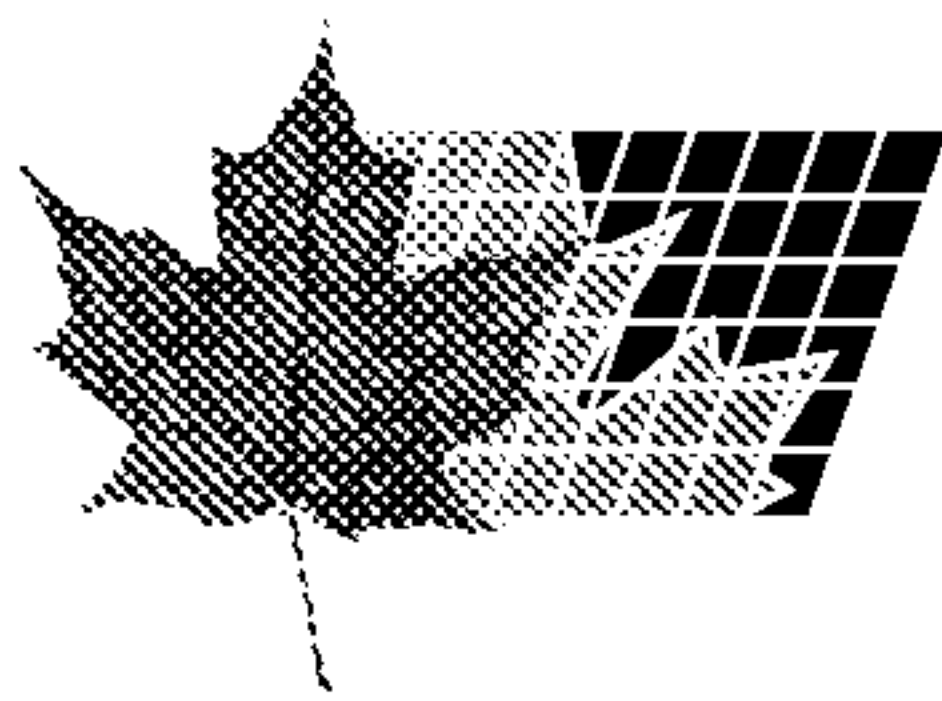




(72) TINDAL, MICHAEL HOWARD, US

(72) HAQI, TARIQ MEHMOOD, US

(71) THE PROCTER & GAMBLE COMPANY, US

(71) CASE WESTERN RESERVE UNIVERSITY, US

(51) Int.Cl.<sup>6</sup> C12N 15/57, C12Q 1/68, C12N 9/64, G01N 33/573, C12Q 1/37

(30) 1997/02/25 (08/810,153) US

(30) 1997/02/28 (PCT/US97/03217) US

(54) **UTILISATION D'UNE NOUVELLE PROTEINE**

**METALLOPROTEASE DISINTEGRINE, MUTANTS,**

**FRAGMENTS ET ANALOGUES DE LADITE PROTEINE**

(54) **USE OF A NOVEL DISINTEGRIN METALLOPROTEASE**

(57) La présente invention concerne un procédé permettant d'identifier des composés susceptibles de se fixer à la protéine disintégrine, et de déterminer la quantité et l'affinité d'un composé susceptible de se fixer à la protéine disintégrine dans un échantillon. L'invention concerne également une cellule hôte contenant un vecteur d'expression recombinant à la protéine disintégrine et un vecteur d'expression recombinant codant la protéine disintégrine, la protéine humaine métalloprotéase disintégrine, son mutant et son fragment. Ladite cellule est utilisée pour répondre aux objectifs de la présente invention. L'invention concerne enfin un procédé in vivo ou in vitro, qui permet de dépister l'ostéo-arthrite dégénérative et d'autres maladies provoquées par la métalloprotéase, et est susceptible d'être fabriqué et utilisé sous la forme d'un kit.

(57) This invention provides a method for identifying compounds capable of binding to the disintegrin protein, and determining the amount and affinity of a compound capable of binding to the disintegrin protein in a sample. This invention also provides a host cell comprising a recombinant expression vector to the disintegrin protein and a recombinant expression vector encoding to the disintegrin protein and the human disintegrin metalloprotease protein, fragment or mutant thereof, useful for these purposes. This invention also provides an in vivo or in vitro method for screening for osteoarthritis and other metalloprotease based diseases, capable of manufacture and use in a kit form.

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

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

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                               |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| <b>(51) International Patent Classification <sup>6</sup> :</b><br><b>C12N 15/57, 9/64, C12Q 1/37,</b><br><b>G01N 33/573, C12Q 1/68</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | <b>A3</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | <b>(11) International Publication Number:</b> <b>WO 98/37092</b><br><br><b>(43) International Publication Date:</b> 27 August 1998 (27.08.98) |
| <b>(21) International Application Number:</b> PCT/US98/03490<br><br><b>(22) International Filing Date:</b> 25 February 1998 (25.02.98)<br><br><b>(30) Priority Data:</b><br>08/810,153      25 February 1997 (25.02.97)      US<br>PCT/US97/03217      28 February 1997 (28.02.97)      WO<br><b>(34) Countries for which the regional or international application was filed:</b> US et al.<br><br><b>(71) Applicants:</b> THE PROCTER & GAMBLE COMPANY<br>[US/US]; One Procter & Gamble Plaza, Cincinnati, OH 45202 (US). CASE WESTERN RESERVE UNIVERSITY [US/US]; 10900 Euclid Avenue, Cleveland, OH 44106-7020 (US).<br><br><b>(72) Inventors:</b> TINDAL, Michael, Howard; 620 Cody Pass, Wyoming, OH 45215 (US). HAQQI, Tariq, Mehmood; 2625 Canterbury Road, Cleveland Heights, OH 44118 (US).<br><br><b>(74) Agents:</b> REED, T., David et al.; The Procter & Gamble Company, 5299 Spring Grove Avenue, Cincinnati, OH 45217 (US). | <b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).<br><br><b>Published</b><br><i>With international search report.</i><br><br><b>(88) Date of publication of the international search report:</b><br>5 November 1998 (05.11.98) |                                                                                                                                               |
| <b>(54) Title:</b> USE OF A NOVEL DISINTEGRIN METALLOPROTEASE<br><br><b>(57) Abstract</b><br><br><p>This invention provides a method for identifying compounds capable of binding to the disintegrin protein, and determining the amount and affinity of a compound capable of binding to the disintegrin protein in a sample. This invention also provides a host cell comprising a recombinant expression vector to the disintegrin protein and a recombinant expression vector encoding to the disintegrin protein and the human disintegrin metalloprotease protein, fragment or mutant thereof, useful for these purposes. This invention also provides an <i>in vivo</i> or <i>in vitro</i> method for screening for osteoarthritis and other metalloprotease based diseases, capable of manufacture and use in a kit form.</p>                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                               |

5 USE OF A NOVEL DISINTEGRIN METALLOPROTEASE, MUTANTS,  
FRAGMENTS AND THE LIKE

Field of the invention

10 The invention relates to a novel protein, its fragments and mutants and to its use in detecting and testing drugs for ailments, including osteoarthritis and others characterized by up regulation of metalloproteases.

Background

15 A number of enzymes effect the breakdown of structural proteins and are structurally related metalloproteases. These include human skin fibroblast collagenase, human skin fibroblast gelatinase, human sputum collagenase and gelatinase, and human stromelysin. See e.g., S.E. Whitham et al., Comparison of human stromelysin and collagenase by cloning and sequence analysis" *Biochem J.* 240:913 (1986). See also G.I. Goldberg et al., "Human Fibroblast Collagenase" *J.*  
20 *Biol. Chem.* 261:660 (1986). Metal dependence (e.g., zinc) is a common feature of these structurally related enzymes known as "metalloproteases."

Controlled production and activity of these enzymes plays an important role in the normal development of tissue architecture. In excess, however, these enzymes can cause pathologic destruction of connective tissues. See generally, J. Saus et al., "The  
25 Complete Primary Structure of Human Matrix Metalloprotease-3" *J. Biol. Chem.* 263:6742 (1988). Many of these are zinc-containing metalloprotease enzymes, as are the angiotensin-converting enzymes and the enkephalinases. Collagenase, stromelysin and related enzymes are important in mediating the symptomatology of a number of diseases, including rheumatoid arthritis (Mullins, D. E., et al., *Biochim*  
30 *Biophys Acta* (1983) 695:117-214); osteoarthritis (Henderson, B., et al., *Drugs of the Future* (1990) 15:495-508); the metastasis of tumor cells (ibid, Broadhurst, M. J., et al., *European Patent Application* 276,436 (published 1987), Reich, R., et al., 48 *Cancer Res* 3307-3312 (1988); and various ulcerated conditions. Ulcerative  
35 conditions can result in the cornea as the result of alkali burns or as a result of infection by *Pseudomonas aeruginosa*, *Acanthamoeba*, Herpes simplex and vaccinia viruses.

In fact, measurement of metalloproteases in cancer tissue suggests increased levels of metalloproteases correlate with metastatic potential. See e.g., M. J. Duffy et al., "Assay of matrix metalloproteases types 8 and 9 by ELISA in human breast cancer" *Br. J. Cancer* 71:1025 (1995).

5 Other conditions characterized by undesired metalloprotease activity include periodontal disease, epidermolysis bullosa and scleritis. In view of the involvement of metalloproteases in a number of disease conditions, attempts have been made to prepare inhibitors to these enzymes. A number of such inhibitors are disclosed in the literature. The invention seeks to provide novel inhibitors, preferably specific to this  
10 protease, that have enhanced activity in treating diseases mediated or modulated by this protease.

Inhibitors of metalloproteases are useful in treating diseases caused, at least in part, by breakdown of structural proteins. A variety of inhibitors have been prepared, but there is a continuing need for metalloprotease inhibitor screens to design drugs for  
15 treating such diseases.

Given the involvement of matrix metalloproteases in a number of disease conditions, attempts have been made to identify inhibitors of these enzymes. For Example TapI-2 and 1,10-phenanthroline are known metalloprotease inhibitors. See, e.g., J. Arribas et al., "Diverse Cell Surface Protein Ectodomains Are Shed by a  
20 System Sensitive to Metalloprotease Inhibitors", *J. Biol. Chem.* 271:11376 (1996).

Metalloproteases are a broad class of proteins which have widely varied functions. Disintegrins are zinc metalloproteases, abundant in snake venom. Mammalian disintegrins are a family of proteins with about 18 known subgroups. They act as cell adhesion disrupters and are also known to be active in reproduction  
25 (for example, in fertilization of the egg by the sperm, including fusion thereof, and in sperm maturation).

These proteases and many others are uncovered in molecular biology and biochemistry. As a result, GenBank, a repository for gene sequences, provides several sequences of metalloproteases, including some said to encode fragments of  
30 disintegrins. For example, GenBank accession # Z48444 dated February 25, 1994 discloses 2407 nucleotides of a rat gene said to be a rat disintegrin metalloprotease gene; GenBank accession # Z48579 dated March 2, 1995 discloses 1824 nucleotides of a partial sequence of a gene said to be a human disintegrin metalloprotease gene; GenBank accession # Z21961 dated October 25, 1994, discloses 2397 nucleotides of a  
35 partial sequence of a gene said to be a bovine zinc metalloprotease gene.

Because there is such a wide variety of metalloproteases, there is a continuing need for i) methods that will specifically detect a particular metalloprotease, as well as ii) methods for identifying candidate inhibitors.

It would be advantageous to implicate metalloproteases in specific disease states, and to use these metalloproteases as tools to detect and ultimately cure, control or design cures for such diseases.

#### OBJECTS OF THE INVENTION

It is an object of the present invention to provide a method for identifying compounds capable of binding to the disintegrin protein.

10 It is also an object of the present invention to provide a host cell comprising a recombinant expression vector to the disintegrin protein and a recombinant expression vector encoding to the disintegrin protein.

15 It is also an object of the present invention to provide a method for screening for metalloprotease mediated diseases such as cancer, arthropathies (including ankylosing spondylitis, rheumatoid arthritis, gouty arthritis (gout), inflammatory arthritis, Lyme disease and osteoarthritis).

It is also an object of the present invention to provide an antibody to the protein useful in the screen, in the isolation of the protein or as a targeting moiety for the protein.

20

#### SUMMARY OF THE INVENTION

This invention provides a method for identifying compounds capable of binding to the disintegrin protein, and determining the amount and affinity of a compound capable of binding to the disintegrin protein in a sample.

25 This invention also provides a host cell comprising a recombinant expression vector to the disintegrin protein and a recombinant expression vector encoding to the disintegrin protein and the human disintegrin metalloprotease protein, fragment or mutant thereof, useful for these purposes.

30 This invention also provides an in vivo or in vitro method for screening for osteoarthritis and other metalloprotease based diseases, such as cancer, capable of manufacture and use in a kit form.

#### DETAILED DESCRIPTION

35 The term "gene" refers to a DNA sequence that comprises control and coding sequences necessary for the production of a mature protein or precursor thereof. The protein can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired enzymatic activity is retained.

The term "oligonucleotide" as used herein is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, usually more than three (3).

and typically more than ten (10) and up to one hundred (100) or more (although preferably between twenty and thirty). The exact size will depend on many factors, which in turn depends on the ultimate function or use of the oligonucleotide. The oligonucleotide may be generated in any manner, including chemical synthesis, DNA replication, restriction endonuclease digestion reverse transcription, or a combination thereof.

Because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage, an end of an oligonucleotide is referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends.

When two different, non-overlapping oligonucleotides anneal to different regions of the same linear complementary nucleic acid sequence, and the 3' end of one oligonucleotide points towards the 5' end of the other, the former may be called the "upstream" oligonucleotide and the latter the "downstream" oligonucleotide.

The term "primer" refers to an oligonucleotide which is capable of acting as a point of initiation of synthesis when placed under conditions in which primer extension is initiated. An oligonucleotide "primer" may occur naturally, as in a purified restriction digest or may be produced synthetically.

A primer is selected to be "substantially" complementary to a strand of specific sequence of the template. A primer must be sufficiently complementary to hybridize with a template strand for primer elongation to occur. A primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being substantially complementary to the strand. Non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the template to hybridize and thereby form a template primer complex for synthesis of the extension product of the primer.

"Hybridization" methods involve the annealing of a complementary sequence to the target nucleic acid (the sequence to be detected). The ability of two polymers of nucleic acid containing complementary sequences to find each other and anneal through base pairing interaction is a well-recognized phenomenon. The initial observations of the "hybridization" process by Marmur and Lane, *Proc. Natl. Acad.*

*Sci. USA* 46:453 (1960) and Doty et al., *Proc. Natl. Acad. Sci. USA* 46:461 (1960) have been followed by the refinement of this process into an essential tool of modern biology. Nonetheless, a number of problems have prevented the wide scale use of hybridization as a tool in human diagnostics. Among the more formidable problems  
5 are: 1) the inefficiency of hybridization; 2) low concentration of specific target sequences in a mixture of genomic DNA; and 3) the hybridization of only partially complementary probes and targets.

With regard to efficiency, it is experimentally observed that only a fraction of the possible number of probe-target complexes are formed in a hybridization reaction.  
10 This is particularly true with short oligonucleotide probes (less than 100 bases in length). There are three fundamental causes: a) hybridization cannot occur because of secondary and tertiary structure interactions; b) strands of DNA containing the target sequence have rehybridized (reannealed) to their complementary strand; and c) some target molecules are prevented from hybridization when they are used in  
15 hybridization formats that immobilize the target nucleic acids to a solid surface.

Even where the sequence of a probe is completely complementary to the sequence of the target, i.e., the target's primary structure, the target sequence must be made accessible to the probe via rearrangements of higher-order structure. These higher-order structural rearrangements may concern either the secondary structure or  
20 tertiary structure of the molecule. Secondary structure is determined by intramolecular bonding. In the case of DNA or RNA targets this consists of hybridization within a single, continuous strand of bases (as opposed to hybridization between two different strands). Depending on the extent and position of intramolecular bonding, the probe can be displaced from the target sequence  
25 preventing hybridization.

Solution hybridization of oligonucleotide probes to denatured double-stranded DNA is further complicated by the fact that the longer complementary target strands can renature or reanneal. Again, hybridized probe is displaced by this process. This results in a low yield of hybridization (low "coverage") relative to the starting  
30 concentrations of probe and target.

With regard to low target sequence concentration, the DNA fragment containing the target sequence is usually in relatively low abundance in genomic DNA. This presents great technical difficulties: most conventional methods that use oligonucleotide probes lack the sensitivity necessary to detect hybridization at such  
35 low levels.

One attempt at a solution to the target sequence concentration problem is the amplification of the detection signal. Most often this entails placing one or more

labels on an oligonucleotide probe. In the case of non-radioactive labels, even the highest affinity reagents have been found to be unsuitable for the detection of single copy genes in genomic DNA with oligonucleotide probes. See Wallace et al., *Biochimie* 67:755 (1985). In the case of radioactive oligonucleotide probes, only  
5 extremely high specific activities are found to show satisfactory results. See Studencki and Wallace, *DNA* 3:1 (1984) and Studencki et al., *Human Genetics* 37:42 (1985).

K. B. Mullis et al., U.S. Patent Nos. 4,683,195 and 4,683,202, hereby incorporated by reference, describe a method for increasing the concentration of a  
10 segment of a target sequence in a mixture of any DNA without cloning or purification. This process for amplifying the target sequence (which can be used in conjunction with the present invention to make target molecules) consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of  
15 a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then allowed to anneal to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer  
20 annealing, and primer extension can be repeated many times (i.e., denaturation, annealing and extension constitute one "cycle;" there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this  
25 length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to by the inventors as the "Polymerase Chain Reaction" (hereinafter PCR). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified."

30 With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (e.g., hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of  $^{32}\text{P}$  labeled deoxynucleotide triphosphates, e.g., dCTP or dATP, into the amplified segment). In addition to  
35 genomic DNA, any oligonucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications.

The PCR amplification process is known to reach a plateau concentration of specific target sequences of approximately  $10^{-8}$  M. A typical reaction volume is 100  $\mu$ l, which corresponds to a yield of  $6 \times 10^{11}$  double stranded product molecules.

With regard to complementarity, it is important for some diagnostic applications to determine whether the hybridization represents complete or partial complementarity. For example, where it is desired to detect simply the presence or absence of pathogen DNA or RNA (such as from a virus, bacterium, fungi, mycoplasma, protozoan) it is only important that the hybridization method ensures hybridization when the relevant sequence is present; conditions can be selected where both partially complementary probes and completely complementary probes will hybridize. Other diagnostic applications, however, may require that the hybridization method distinguish between partial and complete complementarity. It may be of interest to detect genetic polymorphisms. For example, human hemoglobin is composed, in part, of four polypeptide chains. Two of these chains are identical chains of 141 amino acids (alpha chains) and two of these chains are identical chains of 146 amino acids (beta chains). The gene encoding the beta chain is known to exhibit polymorphism. The normal allele encodes a beta chain having glutamic acid at the sixth position. The mutant allele encodes a beta chain having valine at the sixth position. This difference in amino acids has a profound (most profound when the individual is homozygous for the mutant allele) physiological impact known clinically as sickle cell anemia. It is well known that the genetic basis of the amino acid change involves a single base difference between the normal allele DNA sequence and the mutant allele DNA sequence.

Unless combined with other techniques (such as restriction enzyme analysis), methods that allow for the same level of hybridization in the case of both partial as well as complete complementarity are typically unsuited for such applications; the probe will hybridize to both the normal and variant target sequence. Hybridization, regardless of the method used, requires some degree of complementarity between the sequence being assayed (the target sequence) and the fragment of DNA used to perform the test (the probe). (Of course, one can obtain binding without any complementarity but this binding is nonspecific and to be avoided.)

The complement of a nucleic acid sequence as used herein refers to an oligonucleotide which, when aligned with the nucleic acid sequence such that the 5' end of one sequence is paired with the 3' end of the other, is in "antiparallel association." Certain bases not commonly found in natural nucleic acids may be included in the nucleic acids of the present invention and include, for example, inosine and 7-deazaguanine complementarity need not be perfect; stable duplexes may

contain mismatched base pairs or unmatched bases. Those skilled in the art of nucleic acid technology can determine duplex stability empirically considering a number of variables including, for example, the length of the oligonucleotide, base composition and sequence of the oligonucleotide, ionic strength and incidence of mismatched base pairs.

Stability of a nucleic acid duplex is measured by the melting temperature, or " $T_m$ ." The  $T_m$  of a particular nucleic acid duplex under specific conditions is the temperature at which on average half of the base pairs have disassociated. The equation for calculating the  $T_m$  of nucleic acids is well known in the art. As indicated by strand references, an estimate of the  $T_m$  value may be calculated by the equation:

$$T_m = 81.5^{\circ}\text{C} + 16.6 \log M + .41(\%GC) - 0.61(\% \text{ form}) - 500/L$$

where  $M$  is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, %form is the percentage of formamide in the hybridization solution, and  $L$  = length of the hybrid in base pairs [See, e.g., Guide to Molecular Cloning Techniques, Ed. S.L. Berger and A.R. Kimmel, in Methods in Enzymology Vol. 152, 401 (1987)]. Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of  $T_m$ .

The term "probe" as used herein refers to a labeled oligonucleotide which forms a duplex structure with a sequence in another nucleic acid, due to complementarity of at least one sequence in the probe with a sequence in the other nucleic acid.

The term "label" as used herein refers to any atom or molecule which can be used to provide a detectable (preferably quantifiable) signal, and which can be attached to a nucleic acid or protein. Labels may provide signals detectable fluorescence, radioactivity, colorimetry, gravimetry, X-ray diffraction or absorption, magnetism, enzymatic activity, and the like. Such labels can be added to the oligonucleotides of the present invention.

The terms "nucleic acid substrate" and "nucleic acid template" are used herein interchangeably and refer to a nucleic acid molecule which may comprise single- or double-stranded DNA or RNA.

The term "substantially single-stranded" when used in reference to a nucleic acid substrate means that the substrate molecule exists primarily as a single strand of nucleic acid in contrast to a double-stranded substrate which exists as two strands of nucleic acid which are held together by inter-strand base pairing interactions.

The term "sequence variation" as used herein refers to differences in nucleic acid sequence between two nucleic acid templates. For example, a wild-type structural gene and a mutant form of this wild-type structural gene may vary in

sequence by the presence of single base substitutions and/or deletions or insertions of one or more nucleotides. These two forms of the structural gene are said to vary in sequence from one another. A second mutant form of the structural gene may exist. This second mutant form is said to vary in sequence from both the wild-type gene and the first mutant form of the gene. It should be noted that, while the invention does not require that a comparison be made between one or more forms of a gene to detect sequence variations, such comparisons are possible with the oligo/solid support matrix of the present invention using particular hybridization conditions as described in U.S. Pat. Appl. Ser. No. 08/231,440, hereby incorporated by reference.

10 "Oligonucleotide primers matching or complementary to a gene sequence" refers to oligonucleotide primers capable of facilitating the template-dependent synthesis of single or double-stranded nucleic acids. Oligonucleotide primers matching or complementary to a gene sequence may be used in PCRs, reverse transcriptase-PCR (RT-PCRs) and the like.

15 A "consensus gene sequence" refers to a gene sequence which is derived by comparison of two or more gene sequences and which describes the nucleotides most often present in a given segment of the genes; the consensus sequence is the canonical sequence.

As used herein, the terms "protein" and "protease" refer to metalloprotease.

20 The term "metalloprotease" refers to a native metal dependent protease, a fragment thereof, a mutant or homologue which still retains its function. The invention contemplates metalloproteases (or "disintegrins") from differing species, and those prepared by recombinant methods, *in vitro* methods, or standard peptide synthesis. Preferably the protein is a human disintegrin or mutant thereof. For the purposes of

25 defining the mutants of the protein the preferred "native" protein is partially described in Gen Bank accession #Z48579, incorporated herein by reference and referred to in the sequence below. Homologue disintegrins include whole proteins with at least 90% homology as understood by the art, or fragments thereof. It is recognized that some interspecies variation may occur including insertions or deletions which may or

30 may not alter function. For example, a rat protein which is 95% homologous to the protein based on the peptide sequence, and a bovine protein (based on DNA sequence) being 97-98% homologous based on the first 300 base pairs are both considered homologues. For reference GenBank accession #Z48444 dated February 25, 1994 discloses 2407 bases of a rat gene said to be a rat disintegrin metalloprotease gene;

35 GenBank accession #Z21961 dated October 25, 1994, discloses 2397 bases of a partial sequence of a gene said to be a bovine zinc metalloprotease gene. Preferably this metalloprotease is a human disintegrin as described below.

The term "antibody" refers to an antibody to a disintegrin, or fragment thereof. These many be monoclonal or polyclonal, and can be from any of several sources. The invention also contemplates fragments of these antibodies made by any method in the protein or peptide art.

5       The term "disease screen" refers to a screen for a disease or disease state. A disease state is the physiological or cellular or biochemical manifestation of the disease. Preferably this screen is used on body tissues or fluids of an animal or cell culture, using standard techniques, such as ELISA. It also contemplates "mapping" of disease in a whole body, such as by labeled antibody as described above given  
10       systemically: regardless of the detection method, preferable such detection methods include fluorescence, X-ray (including CAT scan), NMR (Including MRI), and the like.

      The term "compound screen" is related to the methods and screens related to finding compounds, determining their affinity for the protease, or designing or  
15       selecting compounds based on the screen. In another embodiment, it contemplates the use of the three dimensional structure for drug design, preferable "rational drug design", as understood by the art. It may be preferred that the protease is in "essentially pure form", which refers to a protein reasonably free of other impurities, so as to make it useful for experiments or characterization. Use of this screening  
20       method assists the skilled artisan in finding novel structures, whether made by the chemist or by nature, which bind to and preferably inhibit the protease. These "inhibitors" may be useful in regulating or modulating the activity of the protease, and may be used to thus modulate the biological cascade that they function in. This approach affords new pharmaceutically useful compounds.

25       The term "disintegrin" refers to a disintegrin, a fragment thereof, a mutant thereof or a homologue which still retains its function. This term contemplates aggrecanase, and other proteases which are involved in or modulate tissue remodeling. This contemplates disintegrins from differing species, and those prepared by recombinant methods, in vitro methods, or standard peptide synthesis. Preferably the  
30       protein is a human disintegrin or mutant thereof. For the purposes of defining the mutants, with reference to a protein is partially described in GenBank accession # Z48579, incorporated herein by reference and referred to in the sequence below. SEQ ID NO:1 describes a fragment of that DNA sequence and its transcript and SEQ ID NO:2 describes the protein coded by the gene. Homologue disintegrins include whole  
35       proteins with at least 90% homology as understood by the art, or fragments thereof. For example, a rat protein which is 95% homologous to that of SEQ ID NO:2 based on the amino acid sequence derived from the DNA or cDNA sequence containing

SEQ ID NO:1, and a bovine protein (similarly derived) being 97-98% homologous, are both considered homologues. Thus homologous cDNAs cloned from other organisms give rise to homologous proteins.

Likewise proteins may be considered homologues based on the amino acid sequence alone. Practical limitations of amino acid sequencing would allow one to determine that a protein is homologous to another using, for example, comparison of the first 50 amino acids of the protein. Hence 90% homology in would allow for 5 differing amino acids in the chain of the first 50 amino acids of the homologous protein.

The skilled artisan will appreciate that the degeneracy of the genetic code provides for differing DNA sequences to provide the equivalent transcript, and thus the same protein. In certain cases preparing the DNA sequence, which encodes for the same protein, but differs from the native DNA include;

- ease of sequencing or synthesis;
- increased expression of the protein; and
- preference of certain heterologous hosts for certain codons over others.

These practical considerations are widely known and provide embodiments that may be advantageous to the user of the invention. Thus it is clearly contemplated that the native DNA is not the only embodiment envisioned in this invention.

In addition it is apparent to the skilled artisan that fragments of the protein may be used in screening, drug design and the like, and that the entire protein may not be required for the purposes of using the invention. Thus it is clearly contemplated that the skilled artisan will understand that the disclosure of the protein and its uses contemplates the useful peptide fragments.

The practical considerations of protein expression, purification yield, stability, solubility, and the like, are considered by the skilled artisan when choosing whether to use a fragment, and the fragment to be used. As a result, using routine practices in the art, the artisan can, given this disclosure practice the invention using fragments of the protein as well.

Thus, the present invention specifically contemplates the use of less than the entire nucleic acid sequence for the gene and less than the entire amino acid sequence of the protein. Fragments of the protein may be used in screening, drug design and the like, and that the entire protein may not be required for the purposes of using the invention. The protein itself can be used to determine the binding activity of small molecules to the protein. Drug screening using enzymatic targets is used in the art and can be employed using automated, high throughput technologies.

The protein or protease itself can be used to determine the binding activity of small molecules to the protein. Drug screening using enzymatic targets is used in the art and can be employed using automated, high throughput technologies.

5 The inhibition of disintegrin activity may be a predictor of efficacy in the treatment of osteoarthritis, and other diseases involving degeneration of articular cartilage and other tissues having matrix degradation, such as tissue remodeling and the like.

#### Gene therapy

10 Without being bound by theory it is thought that the metalloprotease is up regulated during osteoarthritis in tissues. We have surprisingly found that a human disintegrin is up-regulated in human chondrocytes during osteoarthritic conditions. Inhibition of signal transduction mechanism is efficacious in disrupting the cascade of events in osteoarthritis and other diseases involving cartilage degeneration. The skilled artisan will recognize that if up-regulation is a cause of the onset of arthritis, 15 then interfering with the activity of this gene may be useful in treating osteoarthritis.

This is done by any of several methods, including gene (i.e., antisense) therapy.

#### Purification of the protease

20 Media, cell extracts or inclusion bodies from mammalian, yeast, insect or eukaryotic cells containing recombinant disintegrin or fragments of the full length protein are used for purification of disintegrin or fragments of disintegrin. Solutions consisting of denatured disintegrin may be refolded prior to purification across successive chromatographic resins or following the final stage of separation. Media, cell extracts, or solubilized disintegrin are prepared in the presence of one or a 25 combination of detergents, denaturants or organic solvents, such as octylglucoside, urea or dimethylsulfoxide, as required. Ion exchange and hydrophobic interaction chromatography are used individually or in combination for the separation of recombinant disintegrin from contaminating cell material. Such material is applied to the column and disintegrin is eluted by adjustment of pH, changes in ionic strength, 30 addition of denaturant and/or use of organic solvent. Typically, solutions containing disintegrin are then passed over an antibody affinity column or ligand affinity column for site specific purification of disintegrin. The immunoaffinity column contains an antibody specific for disintegrin immobilized on a solid support such as Sepharose 4B (Pharmacia) or other similar materials. Preferably, the column is washed to remove

unbound proteins and the disintegrin is eluted via low pH glycine buffer or high ionic strength. The ligand affinity column may have specificity for the active site of disintegrin or to a portion of the molecule adjacent or removed from the active site. The column is washed and disintegrin is eluted by addition of a competing molecule  
5 to the elution buffer. Preferably, a protease inhibitor cocktail containing one or more protease inhibitors, such as benzamidine, leupeptin, phosphoramidon, phenylmethylsulfonyl fluoride, and 1,10-phenanthroline is present throughout the purification procedure. Various detergents such as octylthioglucoside and Triton X-100 or chemical agents such as glycerol may be added to increase disintegrin  
10 solubility and stability. Final purification of the protein is achieved by gel filtration across a chromatographic support, if required.

#### Inhibitors of the protease

The protease of the invention can be used to find inhibitors of the protease. Hence it is useful as a screening tool or for rational drug design. Without being bound  
15 by theory, the protease may modulate cellular remodeling and in fact may enhance extracellular matrix remodeling and thus enhance tissue breakdown. Hence inhibition of disintegrin provides a therapeutic route for treatment of diseases characterized by these processes.

In screening, a drug compound can be used to determine both the quality and  
20 quantity of inhibition. As a result such screening provides information for selection of actives, preferably small molecule actives, which are useful in treating these diseases.

In therapy, inhibition of disintegrin metalloprotease activity via binding of small molecular weight, synthetic metalloprotease inhibitors, such as those used to  
25 inhibit the matrix metalloproteases would be used to inhibit extracellular matrix remodeling.

#### Antibodies to the protein

Metalloproteases can be targeted by conjugating a metalloprotease inhibitor to a to an antibody or fragment thereof. Conjugation methods are known in the art.  
30 These antibodies are then useful both in therapy and in monitoring the dosage of the inhibitors.

The antibody of the invention can also be conjugated to solid supports. These conjugates can be used as affinity reagents for the purification of a desired metalloprotease, preferably a disintegrin.

In another aspect, the antibody of the invention is directly conjugated to a  
5 label. As the antibody binds to the metalloprotease, the label can be used to detect the presence of relatively high levels of metalloprotease in vivo or in vitro cell culture.

For example, targeting ligand which specifically reacts with a marker for the intended target tissue can be used. Methods for coupling the invention compound to the targeting ligand are well known and are similar to those described below for  
10 coupling to carrier. The conjugates are formulated and administered as described above.

#### Preparation and Use of Antibodies:

Antibodies may be made by several methods, for example, the protein may be injected into suitable (e.g., mammalian) subjects including mice, rabbits, and the like.  
15 Preferred protocols involve repeated injection of the immunogen in the presence of adjuvants according to a schedule which boosts production of antibodies in the serum. The titers of the immune serum can readily be measured using immunoassay procedures, now standard in the art.

The antisera obtained can be used directly or monoclonal antibodies may be  
20 obtained by harvesting the peripheral blood lymphocytes or the spleen of the immunized animal and immortalizing the antibody-producing cells, followed by identifying the suitable antibody producers using standard immunoassay techniques.

Polyclonal or monoclonal preparations are useful in monitoring therapy or prophylaxis regimens involving the compounds of the invention. Suitable samples  
25 such as those derived from blood, serum, urine, or saliva can be tested for the presence of the protein at various times during the treatment protocol using standard immunoassay techniques which employ the antibody preparations of the invention.

These antibodies can also be coupled to labels such as scintigraphic labels, e.g., Tc-99 or I-131, using standard coupling methods. The labeled compounds are  
30 administered to subjects to determine the locations of excess amounts of one or more metalloproteases in vivo. Hence a labeled antibody to the protein would operate as a screening tool for such enhanced expression, indicating the disease.

The ability of the antibodies to bind metalloprotease selectively is thus taken advantage of to map the distribution of these enzymes in situ. The techniques can also

be employed in histological procedures and the labeled antibodies can be used in competitive immunoassays.

Antibodies are advantageously coupled to other compounds or materials using known methods. For example, materials having a carboxyl functionality, the carboxyl  
5 residue can be reduced to an aldehyde and coupled to carrier through reaction with side chain amino groups, optionally followed by reduction of imino linkage formed. The carboxyl residue can also be reacted with side chain amino groups using condensing agents such as dicyclohexyl carbodiimide or other carbodiimide dehydrating agents. Linker compounds can also be used to effect the coupling; both  
10 homobifunctional and heterobifunctional linkers are available from Pierce Chemical Company, Rockford, Ill.

These antibodies, when conjugated to a suitable chromatography material are useful in isolating the protein. Separation methods using affinity chromatography are well known in the art, and are within the purview of the skilled artisan.

15 Disease marker

As noted above, the present invention contemplates detecting expression of metalloprotease genes in samples, including samples of diseased tissue. It is not intended that the present invention be limited by the nature of the source of nucleic acid (whether DNA or RNA); a variety of sources is contemplated, including but not  
20 limited to mammalian (e.g., cancer tissue, lymphocytes, etc.), sources.

Without being bound by theory, expression of genes, and preferably this gene may have a restricted tissue distribution and its expression is up regulated by potential osteoarthritis mediators. Enhanced expression of this gene (and hence its protein) for example, in articular chondrocytes provides a marker to monitor the development,  
25 including the earliest, asymptomatic stages, and the progression of osteoarthritis. Hence an antibody raised to the protein would operate a screening tool for such enhanced expression, indicating the disease.

In addition, when used in a disease screen, antibodies can be conjugated to chromophore or fluorophore containing materials, or can be conjugated to enzymes  
30 which produce chromophores or fluorophores in certain conditions. These conjugating materials and methods are well known in the art. When used in this manner detection of the protein by immunoassay is straightforward to the skilled artisan. Body fluids, (serum, urine, synovial fluid) for example can be screened in this manner for calibration, and detection of distribution of metalloproteases, or increased  
35 levels of these proteases.

When used in this way the invention is a useful diagnostic and/or clinical marker for metalloprotease mediated diseases, such as osteoarthritis or other articular

cartilage degenerative diseases or other diseases characterized by degradation or remodeling of extracellular matrix. When disease is detected, it may be treated before the onset of symptom or debilitation.

Furthermore, such antibodies can be used to target diseased tissue, for  
5 detection or treatment as described above.

#### Nucleic Acid Derived Tools

The nucleic acid content of cells consists of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The DNA contains the genetic blueprint of the cell. RNA is involved as an intermediary in the production of proteins based on the DNA sequence.  
10 RNA exists in three forms within cells, structural RNA (i.e., ribosomal RNA "rRNA"), transfer RNA ("tRNA"), which is involved in translation, and messenger RNA ("mRNA"). Since the mRNA is the intermediate molecule between the genetic information encoded in the DNA, and the corresponding proteins, the cell's mRNA component at any given time is representative of the physiological state of the cell. In  
15 order to study and utilize the molecular biology of the cell, it is therefore important to be able to purify mRNA, including purifying mRNA from the total nucleic acid of a sample.

The preparation of RNA is complicated by the presence of ribonucleases that degrade RNA (e.g., T. Maniatis et al., *Molecular Cloning*, pp. 188-190, Cold Spring  
20 Harbor Laboratory [1982]). Furthermore, the preparation of amplifiable RNA is made difficult by the presence of ribonucleoproteins in association with RNA. (See, R. J. Slater, In: Techniques in Molecular Biology, J. M. Walter and W. Gaastra, eds., Macmillan, NY, pp. 113-120 [1983]).

Typically, the steps involved in purification of nucleic acid from cells include  
25 1) cell lysis; 2) inactivation of cellular nucleases; and 3) separation of the desired nucleic acid from the cellular debris and other nucleic acid. Cell lysis may be achieved through various methods, including enzymatic, detergent or chaotropic agent treatment. Inactivation of cellular nucleases may be achieved by the use of proteases and/or the use of strong salts. Finally, separation of the desired nucleic acid is  
30 typically achieved by extraction of the nucleic acid with phenol or phenol-chloroform; this method partitions the sample into an aqueous phase (which contains the nucleic acids) and an organic phase (which contains other cellular components, including proteins). Commonly used protocols require the use of salts in conjunction with phenol (P. Chomczynski and N. Sacchi, *Anal. Biochem.* 162:156 [1987]), or employ a  
35 centrifugation step to remove the protein (R. J. Slater, *supra*).

Once the nucleic acid fraction has been isolated from the cell, the structure of the mRNA molecule may be used to assist in the purification of mRNA from DNA

and other RNA molecules. Because the mRNA of higher organisms is usually polyadenylated on its 3' end ("poly-A tail" or "poly-A track"), one means of isolating RNA from cells has been based on binding the poly-A tail with its complementary sequence (i.e., oligo-dT), that has been linked to a support such as cellulose. Commonly, the hybridized mRNA/oligo-dT is separated from the other components present in the sample through centrifugation or, in the case of magnetic formats, exposure to a magnetic field. Once the hybridized mRNA/oligo-dT is separated from the other sample components, the mRNA is usually removed from the oligo-dT. However, for some applications, the mRNA may remain bound to the oligo-dT that is linked to a solid support.

A wide variety of solid supports with linked oligo-dT have been developed and are commercially available. Cellulose remains the most common support for most oligo-dT systems, although formats with oligo-dT covalently linked to latex beads and paramagnetic particles have also been developed and are commercially available. The paramagnetic particles may be used in a biotin-avidin system, in which biotinylated oligo-dT is annealed in solution to mRNA. The hybrids are then captured with streptavidin-coated paramagnetic particles, and separated using a magnetic field. In addition to these methods, variations exist, such as affinity purification of polyadenylated RNA from eukaryotic total RNA in a spun-column format. These approaches allow for hybridization of poly-A mRNA, but vary in efficiency and sensitivity.

In one embodiment, the mRNA is treated with reverse transcriptase to make cDNA. The cDNA can be used in primer extension and PCR using the primers described below. Thus, the present invention contemplates nucleic acid molecules detectable by primer extension using the primers described below. Primer extension (and PCR for that matter) can be carried out under conditions (so-called "high stringency conditions") such that only complementary nucleic acid will hybridize (as opposed to hybridization with partially complementary nucleic acid). These conditions including annealing at or near the melting temperature of the duplex.

#### 30 Primers Directed To A Specific Disintegrin Metalloprotease Gene

The invention provides a partial nucleic acid full length protein coding region sequence of a novel disintegrin metalloprotease gene useful for, among other things, the detection of disintegrin metalloprotease gene expression. In one embodiment, primers directed to a portion of this partial sequence are used to detect the presence or absence of the gene sequence. These primers can also be used for the identification of a cDNA clone representing the entire gene, allowing for recombinant expression in

a host cell of the nucleic acid sequence encoding the disintegrin metalloprotease or fragments (or mutants) thereof.

Preferred primers are primer SEQ ID NO:9 (5'-AGCCTGTGTC-3') and SEQ ID NO:10 (5'-AGCCTGTGTCTGAACCACT-3'). However, other primers can be readily designed from the sequences set forth in SEQ ID NO:5 and SEQ ID NO:1.

#### Method of Comparing Biological Samples by Differential Display

Successful amplification can be confirmed by characterization of the product(s) from the reaction. It is not intended that the present invention be limited by the method by which extension products or PCR products are detected. In one embodiment, the PCR products are analyzed by high resolution agarose gel electrophoresis using 2% agarose gels (BRL) and the amplified DNA fragments are visualized by ethidium bromide staining and UV transillumination. The present invention contemplates, in one embodiment, using electrophoresis to confirm product formation and compare the results between samples.

Hence, the present invention contemplates detection of sequences of the novel disintegrin metalloprotease gene in mixtures of nucleic acid (e.g., cDNA or RT-mRNA). By carrying out PCR on a mixture of nucleic acid and running the products on gels, nucleic acid comprising a sequence that is defined by the primers is "isolated." The product can thereafter be "purified" by cutting the band from the gel (or by other suitable methods such as electroelution).

#### Synopsis of the Sequence Listing

For the aid of the reader, the inter-relation of the sequence listings are described hereinbelow:

SEQ ID NO:1 is a fragmentary DNA sequence, and is part of SEQ ID NO:3. The first base (Cytosine or C) of SEQ ID NO:1 is base 940 of SEQ ID NO:3. The DNA sequences are identical where they overlap.

SEQ ID NO:2 and SEQ ID NO:4, are the expressed amino acid sequences of SEQ ID NO:1 and SEQ ID NO:3 respectively. The first amino acid of SEQ ID NO:2, Gln, is the 309th amino acid in SEQ ID NO:4. The two sequences are homologous to the carboxy terminus of the protein.

SEQ ID NO:7 is a sense strand of DNA provided by differential display experiments. The first base of SEQ ID NO:7 corresponds to base 1371 of SEQ ID NO:1, and to base 2310 of SEQ ID NO:3. These sequences are homologous for 452 bases, to base 1822 of SEQ ID NO:1 and to base 2761 of SEQ ID NO:3. The

difference in the last two bases of SEQ ID NO:1 and SEQ ID NO:3 may be due to errors in sequencing or a common replicatory error found in PCR, or may be part of a cloning vector. SEQ ID NO:7 continues some 284 bases beyond the homology, and thus well beyond the terminus of SEQ ID NO:1 and SEQ ID NO:3.

5 In addition, bases 477 to 716 of SEQ ID NO:7 are the SEQ ID NO 6. SEQ ID NO 6 is the sense strand of SEQ ID NO:5, which is an antisense strand found via differential display cloning. Hence SEQ ID NO: 6 shows the DNA orientation as it would appear in the mRNA. These two sequences are found near the 3' end of this gene.

10 Although bases 452 to the 3' end of SEQ ID NO:7 differ from SEQ ID NO:1 and SEQ ID NO:3, SEQ ID NO:7 is nonetheless valid. It is essential to note that the expressed peptide sequence is not affected by this difference. It is likely these bases do not appear in SEQ ID NO:1 and SEQ ID NO:3 because of the use of an alternative polyadenylation signal.

15 SEQ ID NO 8 is a novel full length DNA sequence. SEQ ID NO:9 is the novel expressed protein of SEQ ID NO:8. SEQ ID NO:9 differs from SEQ ID NO:4 in that amino acids 162 (Ser)-213 (Tyr) of SEQ ID NO: 4 is replaced by a single residue, Asn, at position 162 of SEQ ID NO:9. That change is reflected in the DNA by a deletion bases 501-654 for a total of 153 bases, leaving the reading frame intact but  
20 changing one residue and deleting the 51 amino acids present in SEQ ID NO:4.

SEQ ID NO:10 and SEQ ID NO:11 are antisense primers useful in PCR, and are the inverse of the 3' terminus of SEQ ID NO:7, other sequences for primers are discernible by the skilled artisan using sequences referred to herein.

### EXAMPLES

25 The following non-limiting examples illustrate a preferred embodiment of the present invention, and briefly describe the uses of the present invention. These examples are provided for the guidance of the skilled artisan, and do not limit the invention in any way. Armed with this disclosure and these examples the skilled artisan is capable of making and using the claimed invention.

30 Standard starting materials are used for these examples. Many of these materials are known and commercially available. For example, E. coli CJ236 and JM101 are known strains, pUB110 is a known plasmid and Kunkel method

mutagenesis is also well known in the art. In addition certain cell lines and cDNA may be commercially available, for example U-937, available from Clontech Inc., Palo Alto, California.

5 Variants may be made by expression systems and by various methods in various hosts, these methods are within the scope of the practice of the skilled artisan in molecular biology, biochemistry or other arts related to biotechnology.

#### **Example 1**

RNA is isolated from unstimulated and interleukin-1 stimulated cultures of normal human articular chondrocytes. The RNA is reverse transcribed into cDNA.  
10 The cDNA is subjected to a modified differential display procedure using a series of random primers.

PCR samples generated from both stimulated and unstimulated chondrocytes are electrophoresed in adjacent lanes on polyacrylamide gels. The differentially expressed band is excised from the gel, cloned, and sequenced. The differential  
15 expression of the gene is confirmed by RNAase protection and nuclear run on experiments.

#### **Example 2**

A novel partial human cDNA coding the protein is cloned from primary cultures of interleukin-1 stimulated human articular (femoral head) chondrocytes,  
20 using known methods. The same sequence is found, and the gene completed by screening of human cDNA libraries to obtain full length clones.

#### **Example 3**

The cloned DNA of example 2 is placed in pUB110 using known methods.

This plasmid is used to transform *E. coli* and provides a template for site-  
25 directed mutagenesis to create new mutants. Kunkel method mutagenesis was performed altering the Gln 1 to Ala.

#### **Example 4**

[<sup>125</sup>I] disintegrin antibody is prepared using IODOBEADS (Pierce, Rockford, IL; immobilized chloramine-T on nonporous polystyrene beads). Lyophilized antibody (2  
30 µg) is taken up in 50 µl of 10 mM acetic acid and added to 450 µl of phosphate-buffered saline (PBS) (Sigma, St. Louis, MO) on ice. To the tube is added 500 µCurie of <sup>125</sup>I (Amersham, Arlington Heights, IL) (2200Ci/mmol) in 5 µl, and one IODOBEAD. The reaction is incubated on ice for 10 min with occasional shaking. The reaction is then terminated by removal of the reaction from the IODOBEAD. To  
35 remove unreacted <sup>125</sup>I, the mixture is applied to a PD-10 gel filtration column.

#### **Example 5**

A fluorogenic disintegrin metalloprotease substrate peptide (Bachem, Guelph Mills, King of Prussia, Pa) is mixed with the disintegrin and change in the fluorescence is evaluated at 2 min, as a control. Then the fluorogenic peptide is mixed with the disintegrin in the presence of the compound (metalloprotease inhibitor) in evaluation in a separate run, with evaluation at various time points over 2 to 12 hours. Data are evaluated using standard methodology to provide relative binding of the evaluated compound.

#### **Example 6**

0.5 ml of synovial fluid from the left knee of a patient is withdrawn and tested for elevated levels disintegrin by ELISA. The results indicate higher than normal disintegrin level. The patient is prescribed a prophylactic dose of a disintegrin inhibitor administered orally over time or is administered an injection of same in the left knee before leaving the clinician's office.

#### **Example 7**

Inhibition of extracellular matrix remodeling is explored via inhibition of disintegrin metalloprotease activity. Using a small molecular weight, synthetic metalloprotease inhibitor, such as those used to inhibit the matrix metalloproteases, tissue integrity and proteoglycan is monitored.

A sample of IL-1 stimulated bovine nasal cartilage derived articular cartilage is grown in a 1 micromolar solution of a small molecular weight disintegrin inhibitor. The experiment is controlled and compared to an identical culture grown with no inhibitor.

The assay of the culture after 7 days shows that the inhibited culture has less tissue breakdown and less proteoglycan present in the serum of the culture. The result is consistent with the inhibited aggrecanase activity. Inhibition of aggrecanase would inhibit tissue breakdown and reduce the release of proteoglycan.

#### **Example 8**

Inhibition of proteolytic processing resulting in the release from the membrane bound form of the disintegrin metalloprotease domain inhibits "second messenger" signaling of the membrane bound disintegrin molecule. Such second messenger signaling would result in cellular phenotypic changes, changes in gene expression, changes in mitotic activity, and the like.

Cells known to contain disintegrin are treated with a serine protease. Proteins released from the cell are measured by standard methods. Specifically the metalloprotease activity is monitored via literature methods. The amount of metalloprotease released is correlated to the amount of serine protease used to treat the cells.

Increases, versus control, in src tyrosine kinase activity are measured by Western blot analysis of intracellular proteins using monoclonal antibodies specific for phosphotyrosine following cleavage and release of the disintegrin metalloprotease. Controls are cells that have not been treated with serine protease.

5 src tyrosine kinase activity in the cell (or is it cell culture) is measured by literature methods. Release of the metalloprotease domain of the disintegrin is also monitored via literature methods. There is a direct correlation between release of the metalloprotease domain and increases in intracellular src tyrosine kinase activity. This result is consistent with stimulation of disintegrin-mediated cell signaling by  
10 stimulation of the src tyrosine kinase cascade.

#### **Example 9**

Integrin binding is measured with a peptide containing the sequence RGD. Inhibition of intercellular adhesion molecules, or extracellular matrix components results in the inhibition of phenotypic changes, including changes in cell shape,  
15 associated with such interactions. Integrin binding is measured via competitive assay, using cellular changes in shape visible via microscopy. The peptide inhibits such cellular changes.

This result is consistent with competition with or blocking of the interaction of disintegrin. The RGD peptide inhibits cellular changes in chondrocytes. The  
20 osteoarthritis phenotype, characterized by increased matrix synthesis and accelerated matrix metalloprotease activity does not occur. Other readily assayable cellular changes can be used to monitor this result, including gene expression, changes in mitotic activity, and the like.

#### **Example 10**

25 A small molecular weight metalloprotease inhibitor is used to treat a tissue culture according to the method of Example 7. The release of TNF- $\alpha$  from the cell membrane is measured by literature methods. The inhibitor of Example 7 also decreases the amount of TNF- $\alpha$  secreted from the cell membrane.

Hence it is contemplated that inhibition of disintegrin metalloprotease activity  
30 will result in the inhibition of a disintegrin associated inflammation cascade and secretase activity. It is contemplated that monitoring the release of cytokines or IL-1 from the cell membrane, and the like, will produce the same result.

#### **Example 11**

##### **Differential Display Screening for Disease**

35 RNA is isolated from unstimulated and interleukin-1 stimulated cultures of normal human articular chondrocytes. The RNA is reverse transcribed into cDNA. The cDNA is subjected to amplification (PCR) using the above-named primers. PCR

samples generated from both stimulated and unstimulated chondrocytes are electrophoresed in adjacent lanes on polyacrylamide gels. A differentially expressed band (i.e., a band found only in the stimulated cells and not expressed at significant or detectable levels in the unstimulated cells) is excised from the gel, cloned, and partially sequenced. The partial sequence is shown in SEQ ID NO:5. the sequence is found to exhibit approximately 60% homology to a rat metalloprotease (see above). The sequence is found to exhibit approximately 85% homology to a human metalloprotease (see Gen Bank Accession #Z48597, see Figure 2).

#### **Example 12**

##### **Screening for Metastatic Potential of Tumors**

Cancer tissue is tested for metalloprotease gene expression. The above-named primers are used in PCR on extracted nucleic acid from the sample. High levels of transcripts suggest metastatic potential.

#### **Example 13**

##### **Drug Screen for Expression Inhibitors**

Candidate inhibitors of metalloprotease gene expression are screened *in vitro*. Interleukin-1 stimulated cultures of normal human articular chondrocytes are exposed *in vitro* to candidate inhibitors. The RNA is isolated and reverse transcribed into cDNA. the cDNA is subjected to amplification (PCR) using the above-named primers. PCR samples generated from both chondrocytes exposed to inhibitors and uninhibited chondrocytes are electrophoreses in adjacent lanes on polyacrylamide gels. Reduced levels of PCR product identifies an inhibitor.

#### **Example 14**

##### **Drug Screen For Metalloprotease Inhibitors**

Candidate inhibitors of the metalloprotease itself are screened *in vitro*. The culture supernatant of Interleukin-1 stimulated cultures of normal human articular chondrocytes are assayed on suitable metalloprotease substrates (e.g., matrix proteins) in the presence and absence of candidate inhibitors. Known inhibitors are used as controls (e.g., 1,10-phenanthroline available commercially from Sigma Co., St. Louis). Reduced levels of substrate (e.g., fluorogenic disintegrin metalloprotease substrate) degradation identifies an inhibitor.

#### **Example 15**

A 1400 BP clone is isolated via standard screening techniques from U-937, a monocyte-like cell cDNA line library. The initial sequence is a truncated clone, missing a portion of the 5' end. The 5' end is generated using 5' R.A.C.E. (Rapid Amplification of 5 c-DNA Ends, see for example, Chapter 4 (pages 28-38), and

references therein of PCR Protocols, A Guide to Methods and Applications, Innis, et al, eds. 1990 Academic Press), a known technique, generating a 1600 bp clone containing the remaining 5' sequence. These two sequences together provide SEQ ID NO:8, from which the peptide sequence is derived.

5

#### Example 16

Primers SEQ ID NO:9 (5'-AGCCTGTGTC-3') and SEQ ID NO:10 (5'-AGCCTGTGTCTGAACCACT-3') are used in differential display of mRNA (ddrd-PCR). 2-5 ng of sscDNA is used in the PCR. The reaction is precooled 0.2  $\mu$ l thin-walled tubes on ice. Each tube containing, 50mM TrisHCl (pH 8.5), 50mM KCl, 1.5 mM MgCl<sub>2</sub> 1mM of each dNTP, 2-5 ng of sscDNA, 10pmoles of each primer above, 05.  $\mu$ l of  $\alpha$ -P<sup>33</sup> dCTP (10  $\mu$ Ci/ $\mu$ l, Amersham) and water to 20  $\mu$ l. The mixture is subjected to 35 cycles of denaturation (94°C for 30 sec.), annealing (36°C for 30 sec.) and extension (72°C for 1 min.) using a Perkin-Elmer System 2400 Thermal Cycler (Perking-Elmer, Norwalk, CT).

15

By this method, IL-1 treated chondrocytes expressed the mRNA associated with this gene, while the untreated (no IL-1) control chondrocytes expressed no detectable mRNA.

#### Example 17

##### Assay system amenable to high throughput screening

20

The protease activity of disintegrin is measured in a kinetic enzyme inhibition assay using a fluorescent substrate. Using cloned disintegrin enzyme, and a small MW fluorescently labeled protein as the substrate. Enzyme activity is quantified by measurement of fluorescence after cleavage of the substrate molecule at room temperature. This assay simple and very easy to automate.

25

Using standard techniques, this assay is adapted to 96 or 384 well plates.

All references described herein are hereby incorporated by reference.

While particular embodiments of the subject invention have been described, it will be obvious to those skilled in the art that various changes and modifications of the subject invention can be made without departing from the spirit and scope of the invention. It is intended to cover, in the appended claims, all such modifications that are within the scope of this invention.

30

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

5

(i) APPLICANT: TINDAL, MICHAEL H  
HAQQI, TARIQ M

10

(ii) TITLE OF INVENTION: USE OF A NOVEL DISINTEGRIN  
METALLOPROTEASE, ITS MUTANTS, FRAGMENTS AND THE LIKE

(iii) NUMBER OF SEQUENCES: 11

## (iv) CORRESPONDENCE ADDRESS:

15

(A) ADDRESSEE: THE PROCTER & GAMBLE COMPANY  
(B) STREET: 8700 MASON-MONTGOMERY ROAD  
(C) CITY: MASON  
(D) STATE: OH  
(E) COUNTRY: USA  
(F) ZIP: 45040-9462

20

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

25

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:  
(B) FILING DATE:  
(C) CLASSIFICATION:

30

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: HAKE, RICHARD A.  
(B) REGISTRATION NUMBER: 37,143  
(C) REFERENCE/DOCKET NUMBER 59804

35

WO 98/37092

PCT/US98/03490

26

## (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 513/622-0087

(B) TELEFAX: 513/622-0270

5

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1824 base pairs

10

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

15

## (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 2..1477

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

C CAG ACC ACA GAC TTC TCC GGA ATC CGT AAC ATC AGT TTC ATG GTG 46  
25 Gln Thr Thr Asp Phe Ser Gly Ile Arg Asn Ile Ser Phe Met Val  
1 5 10 15

AAA CGC ATA AGA ATC AAT ACA ACT GCT GAT GAG AAG GAC CCT ACA AAT 94  
Lys Arg Ile Arg Ile Asn Thr Thr Ala Asp Glu Lys Asp Pro Thr Asn  
30 20 25 30

CCT TTC CGT TTC CCA AAT ATT AGT GTG GAG AAG TTT CTG GAA TTG AAT 142  
Pro Phe Arg Phe Pro Asn Ile Ser Val Glu Lys Phe Leu Glu Leu Asn  
35 40 45

35 TCT GAG CAG AAT CAT GAT GAC TAC TGT TTT GCC TAT GTC TTC ACA GAC 190  
Ser Glu Gln Asn His Asp Asp Tyr Cys Leu Ala Tyr Val Phe Thr Asp

WO 98/37092

PCT/US98/03490

27

|    |                                                                 |     |     |     |
|----|-----------------------------------------------------------------|-----|-----|-----|
|    | 50                                                              | 55  | 60  |     |
|    | CGA GAT TTT GAT GAT GGC GTA CTT GGT CTG GCT TGG GTT GGA GCA CCT | 238 |     |     |
|    | Arg Asp Phe Asp Asp Gly Val Leu Gly Leu Ala Trp Val Gly Ala Pro |     |     |     |
| 5  | 65                                                              | 70  | 75  |     |
|    | TCA GGA AGC TCT GGA GGA ATA TGT GAA AAA AGT AAA CTC TAT TCA GAT | 286 |     |     |
|    | Ser Gly Ser Ser Gly Gly Ile Cys Glu Lys Ser Lys Leu Tyr Ser Asp |     |     |     |
|    | 80                                                              | 85  | 90  | 95  |
| 10 |                                                                 |     |     |     |
|    | GGT AAG AAG AAG TCC TTA AAC ACT GGA ATT ATT ACT GTT CAG AAC TAT | 334 |     |     |
|    | Gly Lys Lys Lys Ser Leu Asn Thr Gly Ile Ile Thr Val Gln Asn Tyr |     |     |     |
|    | 100                                                             | 105 | 110 |     |
| 15 |                                                                 |     |     |     |
|    | GGG TCT CAT GTA CCT CCC AAA GTC TCT CAC ATT ACT TTT GCT CAC GAA | 382 |     |     |
|    | Gly Ser His Val Pro Pro Lys Val Ser His Ile Thr Phe Ala His Glu |     |     |     |
|    | 115                                                             | 120 | 125 |     |
|    | GTT GGA CAT AAC TTT GGA TCC CCA CAT GAT TCT GGA ACA GAG TGC ACA | 430 |     |     |
| 20 | Val Gly His Asn Phe Gly Ser Pro His Asp Ser Gly Thr Glu Cys Thr |     |     |     |
|    | 130                                                             | 135 | 140 |     |
|    | CCA GGA GAA TCT AAG AAT TTG GGT CAA AAA GAA AAT GGC AAT TAC ATC | 478 |     |     |
|    | Pro Gly Glu Ser Lys Asn Leu Gly Gln Lys Glu Asn Gly Asn Tyr Ile |     |     |     |
| 25 | 145                                                             | 150 | 155 |     |
|    | ATG TAT GCA AGA GCA ACA TCT GGG GAC AAA CTT AAC AAC AAT AAA TTC | 526 |     |     |
|    | Met Tyr Ala Arg Ala Thr Ser Gly Asp Lys Leu Asn Asn Asn Lys Phe |     |     |     |
|    | 160                                                             | 165 | 170 | 175 |
| 30 |                                                                 |     |     |     |
|    | TCA CTC TGT AGT ATT AGA AAT ATA AGC CAA GTT CTT GAG AAG AAG AGA | 574 |     |     |
|    | Ser Leu Cys Ser Ile Arg Asn Ile Ser Gln Val Leu Glu Lys Lys Arg |     |     |     |
|    | 180                                                             | 185 | 190 |     |
| 35 |                                                                 |     |     |     |
|    | AAC AAC TGT TTT GTT GAA TCT GGC CAA CCT ATT TGT GGA AAT GGA ATG | 622 |     |     |
|    | Asn Asn Cys Phe Val Glu Ser Gly Gln Pro Ile Cys Gly Asn Gly Met |     |     |     |
|    | 195                                                             | 200 | 205 |     |

WO 98/37092

PCT/US98/03490

28

|    |                                                                 |      |
|----|-----------------------------------------------------------------|------|
|    | GTA GAA CAA GGT GAA GAA TGT GAT TGT GGC TAT AGT GAC CAG TGT AAA | 670  |
|    | Val Glu Gln Gly Glu Glu Cys Asp Cys Gly Tyr Ser Asp Gln Cys Lys |      |
|    | 210 215 220                                                     |      |
| 5  | GAT GAA TGC TGC TTC GAT GCA AAT CAA CCA GAG GGA AGA AAA TGC AAA | 718  |
|    | Asp Glu Cys Cys Phe Asp Ala Asn Gln Pro Glu Gly Arg Lys Cys Lys |      |
|    | 225 230 235                                                     |      |
| 10 | CTG AAA CCT GGG AAA CAG TGC AGT CCA AGT CAA GGT CCT TGT TGT ACA | 766  |
|    | Leu Lys Pro Gly Lys Gln Cys Ser Pro Ser Gln Gly Pro Cys Cys Thr |      |
|    | 240 245 250 255                                                 |      |
|    | GCA CAG TGT GCA TTC AAG TCA AAG TCT GAG AAG TGT CGG GAT GAT TCA | 814  |
| 15 | Ala Gln Cys Ala Phe Lys Ser Lys Ser Glu Lys Cys Arg Asp Asp Ser |      |
|    | 260 265 270                                                     |      |
|    | GAC TGT GCA AGG GAA GGA ATA TGT AAT GGC TTC ACA GCT CTC TGC CCA | 862  |
|    | Asp Cys Ala Arg Glu Gly Ile Cys Asn Gly Phe Thr Ala Leu Cys Pro |      |
| 20 | 275 280 285                                                     |      |
|    | GCA TCT GAC CCT AAA CCA AAC TTC ACA GAC TGT AAT AGG CAT ACA CAA | 910  |
|    | Ala Ser Asp Pro Lys Pro Asn Phe Thr Asp Cys Asn Arg His Thr Gln |      |
|    | 290 295 300                                                     |      |
| 25 | GTG TGC ATT AAT GGG CAA TGT GCA GGT TCT ATC TGT GAG AAA TAT GGC | 958  |
|    | Val Cys Ile Asn Gly Gln Cys Ala Gly Ser Ile Cys Glu Lys Tyr Gly |      |
|    | 305 310 315                                                     |      |
| 30 | TTA GAG GAG TGT ACG TGT GCC AGT TCT GAT GGC AAA GAT GAT AAA GAA | 1006 |
|    | Leu Glu Glu Cys Thr Cys Ala Ser Ser Asp Gly Lys Asp Asp Lys Glu |      |
|    | 320 325 330 335                                                 |      |
|    | TTA TGC CAT GTA TGC TGT ATG AAG AAA ATG GAC CCA TCA ACT TGT GCC | 1054 |
| 35 | Leu Cys His Val Cys Cys Met Lys Lys Met Asp Pro Ser Thr Cys Ala |      |
|    | 340 345 350                                                     |      |

WO 98/37092

PCT/US98/03490

29

|    |                                                                  |      |
|----|------------------------------------------------------------------|------|
|    | AGT ACA GGG TCT GTG CAG TGG AGT AGG CAC TTC AGT GGT CGA ACC ATC  | 1102 |
|    | Ser Thr Gly Ser Val Gln Trp Ser Arg His Phe Ser Gly Arg Thr Ile  |      |
|    | 355 360 365                                                      |      |
| 5  | ACC CTG CAA CCT GGA TCC CCT TGC AAC GAT TTT AGA GGT TAC TGT GAT  | 1150 |
|    | Thr Leu Gln Pro Gly Ser Pro Cys Asn Asp Phe Arg Gly Tyr Cys Asp  |      |
|    | 370 375 380                                                      |      |
|    | GTT TTC ATG CGG TGC AGA TTA GTA GAT GCT GAT GGT CCT CTA GCT AGG  | 1198 |
| 10 | Val Phe Met Arg Cys Arg Leu Val Asp Ala Asp Gly Pro Leu Ala Arg  |      |
|    | 385 390 395                                                      |      |
|    | CTT AAA AAA GCA ATT TTT AGT CCA GAG CTC TAT GAA AAC ATT GCT GAA  | 1246 |
|    | Leu Lys Lys Ala Ile Phe Ser Pro Glu Leu Tyr Glu Asn Ile Ala Glu  |      |
| 15 | 400 405 410 415                                                  |      |
|    | TGG ATT GTG GCT CAT TGG TGG GCA GTA TTA CTT ATG GGA ATT GCT CTG  | 1294 |
|    | Trp Ile Val Ala His Trp Trp Ala Val Leu Leu Met Gly Ile Ala Leu  |      |
|    | 420 425 430                                                      |      |
| 20 | ATC ATG CTA ATG GCT GGA TTT ATT AAG ATA TGC AGT GTT CAT ACT CCA  | 1342 |
|    | Ile Met Leu Met Ala Gly Phe Ile Lys Ile Cys Ser Val His Thr Pro  |      |
|    | 435 440 445                                                      |      |
| 25 | AGT AGT AAT CCA AAG TTG CCT CCT CCT AAA CCA CTT CCA GGC ACT TTA  | 1390 |
|    | Ser Ser Asn Pro Lys Leu Pro Pro Pro Lys Pro Leu Pro Gly Thr Leu  |      |
|    | 450 455 460                                                      |      |
|    | AAG AGG AGG AGA CCT CCA CAG CCC ATT CAG CAA CCC CAG CGT CAG CGG  | 1438 |
| 30 | Lys Arg Arg Arg Pro Pro Gln Pro Ile Gln Gln Pro Gln Arg Gln Arg  |      |
|    | 465 470 475                                                      |      |
|    | CCC CGA GAG AGT TAT CAA ATG GGA CAC ATG AGA CGC TAA CTGCAGCTTT   | 1487 |
|    | Pro Arg Glu Ser Tyr Gln Met Gly His Met Arg Arg *                |      |
| 35 | 480 485 490                                                      |      |
|    | TGCCTTGGTT CTCCTAGTG CCTACAATGG GAAAAATTCA CTCCAAAGAG AAACCTATTA | 1547 |

WO 98/37092

PCT/US98/03490

30

AGTCATCATC TCCAAACTAA ACCCTCACAA GTAACAGTTG AAGAAAAAAT GGCAAGAGAT 1607

CATATCCTCA GACCAGGTGG AATTACTTAA ATTTTAAAGC CTGAAAATTC CAATTTGGGG 1667

5 GTGGGAGGTG GAAAAGGAAC CCAATTTTCT TATGAACAGA TATTTTAAAC TTAATGGCAC 1727

AAAGTCTTAG AATATTATTA TGTGCCCCGT GTTCCCTGTT CTCGTTGCT GCATTTTCTT 1787

10 CACTTGCAGG CAAACTTGGC TCTCAATAAA CTTTTCG 1824

(2) INFORMATION FOR SEQ ID NO:2:

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 492 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Gln Thr Thr Asp Phe Ser Gly Ile Arg Asn Ile Ser Phe Met Val Lys  
25 1 5 10 15

Arg Ile Arg Ile Asn Thr Thr Ala Asp Glu Lys Asp Pro Thr Asn Pro  
20 25 30

30 Phe Arg Phe Pro Asn Ile Ser Val Glu Lys Phe Leu Glu Leu Asn Ser  
35 40 45

Glu Gln Asn His Asp Asp Tyr Cys Leu Ala Tyr Val Phe Thr Asp Arg  
50 55 60

35 Asp Phe Asp Asp Gly Val Leu Gly Leu Ala Trp Val Gly Ala Pro Ser  
65 70 75 80

WO 98/37092

PCT/US98/03490

31

Gly Ser Ser Gly Gly Ile Cys Glu Lys Ser Lys Leu Tyr Ser Asp Gly  
 85 90 95

5 Lys Lys Lys Ser Leu Asn Thr Gly Ile Ile Thr Val Gln Asn Tyr Gly  
 100 105 110

Ser His Val Pro Pro Lys Val Ser His Ile Thr Phe Ala His Glu Val  
 115 120 125

10 Gly His Asn Phe Gly Ser Pro His Asp Ser Gly Thr Glu Cys Thr Pro  
 130 135 140

Gly Glu Ser Lys Asn Leu Gly Gln Lys Glu Asn Gly Asn Tyr Ile Met  
 15 145 150 155 160

Tyr Ala Arg Ala Thr Ser Gly Asp Lys Leu Asn Asn Asn Lys Phe Ser  
 165 170 175

20 Leu Cys Ser Ile Arg Asn Ile Ser Gln Val Leu Glu Lys Lys Arg Asn  
 180 185 190

Asn Cys Phe Val Glu Ser Gly Gln Pro Ile Cys Gly Asn Gly Met Val  
 195 200 205

25 Glu Gln Gly Glu Glu Cys Asp Cys Gly Tyr Ser Asp Gln Cys Lys Asp  
 210 215 220

Glu Cys Cys Phe Asp Ala Asn Gln Pro Glu Gly Arg Lys Cys Lys Leu  
 30 225 230 235 240

Lys Pro Gly Lys Gln Cys Ser Pro Ser Gln Gly Pro Cys Cys Thr Ala  
 245 250 255

35 Gln Cys Ala Phe Lys Ser Lys Ser Glu Lys Cys Arg Asp Asp Ser Asp  
 260 265 270

WO 98/37092

PCT/US98/03490

32

Cys Ala Arg Glu Gly Ile Cys Asn Gly Phe Thr Ala Leu Cys Pro Ala  
 275 280 285

Ser Asp Pro Lys Pro Asn Phe Thr Asp Cys Asn Arg His Thr Gln Val  
 5 290 295 300

Cys Ile Asn Gly Gln Cys Ala Gly Ser Ile Cys Glu Lys Tyr Gly Leu  
 305 310 315 320

10 Glu Glu Cys Thr Cys Ala Ser Ser Asp Gly Lys Asp Asp Lys Glu Leu  
 325 330 335

Cys His Val Cys Cys Met Lys Lys Met Asp Pro Ser Thr Cys Ala Ser  
 340 345 350

15 Thr Gly Ser Val Gln Trp Ser Arg His Phe Ser Gly Arg Thr Ile Thr  
 355 360 365

Leu Gln Pro Gly Ser Pro Cys Asn Asp Phe Arg Gly Tyr Cys Asp Val  
 20 370 375 380

Phe Met Arg Cys Arg Leu Val Asp Ala Asp Gly Pro Leu Ala Arg Leu  
 385 390 395 400

25 Lys Lys Ala Ile Phe Ser Pro Glu Leu Tyr Glu Asn Ile Ala Glu Trp  
 405 410 415

Ile Val Ala His Trp Trp Ala Val Leu Leu Met Gly Ile Ala Leu Ile  
 420 425 430

30 Met Leu Met Ala Gly Phe Ile Lys Ile Cys Ser Val His Thr Pro Ser  
 435 440 445

Ser Asn Pro Lys Leu Pro Pro Pro Lys Pro Leu Pro Gly Thr Leu Lys  
 35 450 455 460

Arg Arg Arg Pro Pro Gln Pro Ile Gln Gln Pro Gln Arg Gln Arg Pro

Gln Lys His Gln Arg Ala Lys Arg Ala Val Ser His Glu Asp Gln Phe

WO 98/37092

PCT/US98/03490

34

|    |                                                                 |     |     |     |
|----|-----------------------------------------------------------------|-----|-----|-----|
|    | 540                                                             | 545 | 550 |     |
|    | TTA CGT CTA GAT TTC CAT GCC CAT GGA AGA CAT TTC AAC CTA CGA ATG |     |     | 241 |
|    | Leu Arg Leu Asp Phe His Ala His Gly Arg His Phe Asn Leu Arg Met |     |     |     |
| 5  | 555                                                             | 560 | 565 |     |
|    | AAG AGG GAC ACT TCC CTT TTC AGT GAT GAA TTT AAA GTA GAA ACA TCA |     |     | 289 |
|    | Lys Arg Asp Thr Ser Leu Phe Ser Asp Glu Phe Lys Val Glu Thr Ser |     |     |     |
|    | 570                                                             | 575 | 580 |     |
| 10 |                                                                 |     |     |     |
|    | AAT AAA GTA CTT GAT TAT GAT ACC TCT CAT ATT TAC ACT GGA CAT ATT |     |     | 337 |
|    | Asn Lys Val Leu Asp Tyr Asp Thr Ser His Ile Tyr Thr Gly His Ile |     |     |     |
|    | 585                                                             | 590 | 595 |     |
| 15 |                                                                 |     |     |     |
|    | TAT GGT GAA GAA GGA AGT TTT AGC CAT GGG TCT GTT ATT GAT GGA AGA |     |     | 385 |
|    | Tyr Gly Glu Glu Gly Ser Phe Ser His Gly Ser Val Ile Asp Gly Arg |     |     |     |
|    | 600                                                             | 605 | 610 | 615 |
|    | TTT GAA GGA TTC ATC CAG ACT CGT GGT GGC ACA TTT TAT GTT GAG CCA |     |     | 433 |
| 20 | Phe Glu Gly Phe Ile Gln Thr Arg Gly Gly Thr Phe Tyr Val Glu Pro |     |     |     |
|    | 620                                                             | 625 | 630 |     |
|    | GCA GAG AGA TAT ATT AAA GAC CGA ACT CTG CCA TTT CAC TCT GTC ATT |     |     | 481 |
|    | Ala Glu Arg Tyr Ile Lys Asp Arg Thr Leu Pro Phe His Ser Val Ile |     |     |     |
| 25 | 635                                                             | 640 | 645 |     |
|    | TAT CAT GAA GAT GAT ATT AGT GAA AGG CTT AAA CTG AGG CTT AGA AAA |     |     | 529 |
|    | Tyr His Glu Asp Asp Ile Ser Glu Arg Leu Lys Leu Arg Leu Arg Lys |     |     |     |
|    | 650                                                             | 655 | 660 |     |
| 30 |                                                                 |     |     |     |
|    | CTT ATG TCA CTT GAG TTG TGG ACC TCC TGT TGT TTA CCC TGT GCT CTT |     |     | 577 |
|    | Leu Met Ser Leu Glu Leu Trp Thr Ser Cys Cys Leu Pro Cys Ala Leu |     |     |     |
|    | 665                                                             | 670 | 675 |     |
|    | CTG CTT CAC TCA TGG AAG AAA GCT GTA AAT TCT CAC TGC CTT TAC TTC |     |     | 625 |
| 35 | Leu Leu His Ser Trp Lys Lys Ala Val Asn Ser His Cys Leu Tyr Phe |     |     |     |
|    | 680                                                             | 685 | 690 | 695 |

WO 98/37092

PCT/US98/03490

35

AAG GAT TTC TGG GGC TTT TCT GAA ATC TAC TAT CCC CAT AAA TAC GGT 673  
 Lys Asp Phe Trp Gly Phe Ser Glu Ile Tyr Tyr Pro His Lys Tyr Gly  
 700 705 710

5

CCT CAG GGC GGC TGT GCA GAT CAT TCA GTA TTT GAA AGA ATG AGG AAA 721  
 Pro Gln Gly Gly Cys Ala Asp His Ser Val Phe Glu Arg Met Arg Lys  
 715 720 725

10

TAC CAG ATG ACT GGT GTA GAG GAA GTA ACA CAG ATA CCT CAA GAA GAA 769  
 Tyr Gln Met Thr Gly Val Glu Glu Val Thr Gln Ile Pro Gln Glu Glu  
 730 735 740

15

CAT GCT GCT AAT GGT CCA GAA CTT CTG AGG AAA AGA CGT ACA ACT TCA 817  
 His Ala Ala Asn Gly Pro Glu Leu Leu Arg Lys Arg Arg Thr Thr Ser  
 745 750 755

20

GCT GAA AAA AAT ACT TGT CAG CTT TAT ATT CAG ACT GAT CAT TTG TTC 865  
 Ala Glu Lys Asn Thr Cys Gln Leu Tyr Ile Gln Thr Asp His Leu Phe  
 760 765 770 775

25

TTT AAA TAT TAC GGA ACA CGA GAA GCT GTG ATT GCC CAG ATA TCC AGT 913  
 Phe Lys Tyr Tyr Gly Thr Arg Glu Ala Val Ile Ala Gln Ile Ser Ser  
 780 785 790

30

CAT GTT AAA GCG ATT GAT ACA ATT TAC CAG ACC ACA GAC TTC TCC GGA 961  
 His Val Lys Ala Ile Asp Thr Ile Tyr Gln Thr Thr Asp Phe Ser Gly  
 795 800 805

35

ATC CGT AAC ATC AGT TTC ATG GTG AAA GCG ATA AGA ATC AAT ACA ACT 1009  
 Ile Arg Asn Ile Ser Phe Met Val Lys Arg Ile Arg Ile Asn Thr Thr  
 810 815 820

GCT GAT GAG AAG GAC CCT ACA AAT CCT TTT GTT TTC CCA AAT ATT AGT 1057  
 Ala Asp Glu Lys Asp Pro Thr Asn Pro Phe Arg Phe Pro Asn Ile Ser  
 825 830 835

WO 98/37092

PCT/US98/03490

36

|    |                                                                                |      |
|----|--------------------------------------------------------------------------------|------|
|    | GTG GAG AAG TTT CTG GAA TTG AAT TCT GAG CAG AAT CAT GAT GAC TAC                | 1105 |
|    | Val Glu Lys Phe Leu Glu Leu Asn Ser Glu Gln Asn His Asp Asp Tyr                |      |
|    | 840                      845                      850                      855 |      |
| 5  | TGT TTG GCC TAT GTC TTC ACA GAC CGA GAT TTT GAT GAT GGC GTA CTT                | 1153 |
|    | Cys Leu Ala Tyr Val Phe Thr Asp Arg Asp Phe Asp Asp Gly Val Leu                |      |
|    | 860                      865                      870                          |      |
|    | GGT CTG GCT TGG GTT GGA GCA CCT TCA GGA AGC TCT GGA GGA ATA TGT                | 1201 |
| 10 | Gly Leu Ala Trp Val Gly Ala Pro Ser Gly Ser Ser Gly Gly Ile Cys                |      |
|    | 875                      880                      885                          |      |
|    | GAA AAA AGT AAA CTC TAT TCA GAT GGT AAG AAG AAG TCC TTA AAC ACT                | 1249 |
|    | Glu Lys Ser Lys Leu Tyr Ser Asp Gly Lys Lys Lys Ser Leu Asn Thr                |      |
| 15 | 890                      895                      900                          |      |
|    | GGA ATT ATT ACT GTT CAG AAC TAT GGG TCT CAT GTA CCT CCC AAA GTC                | 1297 |
|    | Gly Ile Ile Thr Val Gln Asn Tyr Gly Ser His Val Pro Pro Lys Val                |      |
|    | 905                      910                      915                          |      |
| 20 |                                                                                |      |
|    | TCT CAC ATT ACT TTT GCT CAC GAA GTT GGA CAT AAC TTT GGA TCC CCA                | 1345 |
|    | Ser His Ile Thr Phe Ala His Glu Val Gly His Asn Phe Gly Ser Pro                |      |
|    | 920                      925                      930                      935 |      |
| 25 | CAT GAT TCT GGA ACA GAG TGC ACA CCA GGA GAA TCT AAG AAT TTG GGT                | 1393 |
|    | His Asp Ser Gly Thr Glu Cys Thr Pro Gly Glu Ser Lys Asn Leu Gly                |      |
|    | 940                      945                      950                          |      |
|    | CAA AAA GAA AAT GGC AAT TAC ATC ATG TAT GCA AGA GCA ACA TCT GGC                | 1441 |
| 30 | Gln Lys Glu Asn Gly Asn Tyr Ile Met Tyr Ala Arg Ala Thr Ser Gly                |      |
|    | 955                      960                      965                          |      |
|    | GAC AAA CTT AAC AAC AAT AAA TTC TCA TTT TGT AGT ATT AGA AAT ATA                | 1489 |
|    | Asp Lys Leu Asn Asn Asn Lys Phe Ser Leu Cys Ser Ile Arg Asn Ile                |      |
| 35 | 970                      975                      980                          |      |
|    | AGC CAA GTT CTT GAG AAG AAG AGA AAC AAT TGT TTT GTT GAA TCT GGC                | 1537 |

WO 98/37092

PCT/US98/03490

37

|    |                                                                                                                                    |      |
|----|------------------------------------------------------------------------------------------------------------------------------------|------|
|    | Ser Gln Val Leu Glu Lys Lys Arg Asn Asn Cys Phe Val Glu Ser Gly                                                                    |      |
|    | 985                                      990                                      995                                              |      |
|    | CAA CCT ATT TGT GGA AAT GGA ATG GTA GAA CAA GGT GAA GAA TGT GAT                                                                    | 1585 |
| 5  | Gln Pro Ile Cys Gly Asn Gly Met Val Glu Gln Gly Glu Glu Cys Asp                                                                    |      |
|    | 1000                                      1005                                      1010                                      1015 |      |
|    | TGT GGC TAT AGT GAC CAG TGT AAA GAT GAA TGC TGC TTC GAT GCA AAT                                                                    | 1633 |
|    | Cys Gly Tyr Ser Asp Gln Cys Lys Asp Glu Cys Cys Phe Asp Ala Asn                                                                    |      |
| 10 | 1020                                      1025                                      1030                                           |      |
|    | CAA CCA GAG GGA AGA AAA TGC AAA CTG AAA CCT GGG AAA CAG TGC AGT                                                                    | 1681 |
|    | Gln Pro Glu Gly Arg Lys Cys Lys Leu Lys Pro Gly Lys Gln Cys Ser                                                                    |      |
|    | 1035                                      1040                                      1045                                           |      |
| 15 |                                                                                                                                    |      |
|    | CCA AGT CAA GGT CCT TGT TGT ACA GCA CAG TGT GCA TTC AAG TCA AAG                                                                    | 1729 |
|    | Pro Ser Gln Gly Pro Cys Cys Thr Ala Gln Cys Ala Phe Lys Ser Lys                                                                    |      |
|    | 1050                                      1055                                      1060                                           |      |
| 20 |                                                                                                                                    |      |
|    | TCT GAG AAG TGT CGG GAT GAT TCA GAC TGT GCA AGG GAA GGA ATA TGT                                                                    | 1777 |
|    | Ser Glu Lys Cys Arg Asp Asp Ser Asp Cys Ala Arg Glu Gly Ile Cys                                                                    |      |
|    | 1065                                      1070                                      1075                                           |      |
|    | AAT GGC TTC ACA GCT CTC TGC CCA GCA TCT GAC CCT AAA CCA AAC TTC                                                                    | 1825 |
| 25 | Asn Gly Phe Thr Ala Leu Cys Pro Ala Ser Asp Pro Lys Pro Asn Phe                                                                    |      |
|    | 1080                                      1085                                      1090                                      1095 |      |
|    | ACA GAC TGT AAT AGG CAT ACA CAA GTG TGC ATT AAT GGG CAA TGT GCA                                                                    | 1873 |
|    | Thr Asp Cys Asn Arg His Thr Gln Val Cys Ile Asn Gly Gln Cys Ala                                                                    |      |
| 30 | 1100                                      1105                                      1110                                           |      |
|    | GGT TCT ATC TGT GAG AAA TAT GGC TTA GAG GAG TGT ACG TGT GCC AGT                                                                    | 1921 |
|    | Gly Ser Ile Cys Glu Lys Tyr Gly Leu Glu Glu Cys Thr Cys Ala Ser                                                                    |      |
|    | 1115                                      1120                                      1125                                           |      |
| 35 |                                                                                                                                    |      |
|    | TCT GAT GGC AAA GAT GAT AAA GAA TTA TGC CAT GTA TGC TGT ATG AAG                                                                    | 1969 |
|    | Ser Asp Gly Lys Asp Asp Lys Glu Leu Cys His Val Cys Cys Met Lys                                                                    |      |

WO 98/37092

PCT/US98/03490

38

|    | 1130                                                            | 1135 | 1140 |      |
|----|-----------------------------------------------------------------|------|------|------|
|    | AAA ATG GAC CCA TCA ACT TGT GCC AGT ACA GGG TCT GTG CAG TGG AGT |      |      | 2017 |
|    | Lys Met Asp Pro Ser Thr Cys Ala Ser Thr Gly Ser Val Gln Trp Ser |      |      |      |
| 5  | 1145                                                            | 1150 | 1155 |      |
|    | AGG CAC TTC AGT GGT CGA ACC ATC ACC CTG CAA CCT GGA TCC CCT TGC |      |      | 2065 |
|    | Arg His Phe Ser Gly Arg Thr Ile Thr Leu Gln Pro Gly Ser Pro Cys |      |      |      |
|    | 1160                                                            | 1165 | 1170 | 1175 |
| 10 |                                                                 |      |      |      |
|    | AAC GAT TTT AGA GGT TAC TGT GAT GTT TTC ATG CGG TGC AGA TTA GTA |      |      | 2113 |
|    | Asn Asp Phe Arg Gly Tyr Cys Asp Val Phe Met Arg Cys Arg Leu Val |      |      |      |
|    |                                                                 | 1180 | 1185 | 1190 |
| 15 |                                                                 |      |      |      |
|    | GAT GCT GAT GGT CCT CTA GCT AGG CTT AAA AAA GCA ATT TTT AGT CCA |      |      | 2161 |
|    | Asp Ala Asp Gly Pro Leu Ala Arg Leu Lys Lys Ala Ile Phe Ser Pro |      |      |      |
|    |                                                                 | 1195 | 1200 | 1205 |
|    | GAG CTC TAT GAA AAC ATT GCT GAA TGG ATT GTG GCT CAT TGG TGG GCA |      |      | 2209 |
| 20 | Glu Leu Tyr Glu Asn Ile Ala Glu Trp Ile Val Ala His Trp Trp Ala |      |      |      |
|    |                                                                 | 1210 | 1215 | 1220 |
|    | GTA TTA CTT ATG GGA ATT GCT CTG ATC ATG CTA ATG GCT GGA TTT ATT |      |      | 2257 |
|    | Val Leu Leu Met Gly Ile Ala Leu Ile Met Leu Met Ala Gly Phe Ile |      |      |      |
| 25 |                                                                 |      |      |      |
|    | 1225                                                            | 1230 | 1235 |      |
|    | AAG ATA TGC AGT GTT CAT ACT CCA AGT AGT AAT CCA AAG TTG CCT CCT |      |      | 2305 |
|    | Lys Ile Cys Ser Val His Thr Pro Ser Ser Asn Pro Lys Leu Pro Pro |      |      |      |
|    | 1240                                                            | 1245 | 1250 | 1255 |
| 30 |                                                                 |      |      |      |
|    | CCT AAA CCA CTT CCA GGC ACT TTA AAG AGG AGG AGA CCT CCA CAG CCC |      |      | 2353 |
|    | Pro Lys Pro Leu Pro Gly Thr Leu Lys Arg Arg Arg Pro Pro Gln Pro |      |      |      |
|    |                                                                 | 1260 | 1265 | 1270 |
| 35 |                                                                 |      |      |      |
|    | ATT CAG CAA CCC CAG CGT CAG CGG CCC CCA CAG AGT TAT CAA ATG GGA |      |      | 2401 |
|    | Ile Gln Gln Pro Gln Arg Gln Arg Pro Arg Glu Ser Tyr Gln Met Gly |      |      |      |
|    |                                                                 | 1275 | 1280 | 1285 |

WO 98/37092

PCT/US98/03490

39

CAC ATG AGA CGC T AACTGCAGCT TTTGCCTTGG TTCTTCCTAG TGCCTACAAT 2454  
His Met Arg Arg  
1290

5

GGGAAACTT CACTCCAAAG AGAAACCTAT TAAGTCATCA TCTCCAACT AAACCCTCAC 2514

AAGTAACAGT TGAAGAAAAA ATGGCAAGAG ATCATATCCT CAGACCAGGT GGAATTACTT 2574

10 AAATTTTAAA GCCTGAAAAT TCCAATTGG GGGTGGGAGG TGGAAAAGGA ACCCAATTTT 2634

CTTATGAACA GATATTTTTA ACTTAATGGC ACAAAGTCTT AGAATATTAT TATGTGCCCC 2694

GTGTTCCCTG TTCTTCGTTG CTGCATTTTC TTCACTTGCA GGCAAACCTG GCTCTCAATA 2754

15 AACTTTTCG 2763

(2) INFORMATION FOR SEQ ID NO:4:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 799 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

30

Met Val Leu Leu Arg Val Leu Ile Leu Leu Leu Ser Trp Ala Ala Gly  
1 5 10 15

Met Gly Gly Gln Tyr Gly Asn Pro Leu Asn Lys Tyr Ile Arg His Tyr  
20 25 30

35

Glu Gly Leu Ser Tyr Asn Val Asp Ser Leu His Gln Lys His Gln Arg  
35 40 45

WO 98/37092

PCT/US98/03490

40

Ala Lys Arg Ala Val Ser His Glu Asp Gln Phe Leu Arg Leu Asp Phe  
50 55 60

5 His Ala His Gly Arg His Phe Asn Leu Arg Met Lys Arg Asp Thr Ser  
65 70 75 80

Leu Phe Ser Asp Glu Phe Lys Val Glu Thr Ser Asn Lys Val Leu Asp  
85 90 95

10 Tyr Asp Thr Ser His Ile Tyr Thr Gly His Ile Tyr Gly Glu Glu Gly  
100 105 110

Ser Phe Ser His Gly Ser Val Ile Asp Gly Arg Phe Glu Gly Phe Ile  
15 115 120 125

Gln Thr Arg Gly Gly Thr Phe Tyr Val Glu Pro Ala Glu Arg Tyr Ile  
130 135 140

20 Lys Asp Arg Thr Leu Pro Phe His Ser Val Ile Tyr His Glu Asp Asp  
145 150 155 160

Ile Ser Glu Arg Leu Lys Leu Arg Leu Arg Lys Leu Met Ser Leu Glu  
165 170 175

25 Leu Trp Thr Ser Cys Cys Leu Pro Cys Ala Leu Leu Leu His Ser Trp  
180 185 190

Lys Lys Ala Val Asn Ser His Cys Leu Tyr Phe Lys Asp Phe Trp Gly  
30 195 200 205

Phe Ser Glu Ile Tyr Tyr Pro His Lys Tyr Gly Pro Gln Gly Gly Cys  
210 215 220

35 Ala Asp His Ser Val Phe Glu Arg Met Asn Lys Tyr Gln Met Thr Gly  
225 230 235 240

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
|    | Val | Glu | Glu | Val | Thr | Gln | Ile | Pro | Gln | Glu | Glu | His | Ala | Ala | Asn | Gly |  |
|    |     |     |     |     |     | 245 |     |     |     | 250 |     |     |     |     |     | 255 |  |
| 5  | Pro | Glu | Leu | Leu | Arg | Lys | Arg | Arg | Thr | Thr | Ser | Ala | Glu | Lys | Asn | Thr |  |
|    |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     |     | 270 |     |  |
|    | Cys | Gln | Leu | Tyr | Ile | Gln | Thr | Asp | His | Leu | Phe | Phe | Lys | Tyr | Tyr | Gly |  |
|    |     |     |     | 275 |     |     |     | 280 |     |     |     |     |     | 285 |     |     |  |
| 10 | Thr | Arg | Glu | Ala | Val | Ile | Ala | Gln | Ile | Ser | Ser | His | Val | Lys | Ala | Ile |  |
|    |     |     |     |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |  |
|    | Asp | Thr | Ile | Tyr | Gln | Thr | Thr | Asp | Phe | Ser | Gly | Ile | Arg | Asn | Ile | Ser |  |
|    | 305 |     |     |     |     | 310 |     |     |     |     |     | 315 |     |     |     | 320 |  |
| 15 | Phe | Met | Val | Lys | Arg | Ile | Arg | Ile | Asn | Thr | Thr | Ala | Asp | Glu | Lys | Asp |  |
|    |     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     |     | 335 |  |
|    | Pro | Thr | Asn | Pro | Phe | Arg | Phe | Pro | Asn | Ile | Ser | Val | Glu | Lys | Phe | Leu |  |
| 20 |     |     |     | 340 |     |     |     |     | 345 |     |     |     |     |     | 350 |     |  |
|    | Glu | Leu | Asn | Ser | Glu | Gln | Asn | His | Asp | Asp | Tyr | Cys | Leu | Ala | Tyr | Val |  |
|    |     |     | 355 |     |     |     |     | 360 |     |     |     |     |     | 365 |     |     |  |
| 25 | Phe | Thr | Asp | Arg | Asp | Phe | Asp | Asp | Gly | Val | Leu | Gly | Leu | Ala | Trp | Val |  |
|    |     |     | 370 |     |     |     | 375 |     |     |     |     |     | 380 |     |     |     |  |
|    | Gly | Ala | Pro | Ser | Gly | Ser | Ser | Gly | Gly | Ile | Cys | Glu | Lys | Ser | Lys | Leu |  |
|    | 385 |     |     |     |     | 390 |     |     |     |     |     | 395 |     |     |     | 400 |  |
| 30 | Tyr | Ser | Asp | Gly | Lys | Lys | Lys | Ser | Leu | Asn | Thr | Gly | Ile | Ile | Thr | Val |  |
|    |     |     |     |     | 405 |     |     |     |     |     | 410 |     |     |     |     | 415 |  |
|    | Gln | Asn | Tyr | Gly | Ser | His | Val | Pro | Pro | Lys | Val | Ser | His | Ile | Thr | Phe |  |
| 35 |     |     |     | 420 |     |     |     |     | 425 |     |     |     |     |     |     | 430 |  |
|    | Ala | His | Glu | Val | Gly | His | Asn | Phe | Gly | Ser | Pro | His | Asp | Ser | Gly | Thr |  |

WO 98/37092

PCT/US98/03490

42

|    |                                                                 |     |     |     |         |
|----|-----------------------------------------------------------------|-----|-----|-----|---------|
|    | 435                                                             |     | 440 |     | 445     |
|    | Glu Cys Thr Pro Gly Glu Ser Lys Asn Leu Gly Gln Lys Glu Asn Gly |     |     |     |         |
|    | 450                                                             |     | 455 |     | 460     |
| 5  | Asn Tyr Ile Met Tyr Ala Arg Ala Thr Ser Gly Asp Lys Leu Asn Asn |     |     |     |         |
|    | 465                                                             |     | 470 |     | 475 480 |
|    | Asn Lys Phe Ser Leu Cys Ser Ile Arg Asn Ile Ser Gln Val Leu Glu |     |     |     |         |
| 10 |                                                                 | 485 |     | 490 | 495     |
|    | Lys Lys Arg Asn Asn Cys Phe Val Glu Ser Gly Gln Pro Ile Cys Gly |     |     |     |         |
|    |                                                                 | 500 |     | 505 | 510     |
| 15 | Asn Gly Met Val Glu Gln Gly Glu Glu Cys Asp Cys Gly Tyr Ser Asp |     |     |     |         |
|    | 515                                                             |     | 520 |     | 525     |
|    | Gln Cys Lys Asp Glu Cys Cys Phe Asp Ala Asn Gln Pro Glu Gly Arg |     |     |     |         |
|    | 530                                                             |     | 535 |     | 540     |
| 20 | Lys Cys Lys Leu Lys Pro Gly Lys Gln Cys Ser Pro Ser Gln Gly Pro |     |     |     |         |
|    | 545                                                             |     | 550 |     | 555 560 |
|    | Cys Cys Thr Ala Gln Cys Ala Phe Lys Ser Lys Ser Glu Lys Cys Arg |     |     |     |         |
| 25 |                                                                 | 565 |     | 570 | 575     |
|    | Asp Asp Ser Asp Cys Ala Arg Glu Gly Ile Cys Asn Gly Phe Thr Ala |     |     |     |         |
|    |                                                                 | 580 |     | 585 | 590     |
| 30 | Leu Cys Pro Ala Ser Asp Pro Lys Pro Asn Phe Thr Asp Cys Asn Arg |     |     |     |         |
|    | 595                                                             |     | 600 |     | 605     |
|    | His Thr Gln Val Cys Ile Asn Gly Gln Cys Ala Gly Ser Ile Cys Glu |     |     |     |         |
|    | 610                                                             |     | 615 |     | 620     |
| 35 | Lys Tyr Gly Leu Glu Glu Cys Thr Cys Ala Ser Ser Asp Gly Lys Asp |     |     |     |         |
|    | 625                                                             |     | 630 |     | 635 640 |

Asp Lys Glu Leu Cys His Val Cys Cys Met Lys Lys Met Asp Pro Ser  
645 650 655

5 Thr Cys Ala Ser Thr Gly Ser Val Gln Trp Ser Arg His Phe Ser Gly  
660 665 670

Arg Thr Ile Thr Leu Gln Pro Gly Ser Pro Cys Asn Asp Phe Arg Gly  
675 680 685

10 Tyr Cys Asp Val Phe Met Arg Cys Arg Leu Val Asp Ala Asp Gly Pro  
690 695 700

Leu Ala Arg Leu Lys Lys Ala Ile Phe Ser Pro Glu Leu Tyr Glu Asn  
15 705 710 715 720

Ile Ala Glu Trp Ile Val Ala His Trp Trp Ala Val Leu Leu Met Gly  
725 730 735

20 Ile Ala Leu Ile Met Leu Met Ala Gly Phe Ile Lys Ile Cys Ser Val  
740 745 750

His Thr Pro Ser Ser Asn Pro Lys Leu Pro Pro Pro Lys Pro Leu Pro  
755 760 765

25 Gly Thr Leu Lys Arg Arg Arg Pro Pro Gln Pro Ile Gln Gln Pro Gln  
770 775 780

Arg Gln Arg Pro Arg Glu Ser Tyr Gln Met Gly His Met Arg Arg  
30 785 790 795

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 239 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single

WO 98/37092

PCT/US98/03490

44

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

5 (iv) ANTI-SENSE: YES

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AATACCACCA TTCTCTGTTA TCCTGAGTAT GTCAATTAAA CAGTAATTTT TAATTAAGAG 60

CGGAAAAATT TTATAATACA AAGAAACATC CATATTGCAA TTTCTGTTTA CAATTGCACA 120

15 CAGAAGTACA GTGTACGTAA GAAATACATG TCTGCATATA ACAAGGTATG TACATTGGCA 180

AGTGATGTCT CCAATGTTGA GGTGGTCGAG CCTCCTAGCC TTGATTGGCA GTTGAAAAA 239

20 (2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 239 base pairs

(B) TYPE: nucleic acid

25 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

35 TTTTCAACT GCCAATCAAG GCTAGGAGGC TCGAATCACT CAACATTGGA GACATCACTT 60

GCCAATGTAC ATACCTTGTT ATATGCAGAC ATGTAATTTCT TACGTACACT GTACTTCTGT 120

GTGCAATTGT AACAGAAAT TGCAATATGG ATGTTTCTTT GTATTATAAA ATTTTCCGC 180

TCTTAATTAA AAATTACTGT TTAATTGACA TACTCAGGAT AACAGAGAAT GGTGGTATT 239

5

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 736 base pairs

10

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

20

AACCACTTCC AGGCACTTTA AAGAGGAGGA GACCTCCACA GCCCATTCAG CAACCCCAGC 60

GTCAGCGGCC CCGAGAGAGT TATCAAATGG GACACATGAG ACGCTAACTG CAGCTTTTGC 120

25

CTTGGTTCTT CCTAGTGCCT ACAATGGGAA AACTTCACTC CAAAGAGAAA CCTATTAAGT 180

CATCATCTCC AAATAAACC CTCACAAGTA ACAGTTGAAG AAAAAATGGC AAGAGATCAT 240

ATCCTCAGAC CAGGTGGAAT TACTTAAATT TTAAAGCCTG AAAATTCCAA TTTGGGGGTG 300

30

GGAGGTGGAA AAGGAACCCA ATTTTCTTAT GAACAGATAT TTTTAACTTA ATGGCACAAA 360

GTCTTAGAAT ATTATTATGT GCCCGTGTT CCTGTTCTT CGTTGCTGCA TTTTCTTCAC 420

35

TTGCAGGCAA ACTTGGCTCT CAATAAAGTT TTAATAAAA TTGAAATAAA TATATTTTTT 480

TCAACTGCCA ATCAAGGCTA GGAGGCTCGA TCAATTAAC ATTGGAGACA ATCACTTGCC 540

WO 98/37092

PCT/US98/03490

46

AATGTACATA CCTTGTTATA TGCAGACATG TATTTCTTAC GTACACTGTA CTTCTGTGTG 600

CAATTGTAAA CAGAAATTGC AATATGGATG TTTCTTTGTA TTATAAAATT TTTCCGCTCT 660

5

TAATTAAAAA TTACTGTTTA ATTGACATAC TCAGGATAAC AGAGAATGGT GGTATTCAGT 720

GGTTCAGACA CAGGCT 736

10 (2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2625 base pairs

(B) TYPE: nucleic acid

15 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 17..2263

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGCGGCGGCA CGGAAG ATG GTG TTG CTG AGA GTG TTA ATT CTG CTC CTC 49

Met Val Leu Leu Arg Val Leu Ile Leu Leu Leu

30

800

805

810

TCC TGG GCG GCG GGG ATG GGA GGT CAG TAT GGG AAT CCT TTA AAT AAA 97

Ser Trp Ala Ala Gly Met Gly Gly Gln Tyr Gly Asn Pro Leu Asn Lys

815

820

825

35

TAT ATC AGA CAT TAT GAA GGA TTA TCT TAT AAT GTG GAT TCA TTA CAC 145

Tyr Ile Arg His Tyr Glu Gly Leu Ser Tyr Asn Val Asp Ser Leu His

WO 98/37092

PCT/US98/03490

47

|    |                                                                 |     |     |     |
|----|-----------------------------------------------------------------|-----|-----|-----|
|    | 830                                                             | 835 | 840 |     |
|    | CAA AAA CAC CAG CGT GCC AAA AGA GCA GTC TCA CAT GAA GAC CAA TTT |     |     | 193 |
|    | Gln Lys His Gln Arg Ala Lys Arg Ala Val Ser His Glu Asp Gln Phe |     |     |     |
| 5  | 845                                                             | 850 | 855 |     |
|    | TTA CGT CTA GAT TTC CAT GCC CAT GGA AGA CAT TTC AAC CTA CGA ATG |     |     | 241 |
|    | Leu Arg Leu Asp Phe His Ala His Gly Arg His Phe Asn Leu Arg Met |     |     |     |
|    | 860                                                             | 865 | 870 |     |
| 10 |                                                                 |     |     |     |
|    | AAG AGG GAC ACT TCC CTT TTC AGT GAT GAA TTT AAA GTA GAA ACA TCA |     |     | 289 |
|    | Lys Arg Asp Thr Ser Leu Phe Ser Asp Glu Phe Lys Val Glu Thr Ser |     |     |     |
|    | 875                                                             | 880 | 885 | 890 |
| 15 | AAT AAA GTA CTT GAT TAT GAT ACC TCT CAT ATT TAC ACT GGA CAT ATT |     |     | 337 |
|    | Asn Lys Val Leu Asp Tyr Asp Thr Ser His Ile Tyr Thr Gly His Ile |     |     |     |
|    | 895                                                             | 900 | 905 |     |
|    | TAT GGT GAA GAA GGA AGT TTT AGC CAT GGG TCT GTT ATT GAT GGA AGA |     |     | 385 |
| 20 | Tyr Gly Glu Glu Gly Ser Phe Ser His Gly Ser Val Ile Asp Gly Arg |     |     |     |
|    | 910                                                             | 915 | 920 |     |
|    | TTT GAA GGA TTC ATC CAG ACT CGT GGT GGC ACA TTT TAT GTT GAG CCA |     |     | 433 |
|    | Phe Glu Gly Phe Ile Gln Thr Arg Gly Gly Thr Phe Tyr Val Glu Pro |     |     |     |
| 25 | 925                                                             | 930 | 935 |     |
|    | GCA GAG AGA TAT ATT AAA GAC CGA ACT CTG CCA TTT CAC TCT GTC ATT |     |     | 481 |
|    | Ala Glu Arg Tyr Ile Lys Asp Arg Thr Leu Pro Phe His Ser Val Ile |     |     |     |
|    | 940                                                             | 945 | 950 |     |
| 30 |                                                                 |     |     |     |
|    | TAT CAT GAA GAT GAT ATT AAC TAT CCC CAT AAA TAC GGT CCT CAG GGC |     |     | 529 |
|    | Tyr His Glu Asp Asp Ile Asn Tyr Pro His Lys Tyr Gly Pro Gln Gly |     |     |     |
|    | 955                                                             | 960 | 965 | 970 |
| 35 | GGC TGT GCA GAT CAT TCA GTA TTT GAA AGA ATG AGG AAA TAC CAG ATG |     |     | 577 |
|    | Gly Cys Ala Asp His Ser Val Phe Glu Arg Met Arg Lys Tyr Gln Met |     |     |     |
|    | 975                                                             | 980 | 985 |     |

WO 98/37092

PCT/US98/03490

48

|    |                                                                 |      |
|----|-----------------------------------------------------------------|------|
|    | ACT GGT GTA GAG GAA GTA ACA CAG ATA CCT CAA GAA GAA CAT GCT GCT | 625  |
|    | Thr Gly Val Glu Glu Val Thr Gln Ile Pro Gln Glu Glu His Ala Ala |      |
|    | 990 995 1000                                                    |      |
| 5  |                                                                 |      |
|    | AAT GGT CCA GAA CTT CTG AGG AAA AGA CGT ACA ACT TCA GCT GAA AAA | 673  |
|    | Asn Gly Pro Glu Leu Leu Arg Lys Arg Arg Thr Thr Ser Ala Glu Lys |      |
|    | 1005 1010 1015                                                  |      |
| 10 |                                                                 |      |
|    | AAT ACT TGT CAG CTT TAT ATT CAG ACT GAT CAT TTG TTC TTT AAA TAT | 721  |
|    | Asn Thr Cys Gln Leu Tyr Ile Gln Thr Asp His Leu Phe Phe Lys Tyr |      |
|    | 1020 1025 1030                                                  |      |
|    |                                                                 |      |
|    | TAC GGA ACA CGA GAA GCT GTG ATT GCC CAG ATA TCC AGT CAT GTT AAA | 769  |
| 15 | Tyr Gly Thr Arg Glu Ala Val Ile Ala Gln Ile Ser Ser His Val Lys |      |
|    | 1035 1040 1045 1050                                             |      |
|    |                                                                 |      |
|    | GCG ATT GAT ACA ATT TAC CAG ACC ACA GAC TTC TCC GGA ATC CGT AAC | 817  |
|    | Ala Ile Asp Thr Ile Tyr Gln Thr Thr Asp Phe Ser Gly Ile Arg Asn |      |
| 20 | 1055 1060 1065                                                  |      |
|    |                                                                 |      |
|    | ATC AGT TTC ATG GTG AAA CGC ATA AGA ATC AAT ACA ACT GCT GAT GAG | 865  |
|    | Ile Ser Phe Met Val Lys Arg Ile Arg Ile Asn Thr Thr Ala Asp Glu |      |
|    | 1070 1075 1080                                                  |      |
| 25 |                                                                 |      |
|    | AAG GAC CCT ACA AAT CCT TTC CGT TTC CCA AAT ATT AGT GTG GAG AAG | 913  |
|    | Lys Asp Pro Thr Asn Pro Phe Arg Phe Pro Asn Ile Ser Val Glu Lys |      |
|    | 1085 1090 1095                                                  |      |
|    |                                                                 |      |
| 30 |                                                                 |      |
|    | TTT CTG GAA TTG AAT TCT GAG CAG AAT CAT GAT GAC TAC TGT TTG GCC | 961  |
|    | Phe Leu Glu Leu Asn Ser Glu Gln Asn His Asp Asp Tyr Cys Leu Ala |      |
|    | 1100 1105 1110                                                  |      |
|    |                                                                 |      |
|    | TAT GTC TTC ACA GAC CGA GAT TTT GAT CAT GGT GTA CTT GGT CTG GCT | 1009 |
| 35 | Tyr Val Phe Thr Asp Arg Asp Phe Asp Asp Gly Val Leu Gly Leu Ala |      |
|    | 1115 1120 1125 1130                                             |      |

WO 98/37092

PCT/US98/03490

49

|    |                                                                 |      |
|----|-----------------------------------------------------------------|------|
|    | TGG GTT GGA GCA CCT TCA GGA AGC TCT GGA GGA ATA TGT GAA AAA AGT | 1057 |
|    | Trp Val Gly Ala Pro Ser Gly Ser Ser Gly Gly Ile Cys Glu Lys Ser |      |
|    | 1135 1140 1145                                                  |      |
| 5  | AAA CTC TAT TCA GAT GGT AAG AAG AAG TCC TTA AAC ACT GGA ATT ATT | 1105 |
|    | Lys Leu Tyr Ser Asp Gly Lys Lys Lys Ser Leu Asn Thr Gly Ile Ile |      |
|    | 1150 1155 1160                                                  |      |
|    | ACT GTT CAG AAC TAT GGG TCT CAT GTA CCT CCC AAA GTC TCT CAC ATT | 1153 |
| 10 | Thr Val Gln Asn Tyr Gly Ser His Val Pro Pro Lys Val Ser His Ile |      |
|    | 1165 1170 1175                                                  |      |
|    | ACT TTT GCT CAC GAA GTT GGA CAT AAC TTT GGA TCC CCA CAT GAT TCT | 1201 |
|    | Thr Phe Ala His Glu Val Gly His Asn Phe Gly Ser Pro His Asp Ser |      |
| 15 | 1180 1185 1190                                                  |      |
|    | GGA ACA GAG TGC ACA CCA GGA GAA TCT AAG AAT TTG GGT CAA AAA GAA | 1249 |
|    | Gly Thr Glu Cys Thr Pro Gly Glu Ser Lys Asn Leu Gly Gln Lys Glu |      |
|    | 1195 1200 1205 1210                                             |      |
| 20 | AAT GGC AAT TAC ATC ATG TAT GCA AGA GCA ACA TCT GGG GAC AAA CTT | 1297 |
|    | Asn Gly Asn Tyr Ile Met Tyr Ala Arg Ala Thr Ser Gly Asp Lys Leu |      |
|    | 1215 1220 1225                                                  |      |
| 25 | AAC AAC AAT AAA TTC TCA CTC TGT AGT ATT AGA AAT ATA AGC CAA GTT | 1345 |
|    | Asn Asn Asn Lys Phe Ser Leu Cys Ser Ile Arg Asn Ile Ser Gln Val |      |
|    | 1230 1235 1240                                                  |      |
|    | CTT GAG AAG AAG AGA AAC AAC TGT TTT GTT GAA TCT GGC CAA CCT ATT | 1393 |
| 30 | Leu Glu Lys Lys Arg Asn Asn Cys Phe Val Glu Ser Gly Gln Pro Ile |      |
|    | 1245 1250 1255                                                  |      |
|    | TGT GGA AAT GGA ATG GTA GAA CAA GGT GAA GAA TGT GAT TGT GGC TAT | 1441 |
|    | Cys Gly Asn Gly Met Val Glu Gln Gly Val Glu Cys Asp Cys Gly Tyr |      |
| 35 | 1260 1265 1270                                                  |      |
|    | AGT GAC CAG TGT AAA GAT GAA TGC TGC TGT TAT TCA AAT CAA CCA GAG | 1489 |

WO 98/37092

PCT/US98/03490

50

|    |                                                                 |                |
|----|-----------------------------------------------------------------|----------------|
|    | Ser Asp Gln Cys Lys Asp Glu Cys Cys Phe Asp Ala Asn Gln Pro Glu |                |
|    | 1275                                                            | 1280 1285 1290 |
|    | GGA AGA AAA TGC AAA CTG AAA CCT GGG AAA CAG TGC AGT CCA AGT CAA | 1537           |
| 5  | Gly Arg Lys Cys Lys Leu Lys Pro Gly Lys Gln Cys Ser Pro Ser Gln |                |
|    | 1295 1300 1305                                                  |                |
|    | GGT CCT TGT TGT ACA GCA CAG TGT GCA TTC AAG TCA AAG TCT GAG AAG | 1585           |
|    | Gly Pro Cys Cys Thr Ala Gln Cys Ala Phe Lys Ser Lys Ser Glu Lys |                |
| 10 | 1310 1315 1320                                                  |                |
|    | TGT CGG GAT GAT TCA GAC TGT GCA AGG GAA GGA ATA TGT AAT GGC TTC | 1633           |
|    | Cys Arg Asp Asp Ser Asp Cys Ala Arg Glu Gly Ile Cys Asn Gly Phe |                |
|    | 1325 1330 1335                                                  |                |
| 15 |                                                                 |                |
|    | ACA GCT CTC TGC CCA GCA TCT GAC CCT AAA CCA AAC TTC ACA GAC TGT | 1681           |
|    | Thr Ala Leu Cys Pro Ala Ser Asp Pro Lys Pro Asn Phe Thr Asp Cys |                |
|    | 1340 1345 1350                                                  |                |
| 20 | AAT AGG CAT ACA CAA GTG TGC ATT AAT GGG CAA TGT GCA GGT TCT ATC | 1729           |
|    | Asn Arg His Thr Gln Val Cys Ile Asn Gly Gln Cys Ala Gly Ser Ile |                |
|    | 1355 1360 1365 1370                                             |                |
|    | TGT GAG AAA TAT GGC TTA GAG GAG TGT ACG TGT GCC AGT TCT GAT GGC | 1777           |
| 25 | Cys Glu Lys Tyr Gly Leu Glu Glu Cys Thr Cys Ala Ser Ser Asp Gly |                |
|    | 1375 1380 1385                                                  |                |
|    | AAA GAT GAT AAA GAA TTA TGC CAT GTA TGC TGT ATG AAG AAA ATG GAC | 1825           |
|    | Lys Asp Asp Lys Glu Leu Cys His Val Cys Cys Met Lys Lys Met Asp |                |
| 30 | 1390 1395 1400                                                  |                |
|    | CCA TCA ACT TGT GCC AGT ACA GGG TCT GTG CAG TGG AGT AGG CAC TTC | 1873           |
|    | Pro Ser Thr Cys Ala Ser Thr Gly Ser Val Ala Trp Ser Arg His Phe |                |
|    | 1405 1410 1415                                                  |                |
| 35 |                                                                 |                |
|    | AGT GGT CGA ACC ATC ACC CTG CAA CCT GAA TGC CCT TGC AAC GAT TTT | 1921           |
|    | Ser Gly Arg Thr Ile Thr Leu Gln Gly Ser Pro Cys Asn Asp Phe     |                |

WO 98/37092

PCT/US98/03490

51

|    | 1420                                                              | 1425 | 1430 |      |
|----|-------------------------------------------------------------------|------|------|------|
|    | AGA GGT TAC TGT GAT GTT TTC ATG CGG TGC AGA TTA GTA GAT GCT GAT   |      |      | 1969 |
|    | Arg Gly Tyr Cys Asp Val Phe Met Arg Cys Arg Leu Val Asp Ala Asp   |      |      |      |
| 5  | 1435                                                              | 1440 | 1445 | 1450 |
|    | GGT CCT CTA GCT AGG CTT AAA AAA GCA ATT TTT AGT CCA GAG CTC TAT   |      |      | 2017 |
|    | Gly Pro Leu Ala Arg Leu Lys Lys Ala Ile Phe Ser Pro Glu Leu Tyr   |      |      |      |
|    |                                                                   | 1455 | 1460 | 1465 |
| 10 |                                                                   |      |      |      |
|    | GAA AAC ATT GCT GAA TGG ATT GTG GCT CAT TGG TGG GCA GTA TTA CTT   |      |      | 2065 |
|    | Glu Asn Ile Ala Glu Trp Ile Val Ala His Trp Trp Ala Val Leu Leu   |      |      |      |
|    |                                                                   | 1470 | 1475 | 1480 |
| 15 |                                                                   |      |      |      |
|    | ATG GGA ATT GCT CTG ATC ATG CTA ATG GCT GGA TTT ATT AAG ATA TGC   |      |      | 2113 |
|    | Met Gly Ile Ala Leu Ile Met Leu Met Ala Gly Phe Ile Lys Ile Cys   |      |      |      |
|    |                                                                   | 1485 | 1490 | 1495 |
|    |                                                                   |      |      |      |
|    | AGT GTT CAT ACT CCA AGT AGT AAT CCA AAG TTG CCT CCT CCT AAA CCA   |      |      | 2161 |
| 20 | Ser Val His Thr Pro Ser Ser Asn Pro Lys Leu Pro Pro Pro Lys Pro   |      |      |      |
|    |                                                                   | 1500 | 1505 | 1510 |
|    |                                                                   |      |      |      |
|    | CTT CCA GGC ACT TTA AAG AGG AGG AGA CCT CCA CAG CCC ATT CAG CAA   |      |      | 2209 |
|    | Leu Pro Gly Thr Leu Lys Arg Arg Arg Pro Pro Gln Pro Ile Gln Gln   |      |      |      |
| 25 | 1515                                                              | 1520 | 1525 | 1530 |
|    |                                                                   |      |      |      |
|    | CCC CAG CGT CAG CGG CCC CGA GAG AGT TAT CAA ATG GGA CAC ATG AGA   |      |      | 2257 |
|    | Pro Gln Arg Gln Arg Pro Arg Glu Ser Tyr Gln Met Gly His Met Arg   |      |      |      |
|    |                                                                   | 1535 | 1540 | 1545 |
| 30 |                                                                   |      |      |      |
|    | CGC TAA CTGCAGCTTT TGCCTTGGTT CTCCTAGTG CCTACAATGG GAAAACTTCA     |      |      | 2313 |
|    | Arg *                                                             |      |      |      |
|    |                                                                   |      |      |      |
| 35 |                                                                   |      |      |      |
|    | CTCCAAAGAG AAACCTATTA AGTCATCATC TCCAAATTA ACCCTCACAA GTAACAGTTG  |      |      | 2373 |
|    |                                                                   |      |      |      |
|    | AAGAAAAAAT GGCAAGAGAT CATATCCTCA GACCAAGTGG AATTACTTAA ATTTTAAAGC |      |      | 2433 |

WO 98/37092

PCT/US98/03490

52

CTGAAAATTC CAATTTGGGG GTGGGAGGTG GAAAAGGAAC CCAATTTTCT TATGAACAGA 2493

TATTTTAAAC TTAATGGCAC AAAGTCTTAG AATATTATTA TGTGCCCCGT GTTCCCTGTT 2553

CTTCGTTGCT GCATTTTCTT CACTTGCAGG CAAACTTGGC TCTCAATAAA CTTTACCAC 2613

AAAAAAAAAA AA 2625

10

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 749 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: protein

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Val Leu Leu Arg Val Leu Ile Leu Leu Leu Ser Trp Ala Ala Gly  
1 5 10 15

Met Gly Gly Gln Tyr Gly Asn Pro Leu Asn Lys Tyr Ile Arg His Tyr  
20 25 30

Glu Gly Leu Ser Tyr Asn Val Asp Ser Leu His Gln Lys His Gln Arg  
35 40 45

Ala Lys Arg Ala Val Ser His Glu Asp Gln Phe Leu Arg Leu Asp Phe  
50 55 60

His Ala His Gly Arg His Phe Asn Leu Arg Met Lys Arg Asp Thr Ser  
65 70 75 80

Leu Phe Ser Asp Glu Phe Lys Val Glu Thr Ser Asn Lys Val Leu Asp

WO 98/37092

PCT/US98/03490

53

85

90

95

Tyr Asp Thr Ser His Ile Tyr Thr Gly His Ile Tyr Gly Glu Glu Gly

100

105

110

5

Ser Phe Ser His Gly Ser Val Ile Asp Gly Arg Phe Glu Gly Phe Ile

115

120

125

Gln Thr Arg Gly Gly Thr Phe Tyr Val Glu Pro Ala Glu Arg Tyr Ile

10

130

135

140

Lys Asp Arg Thr Leu Pro Phe His Ser Val Ile Tyr His Glu Asp Asp

145

150

155

160

15

Ile Asn Tyr Pro His Lys Tyr Gly Pro Gln Gly Gly Cys Ala Asp His

165

170

175

Ser Val Phe Glu Arg Met Arg Lys Tyr Gln Met Thr Gly Val Glu Glu

180

185

190

20

Val Thr Gln Ile Pro Gln Glu Glu His Ala Ala Asn Gly Pro Glu Leu

195

200

205

Leu Arg Lys Arg Arg Thr Thr Ser Ala Glu Lys Asn Thr Cys Gln Leu

25

210

215

220

Tyr Ile Gln Thr Asp His Leu Phe Phe Lys Tyr Tyr Gly Thr Arg Glu

225

230

235

240

30

Ala Val Ile Ala Gln Ile Ser Ser His Val Lys Ala Ile Asp Thr Ile

245

250

255

Tyr Gln Thr Thr Asp Phe Ser Gly Ile Arg Asn Ile Ser Phe Met Val

260

265

270

35

Lys Arg Ile Arg Ile Asn Thr Thr Ala Asp Glu Lys Asp Pro Thr Asn

275

280

285

WO 98/37092

PCT/US98/03490

54

Pro Phe Arg Phe Pro Asn Ile Ser Val Glu Lys Phe Leu Glu Leu Asn  
 290 295 300

5 Ser Glu Gln Asn His Asp Asp Tyr Cys Leu Ala Tyr Val Phe Thr Asp  
 305 310 315 320

Arg Asp Phe Asp Asp Gly Val Leu Gly Leu Ala Trp Val Gly Ala Pro  
 325 330 335

10 Ser Gly Ser Ser Gly Gly Ile Cys Glu Lys Ser Lys Leu Tyr Ser Asp  
 340 345 350

Gly Lys Lys Lys Ser Leu Asn Thr Gly Ile Ile Thr Val Gln Asn Tyr  
 15 355 360 365

Gly Ser His Val Pro Pro Lys Val Ser His Ile Thr Phe Ala His Glu  
 370 375 380

20 Val Gly His Asn Phe Gly Ser Pro His Asp Ser Gly Thr Glu Cys Thr  
 385 390 395 400

Pro Gly Glu Ser Lys Asn Leu Gly Gln Lys Glu Asn Gly Asn Tyr Ile  
 405 410 415

25 Met Tyr Ala Arg Ala Thr Ser Gly Asp Lys Leu Asn Asn Asn Lys Phe  
 420 425 430

Ser Leu Cys Ser Ile Arg Asn Ile Ser Gln Val Leu Glu Lys Lys Arg  
 30 435 440 445

Asn Asn Cys Phe Val Glu Ser Gly Gln Pro Ile Cys Gly Asn Gly Met  
 450 455 460

35 Val Glu Gln Gly Glu Glu Cys Asp Cys Gly Tyr Ser Asp Gln Cys Lys  
 465 470 475 480

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
|    | Asp | Glu | Cys | Cys | Phe | Asp | Ala | Asn | Gln | Pro | Glu | Gly | Arg | Lys | Cys | Lys |  |
|    |     |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     |     | 495 |  |
| 5  | Leu | Lys | Pro | Gly | Lys | Gln | Cys | Ser | Pro | Ser | Gln | Gly | Pro | Cys | Cys | Thr |  |
|    |     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |  |
|    | Ala | Gln | Cys | Ala | Phe | Lys | Ser | Lys | Ser | Glu | Lys | Cys | Arg | Asp | Asp | Ser |  |
|    |     |     | 515 |     |     |     |     | 520 |     |     |     |     | 525 |     |     |     |  |
| 10 | Asp | Cys | Ala | Arg | Glu | Gly | Ile | Cys | Asn | Gly | Phe | Thr | Ala | Leu | Cys | Pro |  |
|    |     | 530 |     |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |  |
|    | Ala | Ser | Asp | Pro | Lys | Pro | Asn | Phe | Thr | Asp | Cys | Asn | Arg | His | Thr | Gln |  |
|    | 545 |     |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |  |
| 15 | Val | Cys | Ile | Asn | Gly | Gln | Cys | Ala | Gly | Ser | Ile | Cys | Glu | Lys | Tyr | Gly |  |
|    |     |     |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |  |
|    | Leu | Glu | Glu | Cys | Thr | Cys | Ala | Ser | Ser | Asp | Gly | Lys | Asp | Asp | Lys | Glu |  |
| 20 |     |     |     | 580 |     |     |     |     | 585 |     |     |     |     | 590 |     |     |  |
|    | Leu | Cys | His | Val | Cys | Cys | Met | Lys | Lys | Met | Asp | Pro | Ser | Thr | Cys | Ala |  |
|    |     |     | 595 |     |     |     |     | 600 |     |     |     |     | 605 |     |     |     |  |
| 25 | Ser | Thr | Gly | Ser | Val | Gln | Trp | Ser | Arg | His | Phe | Ser | Gly | Arg | Thr | Ile |  |
|    |     | 610 |     |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     |  |
|    | Thr | Leu | Gln | Pro | Gly | Ser | Pro | Cys | Asn | Asp | Phe | Arg | Gly | Tyr | Cys | Asp |  |
|    | 625 |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     |     | 640 |  |
| 30 | Val | Phe | Met | Arg | Cys | Arg | Leu | Val | Asp | Ala | Asp | Gly | Pro | Leu | Ala | Arg |  |
|    |     |     |     |     | 645 |     |     |     |     | 650 |     |     |     |     |     | 655 |  |
|    | Leu | Lys | Lys | Ala | Ile | Phe | Ser | Pro | Glu | Leu | Tyr | Glu | Asn | Ile | Ala | Glu |  |
| 35 |     |     |     | 660 |     |     |     |     | 665 |     |     |     |     | 670 |     |     |  |
|    | Trp | Ile | Val | Ala | His | Trp | Trp | Ala | Leu | Leu | Met | Gly | Ile | Ala | Leu |     |  |

WO 98/37092

PCT/US98/03490

56

675

680

685

Ile Met Leu Met Ala Gly Phe Ile Lys Ile Cys Ser Val His Thr Pro

690

695

700

5

Ser Ser Asn Pro Lys Leu Pro Pro Pro Lys Pro Leu Pro Gly Thr Leu

705

710

715

720

Lys Arg Arg Arg Pro Pro Gln Pro Ile Gln Gln Pro Gln Arg Gln Arg

10

725

730

735

Pro Arg Glu Ser Tyr Gln Met Gly His Met Arg Arg \*

740

745

15 (2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 base pairs

(B) TYPE: nucleic acid

20 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGCCTGTGTC

10

(2) INFORMATION FOR SEQ ID NO:11:

30 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: DNA (genomic)

WO 98/37092

PCT/US98/03490

57

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AGCCTGTGTC TGAACCACT

19

## WHAT IS CLAIMED IS:

1. A DNA fragment encoding a human disintegrin of SEQ ID NO:9 or a  
5 fragment thereof expressed differentially during arthritis development, capable as  
being used as a screen for disintegrin antagonism, drug design and screening.
2. The human disintegrin, or fragment thereof encoded by the DNA of Claim 1 in  
essentially pure form.
- 10 3. A screening method for compounds capable of binding to a human disintegrin,  
comprising the disintegrin of Claim 1.
4. A screening kit for compounds capable of binding to a human disintegrin,  
15 comprising the disintegrin of Claim 1.
5. A screening kit for osteoarthritis comprising an antibody, or fragment thereof,  
to human disintegrin of Claim 2.
- 20 6. An expression vector or plasmid comprising the DNA of Claim 1.
7. An isolated nucleic acid molecule of Claim 1, comprising the sequence set  
forth in SEQ ID NO:8.
- 25 8. A nucleic acid molecule of Claim 1 detectable by primer extension using a  
primer selected from the group consisting of SEQ ID NO:10 and SEQ ID NO:11.
9. A method of Claim 3 comprising:
  - 30 A) exposing a portion of Interleukin-1 stimulated cultures of chondrocytes to  
candidate inhibitors of metalloprotease gene expression;
  - B) isolating RNA from said exposed portion, said RNA comprising mRNA  
corresponding to a metalloprotease gene;
  - C) comparing the level of said mRNA of said metalloprotease gene from said  
exposed portion with the level in said unexposed portion; and
  - 35 D) observing reduced levels of said mRNA as indicative of an inhibitor.

10. A method of Claim 3 comprising:

- 5 A) isolating a sample of culture supernatant of Interleukin-1 stimulated cultures of normal human articular chondrocytes grown in the presence of candidate inhibitors and control inhibitors, and in the and absence of any inhibitors;
- B) adding a substrate to each sample, said substrate capable of detecting metalloprotease activity; and
- C) detecting the level of metalloprotease activity for each sample.