgggg 1680
acggcaagac ctgccacttg gtgaaggggc acttcagcca cggccggggg acaggttata 1740
tcgaagctgc cgtcattcct gctggagctc ggaggatccg tgtggtggag gataaacctg 1800
cccacagctt tctggctctc aaagactcgg gtaaggggtc catcaacagt gactggaaga 1860
tagagctccc cggagagttc cagattgcag gcacaactgt tcgctatgtg agaagggggc 1920
tgtgggagaa gatctctgcc aagggaccaa ccaaactacc gctgcacttg atggtgttgt 1980
tatttcacga ccaagattat ggaattcatt atgaatacac tgttcctgta aaccgcactg 2040
cggaaaatca aagcgaacca gaaaaaccgc aggactcttt gttcatctgg acccacagcg 2100
gctgggaagg gtgcagtgtg cagtgcggcg gaggggagcg cagaaccatc gtctcgtgta 2160
cacggattgt caacaagacc acaactctgg tgaacgacag tgactgccct caagcaagcc 2220
gccagagacc ccagggtccga aggtgcaact tgcacccctg ccagtcacgt 2270

```

<210> 6

<211> 756

<212> PRT

<213> Homo sapiens

<400> 6

Gly	Ser	Leu	Val	Ser	Leu	Ser	Ala	Cys	Gly	Ala	Ala	Gly	Gly	Leu	Val
1				5					10					15	

Gly	Leu	Ile	Gln	Leu	Gly	Gln	Glu	Gln	Val	Leu	Ile	Gln	Pro	Leu	Asn
			20					25					30		

Asn	Ser	Gln	Gly	Pro	Phe	Ser	Gly	Arg	Glu	His	Leu	Ile	Arg	Arg	Lys
		35					40					45			

Trp	Ser	Leu	Thr	Pro	Ser	Pro	Ser	Ala	Glu	Ala	Gln	Arg	Pro	Glu	Gln
	50					55					60				

Leu	Cys	Lys	Val	Leu	Thr	Glu	Lys	Lys	Lys	Pro	Thr	Trp	Gly	Arg	Pro
65					70					75					80

Ser	Arg	Asp	Trp	Arg	Glu	Arg	Arg	Asn	Ala	Ile	Arg	Leu	Thr	Ser	Glu
				85					90					95	

His	Thr	Val	Glu	Thr	Leu	Val	Val	Ala	Asp	Ala	Asp	Met	Val	Gln	Tyr
			100					105					110		

His	Gly	Ala	Glu	Ala	Ala	Gln	Arg	Phe	Ile	Leu	Thr	Val	Met	Asn	Met
		115					120					125			

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

Gln	Val	Thr	Lys	Leu	Val	Leu	Leu	Arg	Gln	Arg	Pro	Ala	Lys	Leu	Ser
145					150					155					160

Ile	Gly	His	His	Gly	Glu	Arg	Ser	Leu	Glu	Ser	Phe	Cys	His	Trp	Gln
				165					170					175	

Asn	Glu	Glu	Tyr	Gly	Gly	Ala	Arg	Tyr	Leu	Gly	Asn	Asn	Gln	Val	Pro
			180					185					190		

Gly	Gly	Lys	Asp	Asp	Pro	Pro	Leu	Val	Asp	Ala	Ala	Val	Phe	Val	Thr
		195					200					205			

Arg	Thr	Asp	Phe	Cys	Val	His	Lys	Asp	Glu	Pro	Cys	Asp	Thr	Val	Gly
	210					215					220				

Ile	Ala	Tyr	Leu	Gly	Gly	Val	Cys	Ser	Ala	Lys	Arg	Lys	Cys	Val	Leu
225					230					235					240

Ala	Glu	Asp	Asn	Gly	Leu	Asn	Leu	Ala	Phe	Thr	Ile	Ala	His	Glu	Leu
				245					250					255	

Gly	His	Asn	Leu	Gly	Met	Asn	His	Asp	Asp	Asp	His	Ser	Ser	Cys	Ala
			260					265					270		

Gly	Arg	Ser	His	Ile	Met	Ser	Gly	Glu	Trp	Val	Lys	Gly	Arg	Asn	Pro
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

WO 03/004607

PCT/US02/21056

275					280					285					
Ser	Asp	Leu	Ser	Trp	Ser	Ser	Cys	Ser	Arg	Asp	Asp	Leu	Glu	Asn	Phe
	290					295					300				
Leu	Lys	Ser	Lys	Val	Ser	Thr	Cys	Leu	Leu	Val	Thr	Asp	Pro	Arg	Ser
305					310					315					320
Gln	His	Thr	Val	Arg	Leu	Pro	His	Lys	Leu	Pro	Gly	Met	His	Tyr	Ser
				325					330					335	
Ala	Asn	Glu	Gln	Cys	Gln	Ile	Leu	Phe	Gly	Met	Asn	Ala	Thr	Phe	Cys
			340					345					350		
Arg	Asn	Met	Glu	His	Leu	Met	Cys	Ala	Gly	Leu	Trp	Cys	Leu	Val	Glu
		355					360					365			
Gly	Asp	Thr	Ser	Cys	Lys	Thr	Lys	Leu	Asp	Pro	Pro	Leu	Asp	Gly	Thr
	370					375					380				
Glu	Cys	Gly	Ala	Asp	Lys	Trp	Cys	Arg	Ala	Gly	Glu	Cys	Val	Ser	Lys
385						390					395				400
Thr	Pro	Ile	Pro	Glu	His	Val	Asp	Gly	Asp	Trp	Ser	Pro	Trp	Gly	Ala
				405					410					415	
Trp	Ser	Met	Cys	Ser	Arg	Thr	Cys	Gly	Thr	Gly	Ala	Arg	Phe	Arg	Gln
			420					425					430		
Arg	Lys	Cys	Asp	Asn	Pro	Pro	Pro	Gly	Pro	Gly	Gly	Thr	His	Cys	Pro
		435					440					445			
Gly	Ala	Ser	Val	Glu	His	Ala	Val	Cys	Glu	Asn	Leu	Pro	Cys	Pro	Lys
	450					455					460				
Gly	Leu	Pro	Ser	Phe	Arg	Asp	Gln	Gln	Cys	Gln	Ala	His	Asp	Arg	Leu
465						470					475				480
Ser	Pro	Lys	Lys	Lys	Gly	Leu	Leu	Thr	Ala	Val	Val	Val	Asp	Asp	Lys
				485					490					495	
Pro	Cys	Glu	Leu	Tyr	Cys	Ser	Pro	Leu	Gly	Lys	Glu	Ser	Pro	Leu	Leu
			500					505					510		
Val	Ala	Asp	Arg	Val	Leu	Asp	Gly	Thr	Pro	Cys	Gly	Pro	Tyr	Glu	Thr
		515					520					525			
Asp	Leu	Cys	Val	His	Gly	Lys	Cys	Gln	Lys	Ile	Gly	Cys	Asp	Gly	Ile
	530					535					540				
Ile	Gly	Ser	Ala	Ala	Lys	Glu	Asp	Arg	Cys	Gly	Val	Cys	Ser	Gly	Asp
545						550					555				560
Gly	Lys	Thr	Cys	His	Leu	Val	Lys	Gly	Asp	Phe	Ser	His	Ala	Arg	Gly
				565					570					575	
Thr	Gly	Tyr	Ile	Glu	Ala	Ala	Val	Ile	Pro	Ala	Gly	Ala	Arg	Arg	Ile
			580					585					590		
Arg	Val	Val	Glu	Asp	Lys	Pro	Ala	His	Ser	Phe	Leu	Ala	Leu	Lys	Asp

WO 03/004607

PCT/US02/21056

595 600 605

Ser Gly Lys Gly Ser Ile Asn Ser Asp Trp Lys Ile Glu Leu Pro Gly
610 615 620

Glu Phe Gln Ile Ala Gly Thr Thr Val Arg Tyr Val Arg Arg Gly Leu
625 630 635 640

Trp Glu Lys Ile Ser Ala Lys Gly Pro Thr Lys Leu Pro Leu His Leu
645 650 655

Met Val Leu Leu Phe His Asp Gln Asp Tyr Gly Ile His Tyr Glu Tyr
660 665 670

Thr Val Pro Val Asn Arg Thr Ala Glu Asn Gln Ser Glu Pro Glu Lys
675 680 685

Pro Gln Asp Ser Leu Phe Ile Trp Thr His Ser Gly Trp Glu Gly Cys
690 695 700

Ser Val Gln Cys Gly Gly Gly Glu Arg Arg Thr Ile Val Ser Cys Thr
705 710 715 720

Arg Ile Val Asn Lys Thr Thr Thr Leu Val Asn Asp Ser Asp Cys Pro
725 730 735

Gln Ala Ser Arg Pro Glu Pro Gln Val Arg Arg Cys Asn Leu His Pro
740 745 750

Cys Gln Ser Arg
755

<210> 7

<211> 2339

<212> DNA

<213> Homo sapiens

<400> 7

```

ccggctccct cgtctcgctc agcgcctgcg gcgccgccgg cggcctgggt ggccctcattc 60
agcttggggca ggagcaggtg ctaatccagc ccctcaacaa ctcccagggc ccattcagtg 120
gacgagaaca tctgatcagg cgcaaattgt ccttgacccc cagcccttct gctgaggccc 180
agagacctga gcagctctgc aaggttctaa cagaaaagaa gaagccgacg tggggcaggc 240
cttcgcggga ctggcgggag cggaggaacg ctatccggct caccagcgag cacacgggtg 300
agaccctggt ggtggccgac gccgacatgg tgcagtacca cggggccgag gccgcccaga 360
ggttcatcct gaccgtcatg aacatggtat acaatatgtt tcagcaccag agcctgggga 420
ttaaaattaa cattcaagtg accaagcttg tcctgctacg acaacgtccc gctaagttgt 480
ccattgggca ccatggtgag cggctccctg agagcttctg tcaactggcag aacgaggagt 540
atggaggagc gcgatacctc ggcaataacc aggttcccgg cgggaaggac gacccgcccc 600
tggtggatgc tgctgtgttt gtgaccagga cagatttctg tgtacacaaa gatgaaccgt 660
gtgacactgt tggaattgct tacttaggag gtgtgtgcag tgctaagagg aagtgtgtgc 720
ttgccgaaga caatggtctc aatttggcct ttaccatcgc ccatgagctg ggccacaact 780
tgggcatgaa ccacgacgat gaccactcat cttgcgctgg caggtcccac atcatgtcag 840
gagagtgggt gaaaggccgg aacccaagtg acctctcttg gtcctcctgc agccgagatg 900
accttgaaaa cttcctcaag tcaaaagtca gcacctgctt gctagtcacg gaccccagaa 960
gccagcacac agtacgcctc ccgcacaagc tgccggggcat gcactacagt gccaacgagc 1020
agtgccagat cctgtttggc atgaatgcca ccttctgcag aaacatggag catctaattg 1080
gtgctggact gtggtgcctg gtagaaggag acacatcctg caagaccaag ctggaccctc 1140
ccctggatgg caccgagtgt ggggcagaca agtggtgccg cgcgggggag tgcgtgagca 1200
agacgcccat ccggagcat gtggacggag actggagccc gtggggcgcc tggagcatgt 1260

```

WO 03/004607

PCT/US02/21056

```

gcagccgaac atgtgggacg ggagcccgc tccggcagag gaagatgagc aagccccc 1320
ctgggcctgg aggcacacac tgcccgggtg ccagtgtaga acatgcggtc tgcgagaacc 1380
tgccctgccc caagggtctg cccagcttcc gggaccagca gtgccaggca cacgaccggc 1440
tgagcccca gaagaaaggc ctgctgacag ccgtggtggt tgacgataag ccatgtgaac 1500
tctactgctc gcccctcggg aaggagtccc cactgctggt ggccgacagg gtccctggacg 1560
gtacaccctg cgggcccctac gagactgatc tctgcgtgca cggcaagtgc cagaaaatcg 1620
gctgtgacgg catcatcggg tctgcagcca aagaggacag atgcgggggtc tgcagcgggg 1680
acggcaagac ctgccacttg gtgaagggcg acttcagcca cgcccggggg acagttaaga 1740
atgatctctg tacgaaggta tccacatgtg tgatggcaga ggctgttccc aagtgtttct 1800
catgttatat cgaagctgcc gtcattcctg ctggagctcg gaggatccgt gtggtggagg 1860
ataaacctgc ccacagcttt ctggctctca aagactcggg taaggggtcc atcaacagt 1920
actggaagat agagctcccc ggagagtctc agattgcagg cacaactgtt cgctatgtga 1980
gaaggggggt gtgggagaag atctctgcc aaggaccaac caaactaccg ctgcacttga 2040
tggtgttgtt atttcacgac caagattatg gaattcatta tgaatacact gttcctgtaa 2100
accgcactgc ggaaaatcaa agcgaaccag aaaaaccgca ggactctttg ttcactctga 2160
cccacagcgg ctgggaaggg tgcagtgtgc agtgcggcgg aggggagcgc agaaccatcg 2220
tctcgtgtac acggattgtc aacaagacca caactctggt gaacgacagt gactgccctc 2280
aagcaagccg cccagagccc caggtccgaa ggtgcaactt gcaccctgc cagtcacgt 2339

```

<210> 8

<211> 779

<212> PRT

<213> Homo sapiens

<400> 8

Gly Ser Leu Val Ser Leu Ser Ala Cys Gly Ala Ala Gly Gly Leu Val
1 5 10 15

Gly Leu Ile Gln Leu Gly Gln Glu Gln Val Leu Ile Gln Pro Leu Asn
20 25 30

Asn Ser Gln Gly Pro Phe Ser Gly Arg Glu His Leu Ile Arg Arg Lys
35 40 45

Trp Ser Leu Thr Pro Ser Pro Ser Ala Glu Ala Gln Arg Pro Glu Gln
50 55 60

Leu Cys Lys Val Leu Thr Glu Lys Lys Lys Pro Thr Trp Gly Arg Pro
65 70 75 80

Ser Arg Asp Trp Arg Glu Arg Arg Asn Ala Ile Arg Leu Thr Ser Glu
85 90 95

His Thr Val Glu Thr Leu Val Val Ala Asp Ala Asp Met Val Gln Tyr
100 105 110

His Gly Ala Glu Ala Ala Gln Arg Phe Ile Leu Thr Val Met Asn Met
115 120 125

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

Gln Val Thr Lys Leu Val Leu Leu Arg Gln Arg Pro Ala Lys Leu Ser
145 150 155 160

Ile Gly His His Gly Glu Arg Ser Leu Glu Ser Phe Cys His Trp Gln
165 170 175

Asn Glu Glu Tyr Gly Gly Ala Arg Tyr Leu Gly Asn Asn Gln Val Pro

11

WO 03/004607

PCT/US02/21056

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

<210> 9

<211> 5004

<212> DNA

<213> Homo sapiens

<400> 9

cgcacgcccc	cagccgcccc	gcgcgccccg	cccggagagc	ggcgcctgca	gctgcacctg	120
ccggccttcg	ggcgcgacct	gtaccttcag	ctgcgcgcgc	acctgcgctt	cctgtcccga	180
ggcttcgagg	tggaggaggc	gggcgcggcc	cggcgccgcg	gccgccccgc	cgagctgtgc	240
ttctactcgg	gccgtgtgct	cggccacccc	ggctccctcg	tctcgctcag	cgcttgcggc	300
gccgcccggc	gcctgggttg	cctcattcag	cttggggcagg	agcagggtgct	aatccagccc	360
ctcaacaact	cccagggccc	attcagtgga	cgagaacatc	tgatcaggcg	caaatgggtcc	420
ttgacccccca	gcccttctgc	tgaggcccag	agacctgagc	agctctgcaa	ggttctaaca	480
gaaaagaaga	agccgacgtg	gggcaggcct	tgcggggact	ggcggggagcg	gaggaacgct	540
atccgggtca	ccagcgagca	cacgggtggag	accctgggtg	tggccgacgc	cgacatgggtg	600
cagtaccacg	gggcccaggc	cgcccagagg	ttcatcctga	ccgtcatgaa	catgggtatac	660
aatatgtttc	agcaccagag	cctgggggatt	aaaattaaca	ttcaagtga	caagcttgtc	720
ctgctacgac	aacgtcccgc	taagttgtcc	attgggcacc	atgggtgagcg	gtccctggag	780
agcttctgtc	actggcagaa	cgaggagtat	ggaggagcgc	gatacctcgg	caataaccag	840
gttcccggcg	ggaaggacga	cccgcacctg	gtggatgctg	ctgtgtttgt	gaccaggaca	900
gatttctgtg	tacacaaaga	tgaaccgtgt	gacactgttg	gaattgctta	cttaggaggt	960
gtgtgcagtg	ctaagaggaa	gtgtgtgctt	gccgaagaca	atgggtctcaa	tttggccttt	1020
accatcgccc	atgagctggg	ccacaacttg	ggcatgaacc	acgacgatga	ccactcatct	1080
tgcgctggca	ggtcccacat	catgtcagga	gagtgggtga	aaggccggaa	cccaagtga	1140
ctctcttggg	cctcctgcag	ccgagatgac	cttgaanaact	tcctcaagtc	aaaagtcagc	1200
acctgcttgc	tagtcacgga	ccccagaagc	cagcacacag	tacgcctccc	gcacaagctg	1260
ccgggcatgc	actacagtgc	caacgagcag	tgccagatcc	tgtttggcat	gaatgccacc	1320
ttctgcagaa	acatggagca	tctaattgtgt	gctggactgt	ggtgcctggg	agaaggagac	1380
acatcctgca	agaccaagct	ggaccctccc	ctggatggca	ccgagtgtgg	ggcagacaag	1440
tgggtgccgc	cgggggagtg	cgtgagcaag	acgcccattc	cggagcatgt	ggacggagac	1500
tggagcccgt	ggggcgccctg	gagcatgtgc	agccgaacat	gtgggacggg	agcccgcctc	1560
cggcagagga	aatgtgacaa	ccccccccct	gggcctggag	gcacacactg	cccggtgccc	1620
agtgtagaac	atgcggtctg	cgagaacctg	ccctgccccca	aggggtctgcc	cagcttccgg	1680
gaccagcagt	gccaggcaca	cgaccggctg	agcccccaaga	agaaaggcct	gctgacagcc	1740
gtggtggttg	acgataagcc	atgtgaactc	tactgctcgc	ccctcgggaa	ggagtcccca	1800
ctgctgggtg	ccgacagggg	cctggacggg	acaccctgcg	ggccctacga	gactgatctc	1860
tgcgtgcacg	gcaagtgcc	gaaaatcggc	tgtgacggca	tcacggggtc	tgcagccaaa	1920
gaggacagat	gcgggggtctg	cagcggggac	ggcaagacct	gccacttggg	gaagggcgac	1980
ttcagccacg	cccggggggac	aggttatatc	gaagctgccg	tcattcctgc	tggagctcgg	2040
aggatccgtg	tgggtggagga	taaacctgcc	cacagctttc	tggctctcaa	agactcgggt	2100
aaggggtcca	tcaacagtga	ctggaagata	gagctccccg	gagagtcca	gattgcaggc	2160
acaactgttc	gctatgtgag	aaggggggctg	tgggagaaga	tctctgccaa	gggaccaacc	2220
aaactaccgc	tgcacttgat	ggtgttggtta	tttcacgacc	aagattatgg	aattcattat	2280
gaatacactg	ttcctgtaaa	ccgcactgcg	gaaaatcaaa	gcgaaccaga	aaaaccgcag	2340
gactctttgt	tcactctggac	ccacagcggc	tgggaagggt	gcagtgtgca	gtgcggcgga	2400
ggggagcgca	gaaccatcgt	ctcgtgtaca	cggattgtca	acaagaccac	aactctgggtg	2460
aacgacagtg	actgccctca	agcaagccgc	ccagagcccc	aggtccgaag	gtgcaacttg	2520
cacccttgc	agtcacgktg	ggtggcaggc	ccgtggagcc	cctgctcggc	gacctgtgag	2580
aaaggcttcc	agcaccggga	ggtgacctgc	gtgtaccagc	tgcagaacgg	cacacacgtc	2640
gctacgcggc	ccctctactg	cccggggccc	cggccggcgg	cagtgcagag	ctgtgaaggc	2700
caggactgcc	tgtccatctg	ggaggcgtct	gagtgggtcac	agtgtctgc	cagctgtggg	2760
aaaggggtgt	ggaaacggac	cgtggcgtgc	accaactcac	aagggaaatg	cgacgcatcc	2820
acgaggccga	gagccgagga	ggcctgcgag	gactactcag	gctgtctacga	gtggaaaact	2880
ggggactggg	ctacgtgctc	gtcgacctgc	gggaaggggc	tgcagtcccg	ggtgggtgcag	2940
tgcattgcaca	aggtcacagg	gcgccacggc	agcgagtgcc	ccgccctctc	gaagcctgcc	3000
ccctacagac	agtgtacca	ggaggtctgc	aacgacagga	tcaacgccaa	caccatcacc	3060
tccccccgccc	ttgctgtctc	gacctacaaa	tgcacacgag	accagtggac	ggtatatattg	3120
cgggtcatcc	gagaaaagaa	cctctgccag	gacatgcggg	ggtaccagcg	ctgctgccag	3180
acctgcaggg	acttctatgc	aaacaagatg	cgccagccac	cgccgagctc	gtgacacgca	3240
gtcccaaggg	tcgctcaaag	ctcagactca	ggtctgaaag	ccaccacccc	gcaagcctac	3300
cagccttgtg	gccacacccc	cacccggtcg	ccacaagaat	ccaactgcat	agaacatgag	3360
cgtggacttg	gcgtttgcca	ttagtgcttc	cgtacttaat	atattgttaa	cagccactgg	3420
ctcactttct	acagtgagga	gaaagtaggc	atgagtcaca	aagtaacttc	aatttctagg	3480
atttcaggta	cctcgaaggg	aagcacctct	ggcagacaa	cgtcaagaga	gagacatcat	3540
ttagtgttcc	tgtcttgact	cgcttttgac	atttgaattt	ccagtgtctg	gtatatcatg	3600
gaggaaacat	ccccaaaacg	agacatgcta	gaaaaggctt	tattctaaag	gctttattct	

WO 03/004607

PCT/US02/21056

```

gaaagccggc gacaccctgg agggaggggc aggtgttggt gacccctctg acccgggttc 3660
tctggggagg gccgggctgc ttagcccacg tttctcttca tctaccttct tgaccacatg 3720
agaaccagga cattgcctcc atgcccgtct ctgacaacat agtctctaaa tcctagggtg 3780
tgccttgga gtctcgtgcg tggagtgtaa atctatatat gccagcgagg acagcagtgc 3840
cacgcagttc ataccacccg catgggaaga atgttccaag agagtctggg tttggggaag 3900
catctaattt tcagagctct gctgtccacc gtgtaggga acagaagggc ctctcttcaa 3960
ggtgctgtga cataagaaac ggtaattgcg gtgatggggt tgcttcctaa ggcaaaggta 4020
agcttggggc agcttcactg gggcggatgg gcacctgcc cgccttccgc gagcatccac 4080
tctggcccgc acttcctaaa gctttgtacc ttagagatgc tgtaccacat cccagtggct 4140
ttctaccgac cgtggccatt tatctgaagg taagacgaca tttgggacct ctgaggacac 4200
aggcctagga tctgtagagc aaggcctgac tgctctatcc tggcacggag cagcctgata 4260
tgccgggacc aggggaggaa cgccatctgg ctggcactgc tgcacacccg ccgagccttc 4320
ctgtagcccc agactttgtg gtaccatta tcatcacgcc tgtcatcatt gacctatctt 4380
cttgggtggg caaggatgat gcatgatgaa ggtccttccc tcctgcagcc cccttacgcc 4440
tggcagcaga caagcagagt ggcctcgttg agagcacaga ggatggtagc accctacctg 4500
caaggaggcc gggcagggac cctagatgcc aggaggcctg ttttgctcac caacttgggtg 4560
ggcatttcat gggtgcttat gttctaggac tttaccgtaa ataacacctc ctccctgatt 4620
tcaggcagaa ggtctcactt ggacttccat gggatcatct ccctgtgttt cttgatttat 4680
tggtgctgtg tttctgtgtt ttgttttgtt acatgtcaca accgtagagt tagcttaaata 4740
cagaaagaag cctctctgcc ttctccaccc tgtcttacga gctgtgtttt tgtttttact 4800
accctagagg cagagaagcg gtagggatgt caggaattt actcacttcc acttgaatca 4860
acgagaagtg ttgagaaact tccgtgggtg ctctgtggaa agaaccgagg gtgtcaggat 4920
ggagcggccc accctcgccc cgcggcctgc gcagactgct gtcctcccct tcaggcctgg 4980
ccaccagcag actcccatga attc 5004

```

<210> 10

<211> 1057

<212> PRT

<213> Homo sapiens

<400> 10

```

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

```

WO 03/004607

PCT/US02/21056

145					150					155						
Ile	Arg	Leu	Thr	Ser	Glu	His	Thr	Val	Glu	Thr	Leu	Val	Val	Ala	Asp	
				165					170					175		
Ala	Asp	Met	Val	Gln	Tyr	His	Gly	Ala	Glu	Ala	Ala	Gln	Arg	Phe	Ile	
			180					185					190			
Leu	Thr	Val	Met	Asn	Met	Val	Tyr	Asn	Met	Phe	Gln	His	Gln	Ser	Leu	
		195					200					205				
Gly	Ile	Lys	Ile	Asn	Ile	Gln	Val	Thr	Lys	Leu	Val	Leu	Leu	Arg	Gln	
	210					215					220					
Arg	Pro	Ala	Lys	Leu	Ser	Ile	Gly	His	His	Gly	Glu	Arg	Ser	Leu	Glu	
225					230					235					240	
Ser	Phe	Cys	His	Trp	Gln	Asn	Glu	Glu	Tyr	Gly	Gly	Ala	Arg	Tyr	Leu	
				245					250					255		
Gly	Asn	Asn	Gln	Val	Pro	Gly	Gly	Lys	Asp	Asp	Pro	Pro	Leu	Val	Asp	
			260					265					270			
Ala	Ala	Val	Phe	Val	Thr	Arg	Thr	Asp	Phe	Cys	Val	His	Lys	Asp	Glu	
		275					280					285				
Pro	Cys	Asp	Thr	Val	Gly	Ile	Ala	Tyr	Leu	Gly	Gly	Val	Cys	Ser	Ala	
	290					295					300					
Lys	Arg	Lys	Cys	Val	Leu	Ala	Glu	Asp	Asn	Gly	Leu	Asn	Leu	Ala	Phe	
305					310					315					320	
Thr	Ile	Ala	His	Glu	Leu	Gly	His	Asn	Leu	Gly	Met	Asn	His	Asp	Asp	
				325					330					335		
Asp	His	Ser	Ser	Cys	Ala	Gly	Arg	Ser	His	Ile	Met	Ser	Gly	Glu	Trp	
			340					345					350			
Val	Lys	Gly	Arg	Asn	Pro	Ser	Asp	Leu	Ser	Trp	Ser	Ser	Cys	Ser	Arg	
		355					360					365				
Asp	Asp	Leu	Glu	Asn	Phe	Leu	Lys	Ser	Lys	Val	Ser	Thr	Cys	Leu	Leu	
	370					375					380					
Val	Thr	Asp	Pro	Arg	Ser	Gln	His	Thr	Val	Arg	Leu	Pro	His	Lys	Leu	
385					390					395					400	
Pro	Gly	Met	His	Tyr	Ser	Ala	Asn	Glu	Gln	Cys	Gln	Ile	Leu	Phe	Gly	
				405					410					415		
Met	Asn	Ala	Thr	Phe	Cys	Arg	Asn	Met	Glu	His	Leu	Met	Cys	Ala	Gly	
			420				425						430			
Leu	Trp	Cys	Leu	Val	Glu	Gly	Asp	Thr	Ser	Cys	Lys	Thr	Lys	Leu	Asp	
		435					440					445				
Pro	Pro	Leu	Asp	Gly	Thr	Glu	Cys	Gly	Ala	Asp	Lys	Trp	Cys	Arg	Ala	
	450					455					460					
Gly	Glu	Cys	Val	Ser	Lys	Thr	Pro	Ile	Pro	Glu	His	Val	Asp	Gly	Asp	

WO 03/004607

PCT/US02/21056

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

WO 03/004607

PCT/US02/21056

785		790		795		800									
Asn	Asp	Ser	Asp	Cys	Pro	Gln	Ala	Ser	Arg	Pro	Glu	Pro	Gln	Val	Arg
				805					810					815	
Arg	Cys	Asn	Leu	His	Pro	Cys	Gln	Ser	Arg	Trp	Val	Ala	Gly	Pro	Trp
			820					825					830		
Ser	Pro	Cys	Ser	Ala	Thr	Cys	Glu	Lys	Gly	Phe	Gln	His	Arg	Glu	Val
		835					840					845			
Thr	Cys	Val	Tyr	Gln	Leu	Gln	Asn	Gly	Thr	His	Val	Ala	Thr	Arg	Pro
	850					855					860				
Leu	Tyr	Cys	Pro	Gly	Pro	Arg	Pro	Ala	Ala	Val	Gln	Ser	Cys	Glu	Gly
865					870					875					880
Gln	Asp	Cys	Leu	Ser	Ile	Trp	Glu	Ala	Ser	Glu	Trp	Ser	Gln	Cys	Ser
				885					890					895	
Ala	Ser	Cys	Gly	Lys	Gly	Val	Trp	Lys	Arg	Thr	Val	Ala	Cys	Thr	Asn
			900					905					910		
Ser	Gln	Gly	Lys	Cys	Asp	Ala	Ser	Thr	Arg	Pro	Arg	Ala	Glu	Glu	Ala
		915					920					925			
Cys	Glu	Asp	Tyr	Ser	Gly	Cys	Tyr	Glu	Trp	Lys	Thr	Gly	Asp	Trp	Ser
	930					935					940				
Thr	Cys	Ser	Ser	Thr	Cys	Gly	Lys	Gly	Leu	Gln	Ser	Arg	Val	Val	Gln
945					950					955					960
Cys	Met	His	Lys	Val	Thr	Gly	Arg	His	Gly	Ser	Glu	Cys	Pro	Ala	Leu
				965					970					975	
Ser	Lys	Pro	Ala	Pro	Tyr	Arg	Gln	Cys	Tyr	Gln	Glu	Val	Cys	Asn	Asp
			980					985					990		
Arg	Ile	Asn	Ala	Asn	Thr	Ile	Thr	Ser	Pro	Arg	Leu	Ala	Ala	Leu	Thr
		995					1000					1005			
Tyr	Lys	Cys	Thr	Arg	Asp	Gln	Trp	Thr	Val	Tyr	Cys	Arg	Val	Ile	Arg
	1010					1015					1020				
Glu	Lys	Asn	Leu	Cys	Gln	Asp	Met	Arg	Trp	Tyr	Gln	Arg	Cys	Cys	Gln
1025					1030					1035					1040
Thr	Cys	Arg	Asp	Phe	Tyr	Ala	Asn	Lys	Met	Arg	Gln	Pro	Pro	Pro	Ser
			1045					1050						1055	
Ser															

<210> 11
 <211> 3369
 <212> DNA
 <213> Homo sapiens

<400> 11

atgtgtgacg	gcgccctgct	gcctccgctc	gtcctgcccc	tgctgctgct	gctgggtttgg	60
ggactggacc	cgggcacagc	tgtcggcgac	gcggcgggccg	acgtggaggt	ggtgctcccc	120
tggcgggtgc	gccccgacga	cgtgcacctg	ccgccgctgc	ccgcagcccc	cgggccccga	180
cggcggcgac	gcccccgcac	gccccagacc	gccccgcgcg	cccggccccg	agagcgcgcc	240
ctgctgctgc	acctgccggc	cttcggggcg	gacctgtacc	ttcagctgcg	ccgcgacctg	300
cgcttcctgt	cccgaggctt	cgagggtggag	gaggcgggcg	cgggccccg	ccgcggccgc	360
cccgccgagc	tgtgcttcta	ctcggggcct	gtgctcggcc	accccggtc	cctcgtctcg	420
ctcagcgctt	gcggcgccgc	cggcggcctg	gttggcctca	ttcagcttgg	gcaggagcag	480
gtgctaatac	agccccctca	caactcccag	ggcccattca	gtggacgaga	acatctgac	540
aggcgcaaat	ggtccttgac	ccccagccct	tctgctgagg	cccagagacc	tgagcagctc	600
tgcaagggtc	taacagaaaa	gaagaagccg	acgtggggca	ggccttcgcg	ggactggcgg	660
gagcggagga	acgctatccg	gctcaccagc	gagcacacgg	tggagaccct	ggtgggtggc	720
gacgccgaca	tgggtgcagta	ccacggggcc	gaggccgccc	agaggttcat	cctgaccgtc	780
atgaacatgg	tatacaatat	gtttcagcac	cagagcctgg	ggattaaaat	taacattcaa	840
gtgaccaagc	ttgtcctgct	acgacaacgt	cccgtctaagt	tgtccattgg	gcaccatggt	900
gagcgggtcc	tggagagctt	ctgtcactgg	cagaacgagg	agtatggagg	agcgcgatac	960
ctcgggaata	accagggttc	cggcgggaag	gacgaccgc	ccctgggtgga	tgctgctgtg	1020
tttgtgacca	ggacagattt	ctgtgtacac	aaagatgaac	cgtgtgacac	tggttggatt	1080
gcttacttag	gaggtgtgtg	cagtgttaag	aggaagtgtg	tgcttgccga	agacaatggt	1140
ctcaatttgg	cctttaccat	cgcccatgag	ctggggccaca	acttgggcat	gaaccacgac	1200
gatgaccact	catcttgccg	tggcagggtc	cacatcatgt	caggagagtg	ggtgaaaggc	1260
cggaacccaa	gtgacctctc	ttggtcctcc	tgcagccgag	atgacctga	aaacttcctc	1320
aagtcaaaaag	tcagcacctg	cttgctagtc	acggacccca	gaagccagca	cacagtacgc	1380
ctcccgcaca	agctgccggg	catgcactac	agtgccaacg	agcagtgcc	gatcctgttt	1440
ggcatgaatg	ccaccttctg	cagaaacatg	gagcatctaa	tgtgtgctgg	actgtggtgc	1500
ctggtagaag	gagacacatc	ctgcaagacc	aagctggacc	ctcccttgg	tggcaccgag	1560
tgtggggcag	acaagtgggtg	ccgcgcgggg	gagtgcgtga	gcaagacgcc	catcccggag	1620
catgtggacg	gagactggag	cccgtggggc	gcctggagca	tgtgcagccg	aacatgtggg	1680
acgggagccc	gcttcgggca	gaggaaatgt	gacaaccccc	cccctggggc	tggaggcaca	1740
cactgcccg	gtgccagtgt	agaacatgcg	gtctgcgaga	acctgccctg	ccccagggt	1800
ctgcccagct	tccgggacca	gcagtgccag	gcacacgacc	ggctgagccc	caagaagaaa	1860
ggcctgctga	cagccgtggg	ggttgacgat	aagccatgtg	aactctactg	ctcgccccctc	1920
gggaaggagt	ccccactgct	ggtggccgac	agggtcctgg	acggtacacc	ctgcggggccc	1980
tacgagactg	atctctgcgt	gcacggcaag	tgccagaaaa	tgggctgtga	cggcatcatc	2040
gggtctgcag	ccaaagagga	cagatgcggg	gtctgcagcg	gggacggcaa	gacctgccac	2100
ttggtgaagg	gcgacttcag	ccacgcccgg	gggacagggt	atatcgaagc	tgccgtcatt	2160
cctgctggag	ctcggaggat	ccgtgtgggtg	gaggataaac	ctgcccacag	ctttctggct	2220
ctcaaagact	cgggttaagg	gtccatcaac	agtgactgga	agatagagct	ccccggagag	2280
ttccagattg	caggcacaac	tgttcgctat	gtgagaagg	ggctgtggga	gaagatctct	2340
gccaaggggac	caaccaaact	accgctgcac	ttgatgggtg	tgttatttca	cgaccaagat	2400
tatggaattc	attatgaata	cactgttcct	gtaaaccgca	ctgcggaaaa	tcaaagcgaa	2460
ccagaaaaaac	cgcaggactc	tttgttcatc	tggacccaca	gcggctggga	agggtgcagt	2520
gtgcagtgcg	gcggaggggga	gcgcagaacc	atcgtctcgt	gtacacggat	tgtcaacaag	2580
accacaactc	tgggtgaacga	cagtgcactgc	cctcaagcaa	gccgcccaga	gccccagggtc	2640
cgaagggtgca	acttgcaccc	ctgccagtca	cgktgggtgg	caggcccgtg	gagccccctgc	2700
tcggcgacct	gtgagaaagg	cttccagcac	cgggagggtga	cctgcgtgta	ccagctgcag	2760
aacggcacac	acgtcgctac	gcggccccctc	tactgcccg	gcccccgcc	ggcggcagtg	2820
cagagctgtg	aaggccagga	ctgcctgtcc	atctgggagg	cgtctgagtg	gtcacagtgc	2880
tctgccagct	gtggtaaagg	ggtgtggaaa	cggaccgtgg	cgtgcaccaa	ctcacaagg	2940
aaatgcgacg	catccacgag	gccgagagcc	gaggaggcct	gcgaggacta	ctcaggctgc	3000
tacgagtgg	aaactgggga	ctggtctacg	tgctcgtcga	cctgcgggaa	gggcctgcag	3060
tcccgggtgg	tgcagtgcac	gcacaaggtc	acagggcgcc	acggcagcga	gtgccccgcc	3120
ctctcgaagc	ctgcccccta	cagacagtgc	taccaggagg	tctgcaacga	caggatcaac	3180
gccaacacca	tcacctcccc	ccgccttgct	gctctgacct	acaaatgcac	acgagaccag	3240
tggacgggat	attgccgggt	catccgagaa	aagaacctct	gccaggacat	gcggtggtac	3300
cagcgtgct	gccagacctg	cagggacttc	tatgcaaaaa	agatgcgcca	gccaccgccc	3360
agctcgtga						3369

WO 03/004607

PCT/US02/21056

<210> 12
 <211> 200
 <212> DNA
 <213> Homo sapiens

<400> 12
 ttggtgaagg gcgacttcag ccacgcccgg gggacagtta agaatgatct ctgtacgaag 60
 gtatccacat gtgtgatggc agaggctgtt cccaagtgtt tctcatgtta tatcgaagct 120
 gccgtcattc ctgctggagc tcggaggatc cgtgtggtgg aggataaacc tgcccacage 180
 tttctggctc tcaaagactc 200

<210> 13
 <211> 1122
 <212> PRT
 <213> Homo sapiens

<400> 13
 Met Cys Asp Gly Ala Leu Leu Pro Pro Leu Val Leu Pro Val Leu Leu
 1 5 10 15
 Leu Leu Val Trp Gly Leu Asp Pro Gly Thr Ala Val Gly Asp Ala Ala
 20 25 30
 Ala Asp Val Glu Val Val Leu Pro Trp Arg Val Arg Pro Asp Asp Val
 35 40 45
 His Leu Pro Pro Leu Pro Ala Ala Pro Gly Pro Arg Arg Arg Arg Arg
 50 55 60
 Pro Arg Thr Pro Pro Ala Ala Pro Arg Ala Arg Pro Gly Glu Arg Ala
 65 70 75 80
 Leu Leu Leu His Leu Pro Ala Phe Gly Arg Asp Leu Tyr Leu Gln Leu
 85 90 95
 Arg Arg Asp Leu Arg Phe Leu Ser Arg Gly Phe Glu Val Glu Glu Ala
 100 105 110
 Gly Ala Ala Arg Arg Arg Gly Arg Pro Ala Glu Leu Cys Phe Tyr Ser
 115 120 125
 Gly Arg Val Leu Gly His Pro Gly Ser Leu Val Ser Leu Ser Ala Cys
 130 135 140
 Gly Ala Ala Gly Gly Leu Val Gly Leu Ile Gln Leu Gly Gln Glu Gln
 145 150 155 160
 Val Leu Ile Gln Pro Leu Asn Asn Ser Gln Gly Pro Phe Ser Gly Arg
 165 170 175
 Glu His Leu Ile Arg Arg Lys Trp Ser Leu Thr Pro Ser Pro Ser Ala
 180 185 190
 Glu Ala Gln Arg Pro Glu Gln Leu Cys Lys Val Leu Thr Glu Lys Lys
 195 200 205
 Lys Pro Thr Trp Gly Arg Pro Ser Arg Asp Trp Arg Glu Arg Arg Asn
 210 215 220

Ala 225	Ile	Arg	Leu	Thr	Ser 230	Glu	His	Thr	Val	Glu 235	Thr	Leu	Val	Val	Ala 240
Asp	Ala	Asp	Met	Val 245	Gln	Tyr	His	Gly	Ala 250	Glu	Ala	Ala	Gln	Arg	Phe 255
Ile	Leu	Thr	Val 260	Met	Asn	Met	Val	Tyr 265	Asn	Met	Phe	Gln	His	Gln	Ser
Leu	Gly	Ile	Lys	Ile	Asn	Ile	Gln	Val	Thr	Lys	Leu	Val 285	Leu	Leu	Arg
Gln 290	Arg	Pro	Ala	Lys	Leu	Ser 295	Ile	Gly	His	His	Gly 300	Glu	Arg	Ser	Leu
Glu 305	Ser	Phe	Cys	His	Trp 310	Gln	Asn	Glu	Glu	Tyr 315	Gly	Gly	Ala	Arg	Tyr 320
Leu	Gly	Asn	Asn	Gln 325	Val	Pro	Gly	Gly	Lys 330	Asp	Asp	Pro	Pro	Leu	Val 335
Asp	Ala	Ala	Val 340	Phe	Val	Thr	Arg	Thr 345	Asp	Phe	Cys	Val	His 350	Lys	Asp
Glu	Pro	Cys 355	Asp	Thr	Val	Gly	Ile 360	Ala	Tyr	Leu	Gly	Gly 365	Val	Cys	Ser
Ala 370	Lys	Arg	Lys	Cys	Val	Leu 375	Ala	Glu	Asp	Asn	Gly 380	Leu	Asn	Leu	Ala
Phe 385	Thr	Ile	Ala	His	Glu 390	Leu	Gly	His	Asn	Leu 395	Gly	Met	Asn	His	Asp 400
Asp	Asp	His	Ser	Ser 405	Cys	Ala	Gly	Arg	Ser 410	His	Ile	Met	Ser	Gly 415	Glu
Trp	Val	Lys	Gly 420	Arg	Asn	Pro	Ser	Asp 425	Leu	Ser	Trp	Ser	Ser 430	Cys	Ser
Arg	Asp	Asp 435	Leu	Glu	Asn	Phe	Leu 440	Lys	Ser	Lys	Val	Ser 445	Thr	Cys	Leu
Leu 450	Val	Thr	Asp	Pro	Arg	Ser 455	Gln	His	Thr	Val	Arg 460	Leu	Pro	His	Lys
Leu 465	Pro	Gly	Met	His	Tyr 470	Ser	Ala	Asn	Glu	Gln 475	Cys	Gln	Ile	Leu	Phe 480
Gly	Met	Asn	Ala	Thr 485	Phe	Cys	Arg	Asn	Met 490	Glu	His	Leu	Met	Cys 495	Ala
Gly	Leu	Trp	Cys 500	Leu	Val	Glu	Gly	Asp 505	Thr	Ser	Cys	Lys	Thr 510	Lys	Leu
Asp	Pro	Pro 515	Leu	Asp	Gly	Thr	Glu 520	Cys	Gly	Ala	Asp	Lys 525	Trp	Cys	Arg
Ala 530	Gly	Glu	Cys	Val	Ser	Lys 535	Thr	Pro	Ile	Pro	Glu 540	His	Val	Asp	Gly

Asp 545	Trp	Ser	Pro	Trp	Gly 550	Ala	Trp	Ser	Met	Cys 555	Ser	Arg	Thr	Cys	Gly 560
Thr	Gly	Ala	Arg	Phe 565	Arg	Gln	Arg	Lys	Cys 570	Asp	Asn	Pro	Pro	Pro	Gly 575
Pro	Gly	Gly	Thr 580	His	Cys	Pro	Gly	Ala 585	Ser	Val	Glu	His	Ala 590	Val	Cys
Glu	Asn	Leu 595	Pro	Cys	Pro	Lys	Gly 600	Leu	Pro	Ser	Phe	Arg 605	Asp	Gln	Gln
Cys	Gln 610	Ala	His	Asp	Arg	Leu 615	Ser	Pro	Lys	Lys	Lys	Gly 620	Leu	Leu	Thr
Ala 625	Val	Val	Val	Asp	Asp 630	Lys	Pro	Cys	Glu	Leu 635	Tyr	Cys	Ser	Pro	Leu 640
Gly	Lys	Glu	Ser	Pro 645	Leu	Leu	Val	Ala	Asp 650	Arg	Val	Leu	Asp	Gly 655	Thr
Pro	Cys	Gly	Pro 660	Tyr	Glu	Thr	Asp	Leu 665	Cys	Val	His	Gly	Lys 670	Cys	Gln
Lys	Ile	Gly 675	Cys	Asp	Gly	Ile	Ile 680	Gly	Ser	Ala	Ala	Lys 685	Glu	Asp	Arg
Cys	Gly 690	Val	Cys	Ser	Gly	Asp 695	Gly	Lys	Thr	Cys	His 700	Leu	Val	Lys	Gly
Asp 705	Phe	Ser	His	Ala	Arg 710	Gly	Thr	Gly	Tyr	Ile 715	Glu	Ala	Ala	Val	Ile 720
Pro	Ala	Gly	Ala	Arg 725	Arg	Ile	Arg	Val	Val 730	Glu	Asp	Lys	Pro	Ala 735	His
Ser	Phe	Leu	Ala 740	Leu	Lys	Asp	Ser	Gly 745	Lys	Gly	Ser	Ile	Asn 750	Ser	Asp
Trp	Lys	Ile 755	Glu	Leu	Pro	Gly	Glu 760	Phe	Gln	Ile	Ala	Gly 765	Thr	Thr	Val
Arg 770	Tyr	Val	Arg	Arg	Gly	Leu 775	Trp	Glu	Lys	Ile	Ser 780	Ala	Lys	Gly	Pro
Thr 785	Lys	Leu	Pro	Leu	His 790	Leu	Met	Val	Leu	Leu 795	Phe	His	Asp	Gln	Asp 800
Tyr	Gly	Ile	His	Tyr 805	Glu	Tyr	Thr	Val	Pro 810	Val	Asn	Arg	Thr	Ala 815	Glu
Asn	Gln	Ser	Glu 820	Pro	Glu	Lys	Pro	Gln 825	Asp	Ser	Leu	Phe	Ile 830	Trp	Thr
His	Ser	Gly 835	Trp	Glu	Gly	Cys	Ser 840	Val	Gln	Cys	Gly	Gly 845	Gly	Glu	Arg
Arg 850	Thr	Ile	Val	Ser	Cys	Thr 855	Arg	Ile	Val	Asn	Lys 860	Thr	Thr	Thr	Leu

Val Asn Asp Ser Asp Cys Pro Gln Ala Ser Arg Pro Glu Pro Gln Val
 865 870 875 880
 Arg Arg Cys Asn Leu His Pro Cys Gln Ser Arg Trp Val Ala Gly Pro
 885 890 895
 Trp Ser Pro Cys Ser Ala Thr Cys Glu Lys Gly Phe Gln His Arg Glu
 900 905 910
 Val Thr Cys Val Tyr Gln Leu Gln Asn Gly Thr His Val Ala Thr Arg
 915 920 925
 Pro Leu Tyr Cys Pro Gly Pro Arg Pro Ala Ala Val Gln Ser Cys Glu
 930 935 940
 Gly Gln Asp Cys Leu Ser Ile Trp Glu Ala Ser Glu Trp Ser Gln Cys
 945 950 955 960
 Ser Ala Ser Cys Gly Lys Gly Val Trp Lys Arg Thr Val Ala Cys Thr
 965 970 975
 Asn Ser Gln Gly Lys Cys Asp Ala Ser Thr Arg Pro Arg Ala Glu Glu
 980 985 990
 Ala Cys Glu Asp Tyr Ser Gly Cys Tyr Glu Trp Lys Thr Gly Asp Trp
 995 1000 1005
 Ser Thr Cys Ser Ser Thr Cys Gly Lys Gly Leu Gln Ser Arg Val Val
 1010 1015 1020
 Gln Cys Met His Lys Val Thr Gly Arg His Gly Ser Glu Cys Pro Ala
 1025 1030 1035 1040
 Leu Ser Lys Pro Ala Pro Tyr Arg Gln Cys Tyr Gln Glu Val Cys Asn
 1045 1050 1055
 Asp Arg Ile Asn Ala Asn Thr Ile Thr Ser Pro Arg Leu Ala Ala Leu
 1060 1065 1070
 Thr Tyr Lys Cys Thr Arg Asp Gln Trp Thr Val Tyr Cys Arg Val Ile
 1075 1080 1085
 Arg Glu Lys Asn Leu Cys Gln Asp Met Arg Trp Tyr Gln Arg Cys Cys
 1090 1095 1100
 Gln Thr Cys Arg Asp Phe Tyr Ala Asn Lys Met Arg Gln Pro Pro Pro
 1105 1110 1115 1120
 Ser Ser

<210> 14
 <211> 265
 <212> PRT
 <213> Homo sapiens

<400> 14
 Leu Pro Ser Phe Arg Asp Gln Gln Cys Gln Ala His Asp Arg Leu Ser

```
<220>
<223> Description of Unknown Organism: Illustrative zinc
binding signature region
```

WO 03/004607

PCT/US02/21056

<400> 15

Thr Ala Ala His Glu Leu Gly His Val Lys Phe
 1 5 10

<210> 16

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 16

catgggcagc tcgag

15

<210> 17

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 17

ctgcaggcga gcctgaattc ctgagccat catg

34

<210> 18

<211> 68

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 18

cgagggttaa aaacgtctag gccccccgaa ccacggggac gtgggtttcc tttgaaaaac 60
 acgattgc 68

<210> 19

<211> 3438

<212> DNA

<213> Homo sapiens

<400> 19

atgtgtgacg gcgccctgct gcctccgctc gtcctgcccg tgctgctgct gctgggtttgg 60
 ggactggacc cgggcacagc tgtcggcgac gcggcgggccg acgtggaggt ggtgctcccg 120
 tggcgggtgc gccccgacga cgtgcacctg ccgccgctgc ccgcagcccc cgggccccga 180
 cggcggcgac gcccccgac gccccagcc gccccgcgcg cccggccccg agagcgcgcc 240
 ctgctgctgc acctgccggc cttcggggcg gacctgtacc ttcagctgcg ccgcgacctg 300
 cgcttcctgt cccgaggctt cgagggtggag gaggcggggcg cggccccggcg ccgcggccgc 360
 cccgcccagc tgtgcttcta ctcgggccgt gtgctcggcc accccggctc cctcgtctcg 420
 ctcagcgcct gcggcgccgc cggcggcctg gttggcctca ttcagcttgg gcaggagcag 480
 gtgctaatac agcccctcaa caactcccag ggcccattca gtggacgaga acatctgatc 540

WO 03/004607

PCT/US02/21056

```

aggcgcaaata ggtccttgac cccagccct tctgctgagg cccagagac cccagagac 660
tgcaagggttc taacagaaaa gaagaagccg acgtggggca ggccttcgcg ggactggcgg 720
gagcggagga acgctatccg gctcaccagc gagcacacgg tggagaccct ggtggtggcc 780
gacgccgaca tgggtgcagta ccacggggcc gaggccgccc agaggttcat cctgaccgtc 840
atgaacatgg tatacaatat gtttcagcac cagagcctgg ggattaaaat taacattcaa 900
gtgaccaagc ttgtcctgct acgacaacgt cccgctaagt tgtccattgg gcaccatggt 960
gagcgggtccc tggagagctt ctgtcactgg cagaacgagg agtatggagg agcgcgatac 1020
ctcggcaata accagggttc cggcgggaag gacgaccgc ccctggtgga tgctgctgtg 1080
tttgtgacca ggacagattt ctgtgtacac aaagatgaac cgtgtgacac tgttggaatt 1140
gcttacttag gaggtgtgtg cagtgtctaa aggaagtgtg tgcttgccga agacaatggt 1200
ctcaatttgg cctttaccat cgcccatgag ctggggccaca acttgggcat gaaccacgac 1260
gatgaccact catcttgccg tggcagggtc cacatcatgt caggagagtg ggtgaaaggc 1320
cggaacccaa gtgacctctc ttggtcctcc tgcagccgag atgaccttga aaacttcctc 1380
aagtcaaaag tcagcacctg cttgctagtc acggacccca gaagccagca cacagtacgc 1440
ctcccgacaca agctgccggg catgcactac agtgccaacg agcagtgcc aatcctgttt 1500
ggcatgaatg ccaccttctg cagaaacatg gagcatctaa tgtgtgctgg actgtggtgc 1560
ctggtagaag gagacacatc ctgcaagacc aagctggacc ctcccctgga tggcaccgag 1620
tgtggggcag acaagtgggt ccgcgcgggg gagtgcgtag gcaagacgcc catcccgag 1680
catgtggacg gagactggag cccgtggggc gcctggagca tgtgcagccg aacatgtggg 1740
acgggagccc gcttccggca gaggaatgt gacaaccccc cccctggggc tggaggcaca 1800
cactgcccgg gtgccagtgt agaacatgcg gtctgcgaga acctgccctg cccaagggt 1860
ctgcccagct tccgggacca gcagtgccag gcacacgacc ggctgagccc caagaagaaa 1920
ggcctgctga cagccgtggg ggttgacgat aagccatgtg aactctactg ctgcggggcc 1980
gggaaggagt cccactgct ggtggccgac agggctcctg acggtacacc ctgcggggcc 2040
tacgagactg atctctgcgt gcacggcaag tgccagaaaa tcggctgtga cggcatcatc 2100
gggtctgcag ccaaagagga cagatgcggg gtctgcagcg gggacggcaa gacctgccac 2160
ttggtgaagg ggcacttcag ccacgcccgg gggacagtta agaattgatc ctgtacgaag 2220
gtatccacat gtgtgatggc agaggctgtt cccaagtgtt tctcatgtta tatcgaagct 2280
gccgtcattc ctgctggagc tcggaggatc cgtgtggtgg aggataaacc tgcccacagc 2340
tttctggctc tcaaagactc gggtaagggg tccatcaaca gtgactggaa gatagagctc 2400
cccggagagt tccagattgc aggcacaact gttcgctatg tgagaagggg gctgtgggag 2460
aagatctctg ccaagggacc aaccaaacta ccgctgcact tgatggtgtt gttatttcac 2520
gaccaagatt atggaattca ttatgaatac actgttcctg taaaccgcac tgcggaaaat 2580
caaagcgaac cagaaaaacc gcaggactct ttgttcatct ggacccacag cggctgggaa 2640
gggtgcagtg tgcagtgcgg cggagggggg cgcagaacca tcgtctcgtg tacacggatt 2700
gtcaacaaga ccacaactct ggtgaacgac agtgactgcc ctcaagcaag ccgcccagag 2760
ccccagggtc gaagggtgcaa cttgcacccc tgccagtcac gktgggtggc aggcccgtgg 2820
agcccctgct cggcgacctg tgagaaaggc ttccagcacc gggagggtgac ctgcgtgtac 2880
cagctgcaga acggcacaca cgtcgctacg cggcccctct actgccggg gtctgagtgg 2940
gcggcagtg agagctgtga aggccaggac tgccctgtcc tctgggaggc gtgcaccaac 3000
tcacagtgt ctgccagctg tggtaaaggg gtgtggaaac ggaccgtggc cgaggactac 3060
tcacaaggga aatgcgacgc atccacgagg cagagagccg aggaggcctg ctgcgggaag 3120
tcaggctgt acgagtggaa aactggggac tgggtctacgt gctcgtcgac cggcagcgag 3180
ggcctgcagt cccgggtggg gcagtgcatt cacaaggtca cagggcgcca ctgcaacgac 3240
tgccccgccc tctcgaagcc tgccccctac agacagtgt accaggaggt ctgcaacgac 3300
aggatcaacg ccaacaccat cacctcccc cgccttgctg ctctgacct acaatgcaca 3360
cgagaccagt ggacggtata ttgccgggtc atccgagaaa agaacctctg ccaggacatg 3420
cgggtggtacc agcgtgctg ccagacctgc agggacttct atgcaaacia gatgcgccag 3438
ccaccgccga gctcgtga

```

<210> 20

<211> 1145

<212> PRT

<213> Homo sapiens

<400> 20

Met Cys Asp Gly Ala Leu Leu Pro Pro Leu Val Leu Pro Val Leu Leu
1 5 10 15

WO 03/004607

PCT/US02/21056

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

WO 03/004607

PCT/US02/21056

Asp	Ala	Ala	Val	Phe	~~Val~~	Thr	Arg	Thr	Asp	Phe	Cys	~~Val~~	~~His~~	~~Lys~~	~~Asp~~	
			340					345						350		
Glu	Pro	Cys	Asp	Thr	Val	Gly	Ile	Ala	Tyr	Leu	Gly	Gly	Val	Cys	Ser	
		355					360					365				
Ala	Lys	Arg	Lys	Cys	Val	Leu	Ala	Glu	Asp	Asn	Gly	Leu	Asn	Leu	Ala	
	370					375					380					
Phe	Thr	Ile	Ala	His	Glu	Leu	Gly	His	Asn	Leu	Gly	Met	Asn	His	Asp	
385					390					395					400	
Asp	Asp	His	Ser	Ser	Cys	Ala	Gly	Arg	Ser	His	Ile	Met	Ser	Gly	Glu	
				405					410					415		
Trp	Val	Lys	Gly	Arg	Asn	Pro	Ser	Asp	Leu	Ser	Trp	Ser	Ser	Cys	Ser	
			420					425					430			
Arg	Asp	Asp	Leu	Glu	Asn	Phe	Leu	Lys	Ser	Lys	Val	Ser	Thr	Cys	Leu	
		435					440					445				
Leu	Val	Thr	Asp	Pro	Arg	Ser	Gln	His	Thr	Val	Arg	Leu	Pro	His	Lys	
	450					455					460					
Leu	Pro	Gly	Met	His	Tyr	Ser	Ala	Asn	Glu	Gln	Cys	Gln	Ile	Leu	Phe	
465					470					475					480	
Gly	Met	Asn	Ala	Thr	Phe	Cys	Arg	Asn	Met	Glu	His	Leu	Met	Cys	Ala	
				485					490					495		
Gly	Leu	Trp	Cys	Leu	Val	Glu	Gly	Asp	Thr	Ser	Cys	Lys	Thr	Lys	Leu	
			500					505					510			
Asp	Pro	Pro	Leu	Asp	Gly	Thr	Glu	Cys	Gly	Ala	Asp	Lys	Trp	Cys	Arg	
		515					520					525				
Ala	Gly	Glu	Cys	Val	Ser	Lys	Thr	Pro	Ile	Pro	Glu	His	Val	Asp	Gly	
	530					535					540					
Asp	Trp	Ser	Pro	Trp	Gly	Ala	Trp	Ser	Met	Cys	Ser	Arg	Thr	Cys	Gly	
545					550					555					560	
Thr	Gly	Ala	Arg	Phe	Arg	Gln	Arg	Lys	Cys	Asp	Asn	Pro	Pro	Pro	Gly	
				565					570						575	
Pro	Gly	Gly	Thr	His	Cys	Pro	Gly	Ala	Ser	Val	Glu	His	Ala	Val	Cys	
			580					585					590			
Glu	Asn	Leu	Pro	Cys	Pro	Lys	Gly	Leu	Pro	Ser	Phe	Arg	Asp	Gln	Gln	
		595					600					605				
Cys	Gln	Ala	His	Asp	Arg	Leu	Ser	Pro	Lys	Lys	Lys	Gly	Leu	Leu	Thr	
	610					615					620					
Ala	Val	Val	Val	Asp	Asp	Lys	Pro	Cys	Glu	Leu	Tyr	Cys	Ser	Pro	Leu	
625					630					635					640	
Gly	Lys	Glu	Ser	Pro	Leu	Leu	Val	Ala	Asp	Arg	Val	Leu	Asp	Gly	Thr	
				645					650					655		

Pro	Cys	Gly	Pro	Tyr	~~Glu~~	Thr	Asp	Leu	Cys	Val	His	~~Gly~~	~~Lys~~	~~Cys~~	~~Gln~~
			660					665					670		
Lys	Ile	Gly	Cys	Asp	Gly	Ile	Ile	Gly	Ser	Ala	Ala	Lys	Glu	Asp	Arg
		675					680					685			
Cys	Gly	Val	Cys	Ser	Gly	Asp	Gly	Lys	Thr	Cys	His	Leu	Val	Lys	Gly
	690					695					700				
Asp	Phe	Ser	His	Ala	Arg	Gly	Thr	Val	Lys	Asn	Asp	Leu	Cys	Thr	Lys
705					710					715					720
Val	Ser	Thr	Cys	Val	Met	Ala	Glu	Ala	Val	Pro	Lys	Cys	Phe	Ser	Cys
				725					730					735	
Tyr	Ile	Glu	Ala	Ala	Val	Ile	Pro	Ala	Gly	Ala	Arg	Arg	Ile	Arg	Val
			740					745					750		
Val	Glu	Asp	Lys	Pro	Ala	His	Ser	Phe	Leu	Ala	Leu	Lys	Asp	Ser	Gly
		755					760					765			
Lys	Gly	Ser	Ile	Asn	Ser	Asp	Trp	Lys	Ile	Glu	Leu	Pro	Gly	Glu	Phe
	770					775					780				
Gln	Ile	Ala	Gly	Thr	Thr	Val	Arg	Tyr	Val	Arg	Arg	Gly	Leu	Trp	Glu
785					790					795					800
Lys	Ile	Ser	Ala	Lys	Gly	Pro	Thr	Lys	Leu	Pro	Leu	His	Leu	Met	Val
				805					810					815	
Leu	Leu	Phe	His	Asp	Gln	Asp	Tyr	Gly	Ile	His	Tyr	Glu	Tyr	Thr	Val
			820					825					830		
Pro	Val	Asn	Arg	Thr	Ala	Glu	Asn	Gln	Ser	Glu	Pro	Glu	Lys	Pro	Gln
		835					840					845			
Asp	Ser	Leu	Phe	Ile	Trp	Thr	His	Ser	Gly	Trp	Glu	Gly	Cys	Ser	Val
	850					855					860				
Gln	Cys	Gly	Gly	Gly	Glu	Arg	Arg	Thr	Ile	Val	Ser	Cys	Thr	Arg	Ile
865					870					875					880
Val	Asn	Lys	Thr	Thr	Thr	Leu	Val	Asn	Asp	Ser	Asp	Cys	Pro	Gln	Ala
				885					890					895	
Ser	Arg	Pro	Glu	Pro	Gln	Val	Arg	Arg	Cys	Asn	Leu	His	Pro	Cys	Gln
			900					905					910		
Ser	Arg	Trp	Val	Ala	Gly	Pro	Trp	Ser	Pro	Cys	Ser	Ala	Thr	Cys	Glu
		915					920					925			
Lys	Gly	Phe	Gln	His	Arg	Glu	Val	Thr	Cys	Val	Tyr	Gln	Leu	Gln	Asn
	930					935					940				
Gly	Thr	His	Val	Ala	Thr	Arg	Pro	Leu	Tyr	Cys	Pro	Gly	Pro	Arg	Pro
945					950					955					960
Ala	Ala	Val	Gln	Ser	Cys	Glu	Gly	Gln	Asp	Cys	Leu	Ser	Ile	Trp	Glu
				965					970					975	

WO 03/004607

PCT/US02/21056

Ala Ser Glu Trp Ser Gln Cys Ser Ala Ser Cys Gly Lys Gly Val Trp
 980 985 990

Lys Arg Thr Val Ala Cys Thr Asn Ser Gln Gly Lys Cys Asp Ala Ser
 995 1000 1005

Thr Arg Pro Arg Ala Glu Glu Ala Cys Glu Asp Tyr Ser Gly Cys Tyr
 1010 1015 1020

Glu Trp Lys Thr Gly Asp Trp Ser Thr Cys Ser Ser Thr Cys Gly Lys
 1025 1030 1035 1040

Gly Leu Gln Ser Arg Val Val Gln Cys Met His Lys Val Thr Gly Arg
 1045 1050 1055

His Gly Ser Glu Cys Pro Ala Leu Ser Lys Pro Ala Pro Tyr Arg Gln
 1060 1065 1070

Cys Tyr Gln Glu Val Cys Asn Asp Arg Ile Asn Ala Asn Thr Ile Thr
 1075 1080 1085

Ser Pro Arg Leu Ala Ala Leu Thr Tyr Lys Cys Thr Arg Asp Gln Trp
 1090 1095 1100

Thr Val Tyr Cys Arg Val Ile Arg Glu Lys Asn Leu Cys Gln Asp Met
 1105 1110 1115 1120

Arg Trp Tyr Gln Arg Cys Cys Gln Thr Cys Arg Asp Phe Tyr Ala Asn
 1125 1130 1135

Lys Met Arg Gln Pro Pro Pro Ser Ser
 1140 1145

<210> 21
 <211> 22
 <212> DNA
 <213> Homo sapiens

<400> 21
 ccggctccct cgtctcgctc ag 22

<210> 22
 <211> 25
 <212> DNA
 <213> Homo sapiens

<400> 22
 agcagaaggg ctgggggtca aggac 25

<210> 23
 <211> 24
 <212> DNA
 <213> Homo sapiens

<400> 23
 acgtgactgg caggggtgca agtt 24

<210> 24

WO 03/004607

PCT/US02/21056

<211> 24
 <212> DNA
 <213> Homo sapiens

<400> 24
 cggagcatgt ggacggagac tgga 24

<210> 25
 <211> 24
 <212> DNA
 <213> Homo sapiens

<400> 25
 tctggctctc aaagactcgg gtaa 24

<210> 26
 <211> 23
 <212> DNA
 <213> Homo sapiens

<400> 26
 gcaggcaciaa ctgttcgcta tgt 23

<210> 27
 <211> 19
 <212> DNA
 <213> Homo sapiens

<400> 27
 tcacgagctc ggcggtggc 19

<210> 28
 <211> 23
 <212> DNA
 <213> Homo sapiens

<400> 28
 tcggccacca ccagggtctc cac 23

<210> 29
 <211> 23
 <212> DNA
 <213> Homo sapiens

<400> 29
 gttcctccgc tcccgccagt ccc 23

<210> 30
 <211> 22
 <212> DNA
 <213> Homo sapiens

<400> 30
 ggtcccgggt accatgtgtg ac 22

<210> 31
 <211> 70
 <212> DNA
 <213> Homo sapiens

<400> 31
ctagagccgc caccatgtgt gacggcgccc tgctgcctcc gctcgtcctg cccgtgctgc 60
tgctgctggt 70

<210> 32
<211> 74
<212> DNA
<213> Homo sapiens

<400> 32
gtccccaaac cagcagcagc agcacgggca ggacgagcgg aggcagcagg gcgccgtcac 60
acatggtggc ggct 74

<210> 33
<211> 86
<212> DNA
<213> Homo sapiens

<400> 33
ttggggactg gacccgggca cagctgtcgg cgacgcggcg gccgacgtgg aggtggtgct 60
cccgtggcgg gtgcgccccg acgacg 86

<210> 34
<211> 82
<212> DNA
<213> Homo sapiens

<400> 34
tgcacgtcgt cggggcgcac ccgccacggg agcaccacct ccacgtcggc cgccgcgtcg 60
ccgacagctg tgcccgggtc ca 82

PCT/US02/21056, 20012003

1. An isolated DNA molecule comprising a DNA sequence chosen from:
 - a) the sequence of SEQ ID NO. 5 from nucleotide #1-#2270;
 - b) the sequence of SEQ ID NO. 7 from nucleotide #1-#2339;
 - c) the sequence of SEQ ID NO. 3 from nucleotide #1 to #3899;
 - d) the sequence of SEQ ID NO. 9 from nucleotide #1 to #5001;
 - e) the sequence of SEQ. ID NO. 11 from nucleotide #1 to #3369; and
 - f) naturally occurring human allelic sequences and equivalent degenerative codon sequences of (a) through (e).
2. A vector comprising a DNA molecule of claim 1 in operative association with an expression control sequence therefor.
3. A host cell transformed with the DNA sequence of claim 1.
4. A host cell transformed with a DNA sequence of claim 2.
5. A method for producing a purified human aggrecanase protein, said method comprising:
 - a) culturing a host cell transformed with a DNA molecule according to claim 1; and
 - b) recovering and purifying said aggrecanase protein from the culture medium.
6. The method of claim 5, wherein said host cell is an insect cell.
7. A purified aggrecanase protein comprising an amino acid sequence chosen from:
 - a) the amino acid sequence set forth in SEQ ID NO. 4 from amino acid #1-#807;
 - b) the amino acid sequence set forth in SEQ ID NO. 6 from amino acid #1-#756;

- c) the amino acid sequence set forth in SEQ ID NO. 8 from amino acid #1-#779;
 - d) the amino acid sequence set forth in SEQ ID NO. 10 from amino acid #1-#1057;
 - e) the amino acid sequence set forth in SEQ ID NO. 13 from amino acid #1-#1122; and
 - f) homologous aggrecanase proteins consisting of addition, substitution, and deletion mutants of the sequences of (a) through (e).
8. A purified aggrecanase protein produced by the steps of
- a) culturing a cell transformed with a DNA molecule according to claim 1; and
 - b) recovering and purifying from said culture medium a protein comprising an amino acid sequence chosen from SEQ. ID NO. 4, 6, 8, 10, and 13.
9. An antibody that binds to a purified aggrecanase protein of claim 7.
10. The antibody of claim 9, wherein the antibody inhibits aggrecanase activity.
11. A method for identifying inhibitors of aggrecanase comprising
- a) providing an aggrecanase protein chosen from:
 - i) SEQ ID NO. 4 or a fragment thereof;
 - ii) SEQ ID NO. 6 or a fragment thereof;
 - iii) SEQ ID NO. 8 or a fragment thereof;
 - iv) SEQ. ID NO. 10 or a fragment thereof; and
 - v) SEQ. ID NO. 13 or a fragment thereof;
 - b) combining the aggrecanase with a potential inhibitor and

- c) evaluating whether the potential inhibitor inhibits aggrecanase activity.
12. The method of claim 11 wherein the method comprises evaluating the aggrecanase protein is used in a three dimensional structural analysis prior to combining with the potential inhibitor.
13. The method of claim 11 wherein the method comprises evaluating the aggrecanase protein is used in a computer aided drug design prior to combining with the potential inhibitor.
14. A pharmaceutical composition for inhibiting the proteolytic activity of aggrecanase, wherein the composition comprises an antibody according to claim 9 and a pharmaceutical carrier.
15. A method for inhibiting aggrecanase in a mammal comprising administering to said mammal an effective amount of the composition of claim 14 and allowing the composition to inhibit aggrecanase activity.
16. The method of claim 15, wherein the composition is administered intravenously, subcutaneously, or intramuscularly.
17. The method of claim 15, wherein the composition is administered at a dosage of from 500 $\mu\text{g/kg}$ to 1 mg/kg .

1	CGCACGCCCC	CAGCCGCCCC	GCGCGCCCGG	CCCGGAGAGC	GCGCCCTGCT
51	GCTGCACCTG	CCGGCCTTCG	GGCGCGACCT	GTACCTTCAG	CTGCGCCGCG
101	ACCTGCGCTT	CCTGTCCCGA	GGCTTCGAGG	TGGAGGAGGC	GGGCGCGGCC
151	CGGCGCCGCG	GCCGCCCCGC	CGAGCTGTGC	TTCTACTCGG	GCCGTGTGCT
201	CGGCCACCCC	GGCTCCCTCG	TCTCGCTCAG	CGCCTGCGGC	GCCGCCGGCG
251	GCCTGGTTGG	CCTCATTCAG	CTTGGGCAGG	AGCAGGTGCT	AATCCAGCCC
301	CTCAACAAC	CCCAGGGCCC	ATTGAGTGA	CGAGAACATC	TGATCAGGCG
351	CAAATGGTCC	TTGACCCCCA	GCCCTTCTGC	TGAGGCCAG	AGACCTGAGC
401	AGCTCTGCAA	GGTTCTAACA	GAAAAGAAGA	AGCCGACGTG	GGGCAGGCCT
451	TCGCGGGACT	GGCGGGAGCG	GAGGAACGCT	ATCCGGCTCA	CCAGCGAGCA
501	CACGGTGGAG	ACCCTGGTGG	TGGCCGACGC	CGACATGGTG	CAGTACCACG
551	GGGCCGAGGC	CGCCCAGAGG	TTCATCCTGA	CCGTCATGAA	CATGGTATAC
601	AATATGTTTC	AGCACCAGAG	CCTGGGGATT	AAAATTAACA	TTCAAGTGAC
651	CAAGCTTGTC	CTGCTACGAC	AACGTCCCGC	TAAGTTGTCC	ATTGGGCACC
701	ATGGTGAGCG	GTCCCTGGAG	AGCTTCTGTC	ACTGGCAGAA	CGAGGAGTAT
751	GGAGGAGCGC	GATACCTCGG	CAATAACCAG	GTTCCCGGCG	GGAAGGACGA
801	CCCGCCCCTG	GTGGATGCTG	CTGTGTTTGT	GACCAGGACA	GATTTCTGTG
851	TACACAAAGA	TGAACCGTGT	GACACTGTTG	GAATTGCTTA	CTTAGGAGGT
901	GTGTGCAGTG	CTAAGAGGAA	GTGTGTGCTT	GCCGAAGACA	ATGGTCTCAA
951	TTTGGCCTTT	ACCATCGCCC	ATGAGCTGGG	CCACAACCTG	GGCATGAACC
1001	ACGACGATGA	CCACTCATCT	TGCGCTGGCA	GGTCCCACAT	CATGTCAGGA
1051	GAGTGGGTGA	AAGGCCGGAA	CCCAAGTGAC	CTCTCTTGGT	CCTCCTGCAG
1101	CCGAGATGAC	CTTGAAAAC	TCCTCAAGTC	AAAAGTCAGC	ACCTGCTTGC
1151	TAGTCACGGA	CCCCAGAAGC	CAGCACACAG	TACGCCTCCC	GCACAAGCTG
1201	CCGGGCATGC	ACTACAGTGC	CAACGAGCAG	TGCCAGATCC	TGTTTGGCAT
1251	GAATGCCACC	TTCTGCAGAA	ACATGGAGCA	TCTAATGTGT	GCTGGACTGT
1301	GGTGCTTGGT	AGAAGGAGAC	ACATCCTGCA	AGACCAAGCT	GGACCCTCCC
1351	CTGGATGGCA	CCGAGTGTGG	GGCAGACAAG	TGGTGCCGCG	CGGGGGAGTG
1401	CGTGAGCAAG	ACGCCCATCC	CGGAGCATGT	GGACGGAGAC	TGGAGCCCGT
1451	GGGGCGCCTG	GAGCATGTGC	AGCCGAACAT	GTGGGACGGG	AGCCCGCTTC
1501	CGGCAGAGGA	AATGTGACAA	CCCCCCCCCT	GGGCCTGGAG	GCACACACTG
1551	CCCGGGTGCC	AGTGTAGAAC	ATGCGGTCTG	CGAGAACCTG	CCCTGCCCCA
1601	AGGGTCTGCC	CAGCTTCCGG	GACCAGCAGT	GCCAGGCACA	CGACCGGCTG
1651	AGCCCCAAGA	AGAAAGGCCT	GCTGACAGCC	GTGGTGGTTG	ACGATAAGCC
1701	ATGTGAACTC	TACTGCTCGC	CCCTCGGGAA	GGAGTCCCCA	CTGCTGGTGG
1751	CCGACAGGGT	CCTGGACGGT	ACACCCTGCG	GGCCCTACGA	GACTGATCTC

FIG. 1

1801	TGCGTGCACG	GCAAGTGCCA	GAAAATCGGC	TGTGACGGCA	TCATCGGGTC
1851	TGCAGCCAAA	GAGGACAGAT	GCGGGGTCTG	CAGCGGGGAC	GGCAAGACCT
1901	GCCACTTGGT	GAAGGGCGAC	TTCAGCCACG	CCCGGGGGAC	AGGTTATATC
1951	GAAGCTGCCG	TCATTCCTGC	TGGAGCTCGG	AGGATCCGTG	TGGTGGAGGA
2001	TAAACCTGCC	CACAGCTTTC	TGGCTCTCAA	AGACTCGGGT	AAGGGGTCCA
2051	TCAACAGTGA	CTGGAAGATA	GAGCTCCCCG	GAGAGTTCCA	GATTGCAGGC
2101	ACAAC TGTTT	GCTATGTGAG	AAGGGGGCTG	TGGGAGAAGA	TCTCTGCCAA
2151	GGGACCAACC	AACTACCGC	TGCACTTGAT	GGTGTGTTA	TTTCACGACC
2201	AAGATTATGG	AATTCATTAT	GAATACACTG	TTCCTGTAAA	CCGCACTGCG
2251	GAAAATCAAA	GCGAACCAGA	AAAACCGCAG	GACTCTTTGT	TCATCTGGAC
2301	CCACAGCGGC	TGGGAAGGGT	GCAGTGTGCA	GTGCGGCGGA	GGGGAGCGCA
2351	GAACCATCGT	CTCGTGTACA	CGGATTGTCA	ACAAGACCAC	AACTCTGGTG
2401	AACGACAGTG	ACTGCCCTCA	AGCAAGCCGC	CCAGAGCCCC	AGGTCCGAAG
2451	GTGCAACTTG	CACCCCTGCC	AGTCACGKTG	GGTGGCAGGC	CCGTGGAGCC
2501	CCTGCTCGGC	GACCTGTGAG	AAAGGCTTCC	AGCACCGGGA	GGTGACCTGC
2551	GTGTACCAGC	TGCAGAACGG	CACACACGTC	GCTACGCGGC	CCCTCTACTG
2601	CCCGGGCCCC	CGGCCGGCGG	CAGTGCAGAG	CTGTGAAGGC	CAGGACTGCC
2651	TGTCCATCTG	GGAGGCGTCT	GAGTGGTCAC	AGTGCTCTGC	CAGCTGTGGT
2701	AAAGGGGTGT	GGAAACGGAC	CGTGGCGTGC	ACCAACTCAC	AAGGGAAATG
2751	CGACGCATCC	ACGAGGCCGA	GAGCCGAGGA	GGCCTGCGAG	GACTACTCAG
2801	GCTGCTACGA	GTGGAAAAC	GGGGACTGGT	CTACGTGCTC	GTCGACCTGC
2851	GGGAAGGGCC	TGCAGTCCCG	GGTGGTGCAG	TGCATGCACA	AGGTCACAGG
2901	GCGCCACGGC	AGCGAGTGCC	CCGCCCTCTC	GAAGCCTGCC	CCCTACAGAC
2951	AGTGCTACCA	GGAGGTCTGC	AACGACAGGA	TCAACGCCAA	CACCATCACC
3001	TCCCCCGGCC	TTGCTGCTCT	GACCTACAAA	TGCACACGAG	ACCAGTGGAC
3051	GGTATATTGC	CGGGTCATCC	GAGAAAAGAA	CCTCTGCCAG	GACATGCGGT
3101	GGTACCAGCG	CTGCTGCCAG	ACCTGCAGGG	ACTTCTATGC	AAACAAGATG
3151	CGCCAGCCAC	CGCCGAGCTC	<u>GTGAC</u> ACGCA	GTCCCAAGGG	TCGCTCAAAG
3201	CTCAGACTCA	GGTCTGAAAG	CCACCCACCC	GCAAGCCTAC	CAGCCTTGTG
3251	GCCACACCCC	CACCCGGCTG	CCACAAGAAT	CCAAC TGCA	AGAACATGAG
3301	CGTGGACTTG	GCGTTTGCCA	TTAGTGCTTC	CGTACTTAAT	ATATTGTTAA
3351	CAGCCACTGG	CTCACTTTCT	ACAGTGAGGA	GAAAGTAGGC	ATGAGTCACA
3401	AAGTAACTTC	AATTTCTAGG	ATTTCAGGTA	CCTCGAAGGG	AAGCACCTCT
3451	GGCAGACAAC	CGTCAAGAGA	GAGACATCAT	TTAGTGTTCC	TGTCTTGACT
3501	CGCTTTTGAC	ATTTGAATTT	CCAGTGCTTG	GTATATCATG	GAGGAAACAT
3551	CCCCAAAACG	AGACATGCTA	GAAAAGGCTT	TATTCTAAAG	GCTTTATTCT

FIG. 1 (CONT. -1)

3601	GAAAGCCGGC	GACACCCTGG	AGGGAGGGGC	AGGTGTTGGT	GAGCCTCTGC
3651	CCGTGGCTTC	TCTGGGGAGG	GCCGGGCTGC	TTAGCCCACG	TTTCTCTTCA
3701	TCTACCTTCT	TGACCACATG	AGAACCAGGA	CATTGCCTCC	ATGCCCCTCT
3751	CTGACAACAT	AGTCTCTAAA	TCCTAGGTGT	TGCCTTGGA	GTCTCGTGCG
3801	TGGAGTGTA	ATCTATATAT	GCCAGCGAGG	ACAGCAGTGC	CACGCAGTTC
3851	ATACCACCCG	CATGGGAAGA	ATGTTCCAAG	AGAGTCTGGG	TTTGGGGAAG
3901	CATCTAATTT	TCAGAGCTCT	GCTGTCCACC	GTGTAGGGAA	ACAGAAGGGC
3951	CTCTCTTCAA	GGTGCTGTGA	CATAAGAAAC	GGTAATTGCG	GTGATGGGGT
4001	TGCTTCCTAA	GGCAAAGGTA	AGCTTGGGCC	AGCTTCACTG	GGGCGGATGG
4051	GCACCTGCCC	CGCCTTCCGC	GAGCATCCAC	TCTGGCCCCG	ACTTCCTAAA
4101	GCTTTGTACC	TTAGAGATGC	TGTACCACAT	CCCAGTGGCT	TTCTACCGAC
4151	CGTGGCCATT	TATCTGAAGG	TAAGACGACA	TTTGGGACCT	CTGAGGACAC
4201	AGGCCTAGGA	TCTGTAGAGC	AAGGCCTGAC	TGCTCTATCC	TGGCACGGAG
4251	CAGCCTGATA	TGCCGGGACC	AGGGGAGGAA	CGCCATCTGG	CTGGCACTGC
4301	TGCACACCCG	CCGAGCCTTC	CTGTAGCCCC	AGACTTTGTG	GTACCCATTA
4351	TCATCACGCC	TGTCATCATT	GACCCATCTT	CTTGGTGGGG	CAAGGATGAT
4401	GCATGATGAA	GGTCCTTCCC	TCCTGCAGCC	CCCTTACGCC	TGGCAGCAGA
4451	CAAGCAGAGT	GGCCTCGTTG	AGAGCACAGA	GGATGGTAGC	ACCCTACCTG
4501	CAAGGAGGCC	GGGCAGGGAC	CCTAGATGCC	AGGAGGCCTG	TTTTGCTCAC
4551	CAACTTGGTG	GGCATTTCAT	GGGTGCTTAT	GTTCTAGGAC	TTTACCGTAA
4601	ATAACACCTC	CTCCCTGATT	TCAGGCAGAA	GGTCTCACTT	GGACTTCCAT
4651	GGGATCATCT	CCCTGTGTTT	CTTGATTTAT	TGGTGCTGTG	TTTCTGTGTT
4701	TTGTTTTGTT	ACATGTCACA	ACCGTAGAGT	TAGCTTAAAT	CAGAAAGAAG
4751	CCTCTCTGCC	TTCTCCACCC	TGTCTTACGA	GCTGTGTTTT	TGTTTTTACT
4801	ACCCTAGAGG	CAGAGAAGCG	GTAGGGATGT	CAGGGAATTT	ACTCACTTCC
4851	ACTTGAATCA	ACGAGAAGTG	TTGAGAAACT	TCCGTGGGTG	CTCTGTGGAA
4901	AGAACCGAGG	GTGTCAGGAT	GGAGCGGCCC	ACCCTCGCCC	CGCGGCCTGC
4951	GCAGACTGCT	GTCCTCCCCT	TCAGGCCTGG	CCACCAGCAG	ACTCCCATGA
5001	ATTC				

FIG. 1 (CONT. -2)

1	RTPPAAPRAR	PGERALLLHL	PAFGRDLYLQ	LRRDLRFLSR	GFEVEEAGAA
51	RRRGRPAELC	FYSGRVLGHP	GSLVSLSACG	AAGGLVGLIQ	LGQEQVLIQP
101	LNNSQGPFSG	REHLIRRKWS	LTPSPSAEAQ	RPEQLCKVLT	EKKKPTWGRP
151	SRDWRERRNA	IRLTSEHTVE	TLVVADADMV	QYHGAEAAQR	FILTVMMNVY
201	NMFQHQSLGI	KINIQVTKLV	LLRQRPAKLS	IGHHGERSLE	SFCHWQNEEY
251	GGARYLGNNQ	VPGGKDDPPL	VDAAVFVTRT	DFCVHKDEPC	DTVGIAYLGG
301	VCSAKRKCVL	AEDNGLNLAF	TIAHELGHNL	GMNHDDDHSS	CAGRSHIMSG
351	EWVKGRNPSD	LSWSSCSRDD	LENFLKSKVS	TCLLVTDPRS	QHTVRLPHKL
401	PGMHYSANEQ	CQILFGMNAT	FCRNMEHLMC	AGLWCLVEGD	DSCKTKLDPP
451	LDGTECGADK	WCORAGECVSK	TPIPEHVDGD	WSPWGAWSMC	SRTCCTGARF
501	RQRKCDNPPP	GPGGTHCPGA	SVEHAVCENL	PCPKGLPSFR	DQQCQAHDR
551	SPKKKGLLTA	VVDDKPCEL	YCSPLGKESP	LLVADRVLDG	TPCGPYETDL
601	CVHGKCQKIG	CDGIIGSAAK	EDRCGVCSGD	GKTCHLVKGD	FSHARGTGYI
651	EAAVIPAGAR	RIRVVEDKPA	HSFLALKDSG	KGSINSDWKI	ELPGEFQIAG
701	TTVRYVRRGL	WEKISAKGPT	KLPLHLMVLL	FHDQDYGIHY	EYTVPVNRTA
751	ENQSEPEKPQ	DSLFIWTHSG	WEGCSVQCGG	GERRTIVSCT	RIVNKTTLV
801	NDSDCPQASR	PEPQVRRCNL	HPCQSRWVAG	PWSPCSATCE	KGFQHREVT
851	VYQLQNGTHV	ATRPLYCPGP	RPAAVQSCEG	QDCLSIWEAS	EWSQCSASCG
901	KGVWKRTVAC	TNSQGKCDAS	TRPRAEEACE	DYSGCYEWKT	GDWSTCSSTC
951	GKGLQSRVVQ	CMHKVTGRHG	SECPALSKPA	PYRQCYQEV	NDRINANTIT
1001	SPRLAALTYK	CTRDQWTVYC	RVIREKNLCQ	DMRWYQRCCQ	TCRDFYANKM
1051	RQPPPPSS*				

FIG. 2

1	ATGTGTGACG	GCGCCCTGCT	GCCTCCGCTC	GTCCTGCCCCG	TGCTGCTGCT
51	GCTGGTTTGG	GGACTGGACC	CGGGCACAGC	TGTCGGGCGAC	GCGGCGGCCG
101	ACGTGGAGGT	GGTGCTCCCG	TGGCGGGTGC	GCCCCGACGA	CGTGACACCTG
151	CCGCCGCTGC	CCGCAGCCCC	CGGGCCCCGA	CGGCGGCGAC	GCCCCCGCAC
201	GCCCCCAGCC	GCCCCGCGCG	CCCGGCCCGG	AGAGCGCGCC	CTGCTGCTGC
251	ACCTGCCGGC	CTTCGGGCGC	GACCTGTACC	TTCAGCTGCG	CCGCGACCTG
301	CGCTTCCTGT	CCCGAGGCTT	CGAGGTGGAG	GAGGCGGGCG	CGGCCCGGCG
351	CCGCGGCCGC	CCCGCCGAGC	TGTGCTTCTA	CTCGGGCCGT	GTGCTCGGCC
401	ACCCCGGCTC	CCTCGTCTCG	CTCAGCGCCT	GCGGCGCCGC	CGGCGGCCTG
451	GTTGGCCTCA	TTCAGCTTGG	GCAGGAGCAG	GTGCTAATCC	AGCCCCTCAA
501	CAACTCCCAG	GGCCCATTCA	GTGGACGAGA	ACATCTGATC	AGGCGCAAAT
551	GGTCCTTGAC	CCCCAGCCCT	TCTGCTGAGG	CCCAGAGACC	TGAGCAGCTC
601	TGCAAGGTTT	TAACAGAAAA	GAAGAAGCCG	ACGTGGGGCA	GGCCTTCGCG
651	GGACTGGCGG	GAGCGGAGGA	ACGCTATCCG	GCTCACCAGC	GAGCACACGG
701	TGGAGACCCT	GGTGGTGGCC	GACGCCGACA	TGGTGCAGTA	CCACGGGGCC
751	GAGGCCGCCC	AGAGGTTCAT	CCTGACCGTC	ATGAACATGG	TATACAATAT
801	GTTTCAGCAC	CAGAGCCTGG	GGATTAAAAT	TAACATTCAA	GTGACCAAGC
851	TTGTCCTGCT	ACGACAACGT	CCCGCTAAGT	TGTCCATTGG	GCACCATGGT
901	GAGCGGTCCC	TGGAGAGCTT	CTGTCACTGG	CAGAACGAGG	AGTATGGAGG
951	AGCGCGATAC	CTCGGCAATA	ACCAGGTTC	CGGCGGGAAG	GACGACCCGC
1001	CCCTGGTGG	TGCTGCTGTG	TTTGTGACCA	GGACAGATTT	CTGTGTACAC
1051	AAAGATGAAC	CGTGTGACAC	TGTTGGAATT	GCTTACTTAG	GAGGTGTGTG
1101	CAGTGCTAAG	AGGAAGTGTG	TGCTTGCCGA	AGACAATGGT	CTCAATTTGG
1151	CCTTTACCAT	CGCCCATGAG	CTGGGCCACA	ACTTGGGCAT	GAACCACGAC
1201	GATGACCACT	CATCTTGCGC	TGGCAGGTCC	CACATCATGT	CAGGAGAGTG
1251	GGTGAAAGGC	CGGAACCCAA	GTGACCTCTC	TTGGTCCTCC	TGCAGCCGAG
1301	ATGACCTTGA	AAACTTCCTC	AAGTCAAAAG	TCAGCACCTG	CTTGCTAGTC
1351	ACGGACCCCA	GAAGCCAGCA	CACAGTACGC	CTCCCGCACA	AGCTGCCGGG
1401	CATGCACTAC	AGTGCCAACG	AGCAGTGCCA	GATCCTGTTT	GGCATGAATG
1451	CCACCTTCTG	CAGAAACATG	GAGCATCTAA	TGTGTGCTGG	ACTGTGGTGC
1501	CTGGTAGAAG	GAGACACATC	CTGCAAGACC	AAGCTGGACC	CTCCCCTGGA
1551	TGGCACCGAG	TGTGGGGGAG	ACAAGTGGTG	CCGCGCGGGG	GAGTGCGTGA
1601	GCAAGACGCC	CATCCCGGAG	CATGTGGACG	GAGACTGGAG	CCCGTGGGGC
1651	GCCTGGAGCA	TGTGCAGCCG	AACATGTGGG	ACGGGAGCCC	GCTTCCGGCA
1701	GAGGAAATGT	GACAACCCCC	CCCCTGGGCC	TGGAGGCACA	CACTGCCCGG

FIG. 3

1751	GTGCCAGTGT	AGAACATGCG	GTCTGCGAGA	ACCTGCCCTG	CCCCAAGGGT
1801	CTGCCCAGCT	TCCGGGACCA	GCAGTGCCAG	GCACACGACC	GGCTGAGCCC
1851	CAAGAAGAAA	GGCCTGCTGA	CAGCCGTGGT	GGTTGACGAT	AAGCCATGTG
1901	AACTCTACTG	CTCGCCCCTC	GGGAAGGAGT	CCCCACTGCT	GGTGGCCGAC
1951	AGGGTCCTGG	ACGGTACACC	CTGCGGGCCC	TACGAGACTG	ATCTCTGCGT
2001	GCACGGCAAG	TGCCAGAAAA	TCGGCTGTGA	CGGCATCATC	GGGTCTGCAG
2051	CCAAAGAGGA	CAGATGCGGG	GTCTGCAGCG	GGGACGGCAA	GACCTGCCAC
2101	TTGGTGAAGG	GCGACTTCAG	CCACGCCCGG	GGGACAGGTT	ATATCGAAGC
2151	TGCCGTCATT	CCTGCTGGAG	CTCGGAGGAT	CCGTGTGGTG	GAGGATAAAC
2201	CTGCCCACAG	CTTTCTGGCT	CTCAAAGACT	CGGGTAAGGG	GTCCATCAAC
2251	AGTGACTGGA	AGATAGAGCT	CCCCGGAGAG	TTCCAGATTG	CAGGCACAAC
2301	TGTTCGCTAT	GTGAGAAGGG	GGCTGTGGGA	GAAGATCTCT	GCCAAGGGAC
2351	CAACCAAAC	ACCGCTGCAC	TTGATGGTGT	TGTTATTTCA	CGACCAAGAT
2401	TATGGAATTC	ATTATGAATA	CACTGTTTCT	GTAAACCGCA	CTGCGGAAAA
2451	TCAAAGCGAA	CCAGAAAAAC	CGCAGGACTC	TTTGTTTCATC	TGGACCCACA
2501	GCGGCTGGGA	AGGGTGCAGT	GTGCAGTGCG	GCGGAGGGGA	GCGCAGAACC
2551	ATCGTCTCGT	GTACACGGAT	TGTCAACAAG	ACCACAAC	TGGTGAACGA
2601	CAGTGACTGC	CCTCAAGCAA	GCCGCCCAGA	GCCCCAGGTC	CGAAGGTGCA
2651	ACTTGCACCC	CTGCCAGTCA	CGKTGGGTGG	CAGGCCCGTG	GAGCCCCTGC
2701	TCGGCGACCT	GTGAGAAAGG	CTTCCAGCAC	CGGGAGGTGA	CCTGCGTGTA
2751	CCAGCTGCAG	AACGGCACAC	ACGTGCTTAC	GCGGCCCTC	TACTGCCCGG
2801	GCCCCCGGCC	GGCGGCAGTG	CAGAGCTGTG	AAGGCCAGGA	CTGCCTGTCC

FIG. 3 (CONT. -1)

2851 ATCTGGGAGG CGTCTGAGTG GTCACAGTGC TCTGCCAGCT GTGGTAAAGG
 2901 GGTGTGGAAA CGGACCGTGG CGTGCACCAA CTCACAAGGG AAATGCGACG
 2951 CATCCACGAG GCCGAGAGCC GAGGAGGCCT GCGAGGACTA CTCAGGCTGC
 3001 TACGAGTGGA AAAC TGGGGA CTGGTCTACG TGCTCGTCGA CCTGCGGGAA
 3051 GGGCCTGCAG TCCCGGGTGG TGCAGTGCAT GCACAAGGTC ACAGGGCGCC
 3101 ACGGCAGCGA GTGCCCCGCC CTCTCGAAGC CTGCCCCCTA CAGACAGTGC
 3151 TACCAGGAGG TCTGCAACGA CAGGATCAAC GCCAACACCA TCACCTCCCC
 3201 CCGCCTTGCT GCTCTGACCT ACAAATGCAC ACGAGACCAG TGGACGGTAT
 3251 ATTGCCGGGT CATCCGAGAA AAGAACCTCT GCCAGGACAT GCGGTGGTAC
 3301 CAGCGCTGCT GCCAGACCTG CAGGGACTTC TATGCAAACA AGATGCGCCA
 3351 GCCACCGCCG AGCTCGTGA

FIG. 3 (CONT. -2)

2101 TTGGTGAAGG GCGACTTCAG CCACGCCCGG GGGACAGTTA AGAATGATCT
 2151 CTGTACGAAG GTATCCACAT GTGTGATGGC AGAGGCTGTT CCCAAGTGTT
 2201 TCTCATGTTA TATCGAAGCT GCCGTCATTC CTGCTGGAGC TCGGAGGATC
 2251 CGTGTGGTGG AGGATAAACC TGCCACACAGC TTTCTGGCTC TCAAAGACTC

FIG. 4

1 MCDGALLPPL VLPVLLLLLVW GLDPGTAVGD AAADVEVVLP WRVRPDDVHL
 51 PPLPAAPGPR RRRRPRTPPA APRARPGERA LLLHLPAFGR DLYLQLRRDL
 101 RFLSRGFEVE EAGAARRRGR PAELCFYSGR VLGHPSGLVS LSACGAAGGL
 151 VGLIQLGQEQ VLIQPLNNSQ GPFSGREHLI RRKWSLTPSP SAEAQRPEQL
 201 CKVLTEKKKP TWGRPSRDWR ERRNAIRLTS EHTVETLVVA DADMVQYHGA
 251 EAAQRFILTV MNMVYNMFQH QSLGIKINIQ VTKLVLLRQR PAKLSIGHHG
 301 ERSLESFCHW QNEEYGGARY LGNNQVPGGK DDPPLVDAAV FVTRTDFCVH
 351 KDEPCDTVGI AYLGGVCSAK RKCFLAEDNG LNLAF TIAHE LGHNLGMNHD
 401 DDHSSCAGRS HIMSGE WVKG RNPSDLSWSS CSRDDLENFL KSKVSTCLLV
 451 TDPRSQHTVR LPHKLPGMHY SANEQCQILF GMNATFCRNM EHLMCAGLWC
 501 LVEGDT SCKT KLDPPLDGTE CGADKWCRAG ECVSKTPIPE HVDGDWSPWG
 551 AWSMCSRTC G TGARFRQRKC DNPPPGPGGT HCPGASVEHA VCENLPCPKG
 601 LPSFRDQOCQ AHDRLSPKKK GLLTAVVVDD KPCELYCSPL GKESPLLVD
 651 RVLDGT PCGP YETDLCVHGK CQKIGCDGII GSAAKEDRCG VCSGDGKTCH

FIG. 5

701 LVKGDFSHAR GTGYIEAAVI PAGARRIRVV EDKPAHSFLA LKDSGKGSIN
 751 SDWKIELPGE FQIAGTTVRY VRRGLWEKIS AKGPTKLPLH LMVLLFHDQD
 801 YGIHYEYTVV VNRTAENQSE PEKPQDSLFI WTHSGWEGCS VQCGGGERRT
 851 IVSCTRIVNK TTTLVNDSDC PQASRPEPQV RRCNLHPCQS RWVAGPWSPC
 901 SATCEKGFQH REVTCVYQLQ NGTHVATRPL YCPGPRPAAV QSCEGQDCLS
 951 IWEASEWSQC SASCGKGVWK RTVACTNSQG KCDASTRPRA EEACEDYSGC
 1001 YEWKTGDWST CSSTCGKGLQ SRVVQCMHKV TGRHGSECPA LSKPAPYRQC
 1051 YQEV CNDRIN ANTITSPRLA ALTYKCTRDQ WTVYCRVIRE KNLCQDMRWY
 1101 QRCCQTCRDF YANKMRQPPP SS

FIG. 5 (CONT.- 1)

LPSFRDQQCQ AHDRLSPKKK GLLTAVVVDD KPCELYCSPL GKESPLLVD
 RVLDTGTPCGP YETDL CVHGK CQKIGCDGII GSAAKEDRCG VCSGDGKTCH
 LVKGDFSHAR GTVKNDLCTK VSTCVMAEAV PKCFSCYIEA AVIPAGARRI
 RVVEDKPAHS FLALKDSGKG SINSWKIEL PGEFQIAGTT VRYVRRGLWE
 KISAKGPTKL PLHLMVLLFH DQDYGIHYEY TVPVNRTAEN QSEPEKPQDS
 LFIWTHSGWE GCSVQ

FIG. 6