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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 June 2009 (25.06.2009)

PCT

(10) International Publication Number
WO 2009/077755 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/GB2008/004167

(22) International Filing Date:
17 December 2008 (17.12.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0724556.6 17 December 2007 (17.12.2007) GB

(71) Applicant *(for all designated States except US)*: **QUEEN MARY & WESTFIELD COLLEGE** [GB/GB]; Mile End Road, London E1 4NS (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CHERNAJOVSKY, Yuti** [GB/GB]; William Harvey Research Institute, Barts and The London, Queen Mary's School of Medicine & Dentistry, John Vane Building, Charterhouse Square, London EC1M 6BQ (GB). **ADAMS, Gillian** [GB/GB]; William Harvey Research Institute, Barts and The London, Queen Mary's School of Medicine & Dentistry, John Vane Building, Charterhouse Square, London EC1M 6BQ (GB).

(74) Agents: **BASSIL, Nicholas, Charles** et al.; Kilburn & Strode LLP, 20 Red Lion Street, London WC1R 4PJ (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

WO 2009/077755 A2

(54) **Title:** LATENCY ASSOCIATED PROTEIN CONSTRUCT WITH AGGRECANASE SENSITIVE CLEAVAGE SITE

(57) **Abstract:** The present provides a fusion protein comprising a latency associated peptide (LAP) and a pharmaceutically active agent in which the LAP and the pharmaceutically active agent are connected by an amino acid sequence comprising an aggrecanase proteolytic cleavage site.

LATENCY ASSOCIATED PROTEIN CONSTRUCT WITH AGGRECANASE SENSITIVE CLEAVAGE SITE

The present invention relates to the use of proteins, protein derivatives and DNA constructs that confer latency to pharmaceutically active agents where the pharmaceutically agent is released by the action of aggrecanase. Such products are useful in the treatment of arthritis and cancer.

Most cytokines and growth factors are expressed under tight control mechanisms. Their gene expression is regulated by environmental stimuli such as infection, cell-cell interactions, change in extracellular matrix composition and interactions with adhesion molecules or via stimulation with other cytokines.

In addition to the control at the transcriptional and post-transcriptional level, some cytokines are not released into the medium unless a second signal activates the cell. A third level of regulation for cytokine activity is found in molecules which are secreted in a latent form and become "activated" by releasing the cytokine moiety where processes of inflammation, wound healing and tissue repair takes place (Khalil N, Microbes and Infection, **1,** 1255-1263 (1999). In this latter respect, transforming growth factor beta (TGFβ) has received greatest attention.

TGFβ is synthesized as a dimeric latent cytokine composed of an amino terminal latency associated protein (LAP) and the active TGFβ cytokine at its COOH terminal end (Roberts and Sporn, Peptide Growth Factors and their Receptors: Sporn, MB and Roberts, AB, Springer-Verlag, 419-472 (1996); Roth-Eicchorn et al., Hepatology, **28** 1588-1596 (1998)). The precursor peptide contains a signal peptide (residues 1-29) necessary for protein secretion and guiding the molecule through the Golgi apparatus to become processed by proteolytic cleavage and glycosylation. The LAP domain is separated from TGFβ by proteolytic cleavage at arginines (277-278). Mature TGFβ begins at alanine 279. The LAP, in addition to protect TGFβ, contains important residues necessary for the interaction with other molecules. Mutations in the LAP domain have recently been associated with the autosomal dominant Camurati-Engelmann disease (Janssens et al., Nature Genetics, **26,** 273-275 (2000). Cysteines

224 and 226 are important in the intermolecular disulphide bond between two LAPs. Their mutation to serine renders the molecule "active" (Sanderson et al., Proc. Natl. Acad. Sci . USA, **92**, 2572-2576 (1995); Brunner et al., Mol. Endocrinol. **6**, 1691-1700 (1992); Brunner et al., J. Biol. Chem, **264**, 13660-13664 (1989)). The RGD motif (245-247) facilitates the interaction with integrins (Munger et al., Mol, Biol. of the Cell, **9**, 2627-2638 (1998; Derynck R, TIBS, **19**, 548-553 (1994)). Nucleic acid encoding TGFβ is described in US 5801231.

In most cell types studied, including those of mesenchymal, epithelial and endothelial origin, TGFβ is secreted in a latent form consisting of TGFβ and its latency associated peptide (LAP) propeptide dimers, covalently linked to latent TGFβ-binding proteins (LTBPs). LTBPs are also needed for the secretion and folding of TGFβ (Miyazano et al., EMBO J. **10**, 1091-1101 (1991); Miyazano et al., J. Biol. Chem. **267**, 5668-5675 (1992); Eklov et al., Cancer Res. **53**, 3193-3197 (1993)). Cysteine 33 is important for the disulphide bridge with the third 8 cysteine-rich repeat of latent TGFβ binding protein (LTBP) (Saharinen et al., The EMBO Journal, **15**, 245-253 (1996). Modification of LTBP by enzymes such as thrombospondin (Schultz et al., The Journal of Biological Chemistry, **269**, 26783-26788 (1994); Crawford et al., Cell, **93**, 1159-1170 (1998)), transglutaminase (Nunes et al., J. Cell, Biol. **136**, 1151-1163 (1997); Kojima et al., The Journal of Cell Biology, **121**, 439-448 (1993)) and MMP9, MMP2 (Yu and Stamenkovic, Genes and Dev, **14**, 163-176 (2000)) could release the active portion of TGFβ from the latent complex.

Cytokines are natural products serving as soluble local mediators of cell-cell interactions. They have a variety of pleiotropic actions, some of which can be harnessed for therapeutic purposes. Targeting of cytokines to specific cell types using scFv (Lode et al., Pharmacol. Ther, **80**, 277-292 (1998)) and vWF (Gordon et al., Human Gene Therapy, **8**, 1385-1394 (1997)) have focused entirely on the active cytokine moiety of the cytokine complex.

Pharmacologically active proteins or other medicines based on such agents, which have to be administered at very high concentrations systemically in order to achieve

biologically effective concentrations in the tissue being targeted, tend to give rise to undesirable systemic effects, for example toxicity, which limit their use and efficacy.

The principles underlying the construction of such a system for providing latency to pharmaceutically active agents using the LAP of TGF-β was described in WO 02/055098. The present inventors have now developed an improved means for providing pharmaceutically active agents in latent form based on this system.

According to a first aspect of the invention there is provided a fusion protein comprising a latency associated peptide (LAP) and a pharmaceutically active agent in which the LAP and the pharmaceutically active agent are connected by an amino acid sequence comprising an aggrecanase proteolytic cleavage site.

The fusion protein comprising a LAP, an aggrecanase proteolytic cleavage site and a pharmaceutically active agent may provide for site specific activation of the latent pharmaceutically active agent. The term "site specific activation" as used herein means, in general terms and not limited to the removal or reduction of latency, conferred on a pharmaceutically active agent, by site-specific cleavage at the aggrecanase proteolytic cleavage site.

Site-specific cleavage at the proteolytic cleavage site is expected to take place concomitantly with the restored activation of the pharmaceutically active agent.

The term "latent pharmaceutically active agent" as used herein may include, but is not limited to, pharmaceutically active agents which are latent due to their association with LAP and an aggrecanase proteolytic cleavage site. Specifically, the pharmaceutically active agent may be latent by virtue of its fusion to a LAP associated aggrecanase proteolytic cleavage site to form a latent fusion protein.

The term "protein" in this text means, in general terms, a plurality of amino acid residues joined together by peptide bonds. It is used interchangeably and means the same as peptide, oligopeptide, oligomer or polypeptide, and includes glycoproteins and derivatives thereof. The term "protein" is also intended to include fragments,

analogues and derivatives of a protein wherein the fragment, analogue or derivative retains essentially the same biological activity or function as a reference protein.

The fragment analogue or derivative of the protein as defined in this text, may be at least 6, preferably 10 or 20, or up to 50 or 100 amino acids long.

The fragment, derivative or analogue of the protein may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably, a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence which is employed for purification of the polypeptide. Such fragments, derivatives and analogues are deemed to be within the scope of those skilled in the art from the teachings herein.

Particularly preferred are variants, analogues, derivatives and fragments having the amino acid sequence of the protein in which several e.g. 5 to 10, or 1 to 5, or 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added in any combination. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the protein of the present invention. Also especially preferred in this regard are conservative substitutions.

An example of a variant of the present invention is a fusion protein as defined above, apart from the substitution of one or more amino acids with one or more other amino acids. The skilled person is aware that various amino acids have similar properties. One or more such amino acids of a substance can often be substituted by one or more other such amino acids without eliminating a desired activity of that substance.

Thus the amino acids glycine, alanine, valine, leucine and isoleucine can often be substituted for one another (amino acids having aliphatic side chains). Of these possible substitutions it is preferred that glycine and alanine are used to substitute for one another

(since they have relatively short side chains) and that valine, leucine and isoleucine are used to substitute for one another (since they have larger aliphatic side chains which are hydrophobic). Other amino acids which can often be substituted for one another include: phenylalanine, tyrosine and tryptophan (amino acids having aromatic side chains); lysine, arginine and histidine (amino acids having basic side chains); aspartate and glutamate (amino acids having acidic side chains); asparagine and glutamine (amino acids having amide side chains); and cysteine and methionine (amino acids having sulphur containing side chains).

Substitutions of this nature are often referred to as "conservative" or "semi-conservative" amino acid substitutions.

Amino acid deletions or insertions may also be made relative to the amino acid sequence for the fusion protein referred to above. Thus, for example, amino acids which do not have a substantial effect on the activity of the polypeptide, or at least which do not eliminate such activity, may be deleted. Such deletions can be advantageous since the overall length and the molecular weight of a polypeptide can be reduced whilst still retaining activity. This can enable the amount of polypeptide required for a particular purpose to be reduced - for example, dosage levels can be reduced.

Amino acid insertions relative to the sequence of the fusion protein above can also be made. This may be done to alter the properties of a substance of the present invention (e.g. to assist in identification, purification or expression, as explained above in relation to fusion proteins).

Amino acid changes relative to the sequence for the fusion protein of the invention can be made using any suitable technique e.g. by using site-directed mutagenesis.

It should be appreciated that amino acid substitutions or insertions within the scope of the present invention can be made using naturally occurring or non-naturally occurring amino acids. Whether or not natural or synthetic amino acids are used, it is preferred that only L- amino acids are present.

A protein according to the invention may have additional N-terminal and/or C-terminal amino acid sequences. Such sequences can be provided for various reasons, for example, glycosylation.

The term "fusion protein" in this text means, in general terms, one or more proteins joined together by chemical means, including hydrogen bonds or salt bridges, or by peptide bonds through protein synthesis or both.

The latency associated peptide (LAP) of the present invention may include, but is not limited to, the coding sequence for the precursor domain of TGFβ or a sequence which is substantially identical thereto.

"Identity" as known in the art is the relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, identity also means the degree of sequence relatedness (homology) between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. While there exist a number of methods to measure identity between two polypeptide or two polynucleotide sequences, methods commonly employed to determine identity are codified in computer programs. Preferred computer programs to determine identity between two sequences include, but are not limited to, GCG program package (Devereux, et al., Nucleic acids Research, **12**, 387 (1984), BLASTP, BLASTN, and FASTA (Atschul et al., J. Molec. Biol. **215**, 403 (1990).

The LAP of the present invention may comprise the precursor domain of TGFβ, for example, the precursor peptide of TGFβ-1, 2 or 3 (from human) (Derynck et al., Nature, **316**, 701-705 (1985); De Martin et al., EMBO J. **6** 3673-3677 (1987); Hanks et al., Proc. Natl. Acad. Sci. **85**, 79-82 (1988); Derynck et al., EMBO J. **7**, 3737-3743 (1988); Ten Dyke et al., Proc. Natl. Acad. Sci. USA, **85**, 4715-4719 (1988)) TGFβ-4 (from chicken) (Jakowlew et al., Mol. Endocrinol. **2**, 1186-1195 (1988)) or TGFβ-5 (from xenopus) (Kondaiah et al., J. Biol. Chem. **265**, 1089-1093 (1990)). The term "precursor domain" is defined as a sequence encoding a precursor peptide which does not include the sequence encoding the mature protein. The amino acid sequences of

the precursor domain of TGFβ 1, 2, 3, 4 and 5 (Roberts and Sporn, Peptide Growth Factors and their Receptors: Sporn, MB and Roberts, AB, Springer-Verlag, Chapter 8, 422 (1996)) are shown in Figure 1.

Preferably, the amino acid sequence of the LAP has at least 50% identity, using the default parameters of the BLAST computer program (Atschul et al., J. Mol. Biol. **215**, 403-410 (1990) provided by HGMP (Human Genome Mapping Project), at the amino acid level, to the precursor domain of TGFβ 1, 2, 3, 4 or 5 (Roberts and Sporn, Peptide Growth Factors and their Receptors: Sporn, MB and Roberts, AB, Springer-Verlag, Chapter 8, 422 (1996)) as shown in Figure 1. More preferably, the LAP may have at least 60%, 70%, 80%, 90% and still more preferably 95% (still more preferably at least 99%) identity, at the nucleic acid or amino acid level, to the precursor domain of TGFβ 1, 2, 3, 4 or 5 as shown in Figure 1 which comprises residues 1 to 278.

The LAP may comprise the LAP of TGFβ 1, 2, 3, 4, or 5 (Roberts and Sporn, Peptide Growth Factors and their Receptors: Sporn, MB and Roberts, AB, Springer-Verlag, Chapter 8, 422 (1996)) as shown in Figure 1.

The LAP may contain at least two, for example at least 4, 6, 8, 10 or 20 cysteine residues for the formation of disulphide bonds.

The LAP may provide a protective "shell" around the pharmaceutically active agent thereby shielding it and hindering, or preventing, its interaction with other molecules in the cell surface or molecules important for its activity.

The LAP may also comprise a sequence which has at least 50%, 60%, 70%, 80%, 90%, 95% or 99% identity with a LAP sequence of Figure 1, using the default parameters of the BLAST computer program provided by HGMP, thereto.

The proteolytic cleavage site may comprise any aggrecanase specific cleavage site which is cleavable by an aggrecanase. An aggrecanase cleavage site may comprise a number of amino acid residues recognisable by an aggrecanase. Moreover, the amino

acids of the aggrecanase site may be linked by one or more peptide bonds which are cleavable, proteolytically, by aggrecanase.

Aggrecanases which may cleave the aggrecanase site include, but are not limited to ADAMTS-4 (aggrecanase-1), ADAMTS-5 (aggrecanase-2) and ADAMTS-11 (Tortorella, M.D., et al Osteoarthritis Cartilage, 2001. **9**(6): p. 539-552); Abbaszade, I., et al J Biol Chem, 1999. **274**(33): p. 23443-23450).

The sequences of the protein cleavage sites of ADAMTS-4 (aggrecanase-1) are shown in Figure 2. Suitable ADAMTS-4 sites include:

HNEFRQRETYMVF

DVQEFRGVTAVIR

The consensus ADAMTS-4 cleavage motif can be represented according to Hills et al (J. Biol. Chem. **282** 11101-11109 (2007)) as:

E-[AFVLMY]-$X_{(0,1)}$-[RK]-$X_{(2,3)}$-[ST]-[VYIFWMLA]

Preferably, the aggrecanase proteolytic cleavage site of the present invention is cleaved at sites of a disease diagnosed as arthritis or cancer which can be characterized by inflammation and/or tissue remodelling. More preferably, the aggrecanase proteolytic cleavage site of the present invention is cleaved by ADAMTS-4 (aggrecanase-1), ADAMTS-5 (aggrecanase-2) or ADAMTS-11.

The amino acid sequence of the aggrecanase cleavage site may include a sequence which has at least 50%, 60%, 70%, 80%, 90%, 95% or 99% identity, using the default parameters of the BLAST computer program provided by HGMP, thereto. Preferably, the nucleic acid sequence encoding the aggrecanase cleavage site comprises the minimum number of residues required for recognition and cleavage by an aggrecanase.

The present invention may further provide a "linker" peptide. Preferably the linker peptide is linked to the amino acid sequence of the proteolytic cleavage site. The linker peptide may be provided at the C terminal or N terminal end of the amino acid sequence encoding the proteolytic cleavage site. Preferably, the linker peptide is continuous with the amino acid sequence of the proteolytic cleavage site. The linker peptide may comprise the amino acid sequence GGGGS or a multimer thereof (for example a dimer, a trimer, or a tetramer), a suitable linker may be $(GGGGS)_3$, or a sequence of nucleotides which has at least 50%, 60%, 70%, 80%, 90%, 95% or 99% identity, using the default parameters of the BLAST computer program provided by HGMP, thereto.

The term "linker peptide" is intended to define any sequence of amino acid residues which preferably provide a hydrophilic region when contained in an expressed protein. Such a hydrophilic region may facilitate cleavage by an enzyme at the proteolytic cleavage site.

The term "latency" as used herein, may relate to a shielding effect which may hinder interaction between the fusion protein and other molecules in the cell surface. Alternatively the term latency may be used to describe a reduction in the activity (up to and including ablation of activity) of a molecule/agent associated with the fusion protein. The term latency may also relate to a stabilising effect of the fusion protein. The effect may be in full or partial, where a partial effect is sufficient to achieve the latency of the active agent.

The pharmaceutically active agent may be a pharmaceutically active protein which can include, but is not limited to, a growth factor (e.g. TGFβ, epidermal growth factor (EGF), platelet derived growth factor (PDGF), nerve growth factor (NGF), colony stimulating factor (CSF), hepatocyte growth factor, insulin-like growth factor, placenta growth factor); a differentiation factor; a cytokine e.g. an interleukin, (e.g. IL1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-19, IL-20, IL-21, IL-22, IL-23, IL-24, IL-25, IL-26, IL-27, IL-28, IL-29, IL-30, IL-31, IL-32 or IL-33 or an interferon (e.g. IFN-α, IFN-β and IFN-γ), tumour necrosis factor (TNF), IFN-γ inducing factor (IGIF), a bone

morphogenetic protein (BMP, e.g. BMP-1, BMP-2, BMP-3, BMP-4, BMP-4, BMP-5, BMP-6, BMP-7, BMP-8, BMP-9, BMP10, BMP-11, BMP-12, BMP-13); an interleukin receptor antagonist (e.g. IL-1ra, IL-1RII); a chemokine (e.g. MIPs (Macrophage Inflammatory Proteins) e.g. MIP1α and MIP1β; MCPs (Monocyte Chemotactic Proteins) e.g. MCP1, 2 or 3; RANTES (regulated upon activation normal T-cell expressed and secreted)); a trophic factor; a cytokine inhibitor; a cytokine receptor; a free-radical scavenging enzyme e.g. superoxide dismutase or catalase; a pro-drug converting enzyme (e.g. angiotensin converting enzyme, deaminases, dehydrogenases, reductases, kinases and phosphatases); a peptide mimetic; a protease inhibitor; a tissue inhibitor of metalloproteinases (TIMPs eg. TIMP1, TIMP2, TIMP3 or TIMP4) or a serpin (inhibitors of serine proteases). Preferably, the pharmaceutically active agent will be derived from the species to be treated e.g. human origin for the treatment of humans. Preferably, the pharmaceutically active agent is IFNβ, IL-4, or IL-1ra.

The interleukins and cytokines may be both anti-inflammatory or pro-inflammatory. Anti-inflammatory cytokines and certain interleukins, such as IL-4 and/or IL-10, are suitable for the treatment of arthritis, whereas pro-inflammatory cytokines and other interleukins, such as IL-1 and IL-2, are suitable for the treatment of cancer.

As used herein "peptide mimetics" includes, but is not limited to, agents having a desired peptide backbone conformation embedded into a non-peptide skeleton which holds the peptide in a particular conformation. Peptide mimetics, which do not have some of the drawbacks of peptides, are of interest in those cases where peptides are not suitable in medicine.

Peptide mimetics may comprise a peptide backbone which is of the L- or D-conformation. Examples of peptides mimetics include melanocortin, adrenocorticotrophin hormone (ACTH) and other peptide mimetic agents which play a role in the central nervous system, endocrine system in signal transduction and in infection and immunity.

The pharmaceutically active agent may comprise a chemical compound such as a chemotherapeutic agent or other synthetic drug. Alternatively, the pharmaceutically active agent may comprise an siRNA or a peptide nucleic acid (PNA) sequence e.g. a poly-lysine sequence which binds to nucleic acids and permeabilises lipid bilayers (Wyman et al., Biological Chemistry, **379**, 1045-1052 (1998)) or a KALA peptide which facilitates transfer through lipid bilayers (Wyman et al., Biochemistry, **36**, 3008-3017 (1997)) or a protein transduction domain (PTD) that enables polypeptides to enter cells via the plasma membrane (Pi et al Molecular Therapy **2**, 339-347 (2000)).

The term "associating with" in the context of the present invention is intended to include all means of association including, but not limited to, chemical cross-linking or peptide bond linkage.

In an alternative embodiment, the invention further provides the fusion protein of the present invention optionally in association with latent TGFβ binding protein (LTBP). Typically, the fusion protein is covalently linked to LTBP to form a complex. Preferably, the association is mediated by disulphide bond(s) between Cys No. 33 of LAP and the third 8 Cys residue of LTBP. The LTBP associated with the fusion protein may include, but is not limited to, LTBP 1, 2, 3 or 4 (Kanzaki et al., Cell, **61**, 1051-1061 (1990); Tsuji et al., Proc. Natl. Acad. Sci. USA, **87**, 8835-8839 (1990); Moren et al., J. Biol. Chem. **269**, 32469-32478 (1994); Yin et al., J. Biol. Chem. **270**, 10147-10160 (1995); Gibson et al., Mol. Cell. Biol. **15**, 6932-6942 (1995); Saharinen et al., J. Biol. Chem. **273**, 18459-18469 (1998)), or fragments of LTBP such as that containing the third 8 Cys repeat, or homologues having a sequence of amino acids or nucleotides which has at least 50%, 60%, 70%, 80%, 90%, 95% or 99% identity, using the default parameters of the BLAST computer program provided by HGMP, to that of LTBP.

Cleavage of LTBP may release the fusion protein from the LTBP complex. Enzymes which may cleave LTBP in this manner include, but are not limited to, thrombospondin (Schultz et al., The Journal of Biological Chemistry, **269**, 26783-26788 (1994); Crawford et al., Cell, **93**, 1159-1170 (1998)), transglutaminase (Nunes

et al., J. Cell, Biol. **136**, 1151-1163 (1997); Kojima et al., The Journal of Cell Biology, **121**, 439-448 (1993)) MMP9 and MMP2 (Yu and Stamenkovic, Genes and Dev, **14**, 163-176 (2000)).

The invention further provides nucleic acid encoding the fusion protein of the first aspect of the invention as defined above. A second aspect of the invention provides a nucleic acid construct comprising a first nucleic acid sequence encoding a pharmaceutically active agent, a second nucleic acid sequence encoding a LAP, wherein a nucleic acid sequence encoding an aggrecanase proteolytic cleavage site is provided between the first and second nucleic acid sequences.

The term "nucleic acid construct" generally refers to any length of nucleic acid which may be DNA, cDNA or RNA such as mRNA obtained by cloning or produced by chemical synthesis. The DNA may be single or double stranded. Single stranded DNA may be the coding sense strand, or it may be the non-coding or anti-sense strand. For therapeutic use, the nucleic acid construct is preferably in a form capable of being expressed in the subject to be treated.

Preferably, the first nucleic acid sequence encodes the protein IFNβ, IL-4 or IL-1ra. In one embodiment of the invention, the first nucleic acid sequence encodes IFNβ, IL-4 or IL-1ra from a mouse or a human .

The nucleic acid construct of the second aspect of the invention may be in the form of a vector, for example, an expression vector, and may include, among others, chromosomal, episomal and virus-derived vectors, for example, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculo-viruses, papova-viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. Generally, any vector suitable to maintain, propagate or express nucleic acid to express a polypeptide in a host, may be used for expression in this regard.

The invention further provides a protein encoded by the nucleic acid construct of the second aspect of the invention optionally in association with latent TGFβ binding protein (LTBP) described herein. Typically, the protein encoded by the nucleic acid construct is covalently linked to LTBP to form a complex. Preferably, the association is mediated by disulphide bond(s) between Cys No. 33 of LAP and the third 8 Cys residue of LTBP.

The nucleic acid construct of the second aspect of the invention preferably includes a promoter or other regulatory sequence which controls expression of the nucleic acid. Promoters and other regulatory sequences which control expression of a nucleic acid have been identified and are known in the art. The person skilled in the art will note that it may not be necessary to utilise the whole promoter or other regulatory sequence. Only the minimum essential regulatory element may be required and, in fact, such elements can be used to construct chimeric sequences or other promoters. The essential requirement is, of course, to retain the tissue and/or temporal specificity. The promoter may be any suitable known promoter, for example, the human cytomegalovirus (CMV) promoter, the CMV immediate early promoter, the HSV thymidinekinase, the early and late SV40 promoters or the promoters of retroviral LTRs, such as those of the Rous Sarcoma virus (RSV) and metallothionine promoters such as the mouse metallothionine-I promoter. The promoter may comprise the minimum comprised for promoter activity (such as a TATA elements without enhancer elements) for example, the minimum sequence of the CMV promoter.

Preferably, the promoter is contiguous to the first and/or second nucleic acid sequence.

As stated herein, the nucleic acid construct of the second aspect of the invention may be in the form of a vector. Vectors frequently include one or more expression markers which enable selection of cells transfected (or transformed) with them, and preferably, to enable a selection of cells containing vectors incorporating heterologous DNA. A suitable start and stop signal will generally be present.

One embodiment of the invention relates to a cell comprising the nucleic acid construct of the second aspect of the invention. The cell may be termed a "host" cell, which is useful for the manipulation of the nucleic acid, including cloning. Alternatively, the cell may be a cell in which to obtain expression of the nucleic acid. Representative examples of appropriate host cells for expression of the nucleic acid construct of the invention include virus packaging cells which allow encapsulation of the nucleic acid into a viral vector; bacterial cells, such as *Streptococci, Staphylococci, E.coli, Streptomyces* and *Bacillus Subtilis*; single cells, such as yeast cells, for example, *Saccharomyces Cerevisiae*, and *Aspergillus* cells; insect cells such as *Drosophila S2* and *Spodoptera* Sf9 cells, animal cells such as CHO, COS, C127, 3T3, PHK.293, and Bowes Melanoma cells and other suitable human cells; and plant cells e.g. *Arabidopsis thaliana.*

Induction of an expression vector into the host cell can be affected by calcium phosphate transfection, DEAE-dextran mediated transfection, microinjection, cationic – lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction, infection or other methods. Such methods are described in many standard laboratory manuals, such as Sambrook et al, Molecular Cloning, a Laboratory Manual, Second Edition, Coldspring Harbor Laboratory Press, Coldspring Harbor, N.Y. (1989).

Mature proteins can be expressed in host cells, including mammalian cells such as CHO cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can be employed to produce such proteins using RNAs derived from the nucleic acid construct of the third aspect of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook et al, Molecular Cloning, a Laboratory Manual, Second Edition, Coldspring Harbor Laboratory Press, Coldspring Harbor, N.Y. (1989).

Proteins can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulphate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phoshocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite

chromatography, high performance liquid chromatography, lectin and/or heparin chromatography. For therapy, the nucleic acid construct e.g. in the form of a recombinant vector, may be purified by techniques known in the art, such as by means of column chromatography as described in Sambrook et al, Molecular Cloning, a Laboratory Manual, Second Edition, Coldspring Harbor Laboratory Press, Coldspring Harbor, N.Y. (1989).

According to a third aspect of the invention, there is provided a composition in accordance with the first aspect of the invention for use in the treatment of arthritis or cancer. This aspect of the invention therefore extends to and includes a method for the treatment or arthritis or cancer comprising the administration to a subject of a composition comprising a fusion protein comprising a latency associated peptide (LAP) connected by an aggrecanase proteolytic cleavage site to a pharmaceutically active agent.

The present invention provides a composition as described above for use in the treatment of arthritis or cancer. Arthritis defines a group of disease conditions (or arthropathies) where damage is caused to the joints of the body and includes osteoarthritis (also known as degenerative joint disease) which can occur following trauma to the joint, following an infection of the joint or as a result of aging. Other forms of arthritis include rheumatoid arthritis and psoriatic arthritis, which are autoimmune diseases, and septic arthritis is caused by infection in the joints. Cancer defines a group of diseases characterized by an abnormal proliferation of cells in the body, which can be defined as tumors, for example glioblastoma. Glioblastoma is also sometimes referred to as Grade 4 astrocytoma.

In a fourth aspect, the invention provides a nucleic acid sequence in accordance with the second aspect of the invention for use in the treatment of arthritis or cancer. This aspect therefore extends to and includes a method for the treatment of arthritis or cancer comprising the administration to a subject a nucleic acid construct of the second aspect of the invention. Where the nucleic acid construct is used in the therapeutic method of the invention, the construct may be used as part of an expression construct, e.g. in the form of an expression vector such as a plasmid or

virus. In such a method, the construct may be administered intravenously, intradermally, intramuscularly, orally or by other routes.

The nucleic acid construct of the second aspect of the invention, and proteins derived therefrom, may be employed alone or in conjunction with other compounds, such as therapeutic compounds, e.g. anti-inflammatory drugs, cytotoxic agents, cytostatic agents or antibiotics. The nucleic acid constructs and proteins useful in the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

As used herein, the term "treatment" includes any regime that can benefit a human or a non-human animal. The treatment of "non-human animals" extends to the treatment of domestic animals, including horses and companion animals (e.g. cats and dogs) and farm/agricultural animals including members of the ovine, caprine, porcine, bovine and equine families. The treatment may be in respect of any existing condition or disorder, or may be prophylactic (preventive treatment). The treatment may be of an inherited or an acquired disease. The treatment may be of an acute or chronic condition. Preferably, the treatment is of a condition/disorder associated with inflammation. The first nucleic acid sequence of the nucleic acid construct of the third aspect of the invention may encode a protein for use in the treatment of the disorder, including, but not limited to osteoarthritis, scleroderma, renal disease, rheumatoid arthritis, inflammatory bowel disease, multiple sclerosis, atherosclerosis, cancer, or any inflammatory disease.

The nucleic acid construct of the second aspect of the invention may be used therapeutically in a method of the invention by way of gene therapy. Alternatively, protein encoded by the nucleic acid construct may be directly administered as described herein.

Administration of the nucleic acid construct of the second aspect may be directed to the target site by physical methods. Examples of these include topical administration of the "naked" nucleic acid in the form of a vector in an appropriate vehicle, for example, in solution in a pharmaceutically acceptable excipient, such as phosphate

buffered saline, or administration of a vector by physical method such as particle bombardment according to methods known in the art.

Other physical methods for administering the nucleic acid construct or proteins of the third aspect of the invention directly to the recipient include ultrasound, electrical stimulation, electroporation and microseeding. Further methods of administration include oral administration or administration through inhalation.

Particularly preferred is the microseeding mode of delivery which is a system for delivering genetic material into cells *in situ* in a patient. This method is described in US Patent No. 5697901.

The nucleic acid construct according to the second aspect of the invention may also be administered by means of delivery vectors. These include viral delivery vectors, such as adenovirus, retrovirus or lentivirus delivery vectors known in the art.

Other non-viral delivery vectors include lipid delivery vectors, including liposome delivery vectors known in the art.

Administration may also take place via transformed host cells. Such cells include cells harvested from the subject, into which the nucleic acid construct is transferred by gene transfer methods known in the art. Followed by the growth of the transformed cells in culture and grafting to the subject.

As used herein the term "gene therapy" refers to the introduction of genes by recombinant genetic engineering of body cells (somatic gene therapy) for the benefit of the patient. Furthermore, gene therapy can be divided into *ex vivo* and *in vivo* techniques. *Ex vivo* gene therapy relates to the removal of body cells from a patient, treatment of the removed cells with a vector i.e., a recombinant vector, and subsequent return of the treated cells to the patient. *In vivo* gene therapy relates to the direct administration of the recombinant gene vector by, for example, intravenous or intravascular means.

Preferably the method of gene therapy of the present invention is carried out *ex vivo*.

Preferably in gene therapy, the expression vector of the present invention is administered such that it is expressed in the subject to be treated. Thus for human gene therapy, the promoter is preferably a human promoter from a human gene, or from a gene which is typically expressed in humans, such as the promoter from human CMV.

For gene therapy, the present invention may provide a method for manipulating the somatic cells of human and non-human mammals.

The present invention also provides a gene therapy method which may involve the manipulation of the germ line cells of a non-human mammal.

The present invention therefore provides a method for providing a human with a therapeutic protein comprising introducing mammalian cells into a human, the human cells having been treated *in vitro* to insert therein a nucleic acid construct according to the second aspect of the invention.

Each of the individual steps of the *ex vivo* somatic gene therapy method are also covered by the present invention. For example, the step of manipulating the cells removed from a patient with the nucleic acid construct of the third aspect of the invention in an appropriate vector. As used herein, the term "manipulated cells" covers cells transfected with a recombinant vector.

Also contemplated is the use of the transfected cells in the manufacture of a medicament for the treatment of arthritis or cancer.

The present invention may also find application in veterinary medicine for treatment/prophylaxis of domestic animals including horses and companion animals (e.g. cats and dogs) and farm animals which may include mammals of the ovine, porcine, caprine, bovine and equine families.

The present invention also relates to compositions comprising the nucleic acid construct or proteins of the first or second aspects of the invention. Therefore, the

fusion protein or nucleic acid constructs of the present invention may be employed in combination with the pharmaceutically acceptable carrier or carriers. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, liposomes, water, glycerol, ethanol and combinations thereof.

The pharmaceutical compositions may be administered in any effective, convenient manner effective for treating a patients disease including, for instance, administration by oral, topical, intravenous, intramuscular, intranasal, or intradermal routes among others. In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

For administration to mammals, and particularly humans, it is expected that the daily dosage of the active agent will be from 0.01mg/kg body weight, typically around 1mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual which will be dependent on factors including the age, weight, sex and response of the individual. The above dosages are exemplary of the average case. There can, of course, be instances where higher or lower dosages are merited, and such are within the scope of this invention

A sixth aspect of the invention provides a fusion protein comprising a LAP and an aggrecanase proteolytic cleavage site wherein the fusion protein is associated with a pharmaceutically active agent. The pharmaceutically active agent may be as described above. In some embodiments of this aspect of the invention, the pharmaceutically active agent may be an siRNA or PNA molecule.

The invention further provides a nucleic acid construct encoding the fusion protein of the sixth aspect of the invention. The nucleic acid construct preferably comprises a nucleic acid sequence encoding a LAP adjacent a nucleic acid sequence encoding an aggrecanase proteolytic cleavage site. Preferably, the nucleic acid sequence encoding a LAP is suitably operably linked to a nucleic acid sequence encoding an aggrecanase proteolytic cleavage site.

The invention further provides the fusion protein of the sixth aspect of the invention optionally in association with latent TGFβ binding protein (LTBP) described herein.

The fusion protein of the sixth aspect of the invention may be associated with the pharmaceutically active agent by means of a peptide bond linkage. Alternatively, the fusion protein may be associated with the pharmaceutically active agent by means of a chemical linkage e.g. by cross-linking the fusion protein to a chemical compound such as a chemotherapeutic agent, synthetic drug or PNA.

Preferably, the pharmaceutically active agent is linked to the C-terminal end of the amino acid sequence of the proteolytic cleavage site in the fusion protein of the seventh aspect of the invention. More preferably, the pharmaceutically active agent is continuous with the C-terminal residue of the amino acid sequence of the aggrecanase proteolytic cleavage site.

The fusion protein, and associated pharmaceutically active agent of the sixth aspect of the invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds, e.g. anti-inflammatory drugs, cytotoxic agents, cytostatic agents or antibiotics. Such administration may be simultaneous, separate or sequential. The components may be prepared in the form of a kit which may comprise instructions as appropriate.

Preferably, the fusion protein and associated pharmaceutically active agent of the sixth aspect of the invention are directly administered to a patient as described herein.

The present invention also relates to compositions comprising the fusion protein and associated pharmaceutically active agent of the sixth aspect of the invention. Therefore, the fusion protein and associated pharmaceutically active agent may be employed in combination with the pharmaceutically acceptable carrier or carriers. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, liposomes, water, glycerol, polyethylene glycol, ethanol and combinations thereof.

The pharmaceutical compositions may be administered in any effective, convenient manner effective for treating a disease of a patient including, for instance, administration by oral, topical, intravenous, intramuscular, intranasal, or intradermal routes among others. In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

A seventh aspect of the invention provides a kit of parts comprising a fusion protein of the first aspect of the invention, a nucleic acid construct of the second aspect of the invention, or a fusion protein and associated pharmaceutically active agent according to the sixth aspect of the invention, and an administration vehicle including, but not limited to, tablets for oral administration, inhalers for lung administration and injectable solutions for intravenous administration.

An eighth aspect of the invention provides a process for preparing the fusion protein, of the first aspect of the invention comprising production of the fusion protein recombinantly by expression in a host cell, purification of the expressed fusion protein and association of the pharmaceutically active agent to the purified fusion protein by means of peptide bond linkage, hydrogen or salt bond or chemical cross linking. In some embodiments of this aspect of the invention where the pharmaceutically active agent is a peptide, the fusion protein could be prepared using hydrogen or salt bonds where the peptide is capable or multimerisation, for example dimerisation or trimerisation.

A ninth aspect of the invention provides a process for preparing a nucleic acid construct of the second aspect of the invention comprising ligating together nucleic acid sequences encoding a latency associated peptide, an aggrecanase cleavage sequence, and a pharmaceutically active agent, optionally including a linker sequence on either side of the aggrecanase cleavage site.

A preferred embodiment of the present invention provides a method of providing latency to a pharmaceutically active agent which is a cytokine, preferably interferon or an interleukin, the method comprising constructing a fusion protein having a latency associated peptide, preferably from TGFβ, associated with an aggrecanase

proteolytic cleavage site, preferably an ADAM-TS4 cleavage site, and the pharmaceutically active agent. Preferably, the pharmaceutically active agent is followed by the proteolytic cleavage site and the LAP as follows: LAP-cleavage site-active agent .

All preferred features of the second and subsequent aspects of the invention are as for the first aspect *mutatis mutandis*.

The present invention will now be described by way of example only with reference to the accompanying figures wherein:

FIGURE 1 shows amino acid sequences of the precursor domain of TGFβ 1, 2 and 3 (human, Hu), TGFβ 4 (chicken, Ck), TGFβ (frog, Fg). Arrows indicate the position of the proteolytic processing resulting in cleavage of the signal peptide of TGFβ1 and of the mature TGFβs. N-linked glycosylation sites are underlined, as is the integrin cellular recognition sequence (Roberts and Sporn, Peptide Growth Factors and their Receptors: Sporn, MB and Roberts, AB, Springer-Verlag, Chapter 8, 422 (1996));

FIGURE 2 shows a multiple sequence alignment of the ADAMTS-4 epitope sequences with corresponding average percentage of phagemid cleavage and the derived ADAMTS-4 cleavage motif. Predominant amino acids found at a frequency of greater than 40% in a particular position are illustrated with a black background, in contrast to related amino acids which are shown with a grey background (reproduced from Hills et al J. Biol. Chem. **282** 11101-11109 (2007)).

FIGURE 3 shows the development of ADAMTS-4 sensitive LAP-IL-4. Western blot of supernatants of transiently transfected 293T cells probed with anti-IL-4 antibody. Serum free supernatants were collected 48 hrs after transfection and incubated overnight at 37C without (-) or with (+) recombinant MMP-1 or aggrecanase (both kindly provided by H. Nagase, Kennedy Institute). Note specificity of cleavage LAP-MMP IL-4 is only

cleaved by MMP-1 whilst the Agg1 and Agg2 constructs are cleaved by ADAMTS-4 but not MMP-1. Agg1 and Agg2 correspond to the sequences B05 and B06 from Hills et al. (J. Biol. Chem 282(15):11101-11109; (2007)).

FIGURES 4 and 5 show the full open reading frame and DNA sequence of both aggrecanase constructs. The GGGS sequences before and after the aggrecanase site are italicised, the EcoR1 and Not1 cloning sites are bold and underlined. The aggrecanase sites are in bold. The mIL-4 has a poly Histidine tail and tagged epitope at the COOH end. Figure 4 shows the construct with the B06 cleavage site and Figure 5 shows the construct with the B05 cleavage site.

The invention is now described with reference to the following non-limiting examples;

Experiment 1: Development of ADAMTS-4 sensitive LAP-IL-4

The aggrecanase cleavage sites B05 and B06 reported by Hills et al (J. Biol. Chem. **282**: 11101-11109 (2007)) were flanked on either side with the linker $(GGGGS)_3$ using overlapping oligodeoxynucleotides that were cloned between the EcoR1 and Not1 sites of the LAP-IL-4 construct (see Figure 4 or Figure 5).

To clone the aggrecanase cleavage sites into LAP-mIL-4, the following oligonucleotides coding for the cleavage sites B05 and B06 were purified by HPLC for cloning.

B06: DVQEFRGVTAVIR with GGGGS at either end

Aat II

1. sense (B06)5' aattc GGA GGC GGG GGT TCA GAC GTC CAA GAA TTC CGC GGC GTC ACA GCT GTG ATC CGT GGA GGC GGG GGT TCA gc

2. antisense (B06)5' ggc cgc TGA ACC CCC GCC TCC ACG GAT CAC AGC TGT GAC GCC GCG GAA TTC TTG GAC GTC TGA ACC CCC GCC TCC g

B05: HNEFRQRETYMVF with GGGGS at either end

BamH1

3. sense(B05) 5' aattc GGA GGC GGG GGA TCC CAC AAC GAG TTC CGA CAG CGG GAG ACA TAT ATG GTC TTC GGA GGC GGG GGT TCA gc

4. antisense (B05) 5' ggc cgc TGA ACC CCC GCC TCC GAA GAC CAT ATA TGT CTC CCG CTG TCG GAA CTC GTT GTG GGA TCC CCC GCC TCC g

To prepare the final construct, the LAP-mIL4 construct was cut with EcoR1 and Not1 and the large fragment produced was purified. After annealing the oligonucleotides the cleavage sequences were cloned into them. Positive clones were sent for DNA sequencing. B05 has new Aat II site, and B06 has a BamH1 site.

CLAIMS

1. A fusion protein comprising a latency associated peptide (LAP) and a pharmaceutically active agent in which the LAP and the pharmaceutically active agent are connected by an amino acid sequence comprising an aggrecanase proteolytic cleavage site.

2. A fusion protein as claimed in claim 1, in which the latency associated peptide (LAP) is the precursor domain of a TGFβ.

3. A fusion protein as claimed in claim 2, in which the latency associated peptide (LAP) is the precursor domain of TGFβ-1, -2, -3, -4 or -5.

4. A fusion protein as claimed in any one of claims 1 to 3, in which the aggrecanase cleavage site is cleaved by ADAMTS-4 (aggrecanase-1), ADAMTS-5 (aggrecanase-2) or ADAMTS-11.

5. A fusion protein as claimed in claim 4 in which the aggrecanase cleavage site is cleaved by ADAMTS-4 (aggrecanase-1).

6. A fusion protein as claimed in claim 5, in which the aggrecanase cleavage site is defined by the consensus sequence E-[AFVLMY]-$X_{(0,1)}$-[RK]-$X_{(2,3)}$-[ST]-[VYIFWMLA].

7. A fusion protein as claimed in claim 6, in which the aggrecanase cleavage site is:

HNEFRQRETYMVF, or

DVQEFRGVTAVIR

8. A fusion protein as claimed in any one of claims 1 to 7, in which a peptide linker sequence is present adjacent to the aggrecanase cleavage site.

9. A fusion protein as claimed in claim 8, in which the peptide linker sequence is GGGGS, or a multimer thereof.

10. A fusion protein as claimed in any one of claims 1 to 9, in which the pharmaceutically active agent is a protein.

11. A fusion protein as claimed in claim 10, in which the pharmaceutically active protein is a growth factor, differentiation factor, cytokine, a chemokine, a trophic factor, a cytokine inhibitor, a cytokine receptor, a free-radical scavenging enzyme, a pro-drug converting enzyme, a peptide mimetic, a protease inhibitor, a tissue inhibitor of metalloproteinases or a serine protease inhibitor.

12. A fusion protein as claimed in claim 11, in which the cytokine is an interleukin or an interferon.

13. A fusion protein as claimed in any one of the preceding claims, in which the fusion protein is in association with latent TGFβ binding protein.

14. A fusion protein comprising a LAP and an aggrecanase proteolytic cleavage site wherein the fusion protein is associated with a pharmaceutically active agent.

15. A fusion protein as claimed in any one of claims 1 to 14 for use in the treatment of arthritis or cancer.

16. A pharmaceutical composition comprising a fusion protein of any one claims 1 to 14.

17. A nucleic acid construct comprising a first nucleic acid sequence encoding a pharmaceutically active agent, a second nucleic acid sequence encoding a LAP, wherein a nucleic acid sequence encoding an aggrecanase proteolytic cleavage site is provided between the first and second nucleic acid sequences.

18. A nucleic acid construct as claimed in claim 17, in which the aggrecanase cleavage site is cleaved by ADAMTS-4 (aggrecanase-1), ADAMTS-5 (aggrecanase-2) or ADAMTS-11.

19. A nucleic acid construct as claimed in claim 17 or claim 18, in which the pharmaceutically active agent is a growth factor, a differentiation factor, a cytokine, a chemokine, a trophic factor, a cytokine inhibitor, a cytokine receptor, a free-radical scavenging enzyme, a pro-drug converting enzyme, a peptide mimetic, a protease inhibitor, a tissue inhibitor of metalloproteinases or a serine protease inhibitor.

20. A vector comprising a nucleic acid construct of anyone of claims 17 to 19.

21. A fusion protein encoded by the nucleic acid construct of any one of claims 17 to 19.

22. A cell comprising a vector of claim 21 or a nucleic acid construct of any one of claims 17 to 19.

23. A nucleic acid sequence as claimed in any one of claims 17 to 19 or a cell as claimed in claim 23 for use in the treatment of arthritis or cancer.

24. A pharmaceutical composition comprising a nucleic acid construct as claimed in any one of claims 17 to 19 or a cell as claimed in claim 22.

25. The use of a fusion protein as defined in any one of claims 1 to 14 in the manufacture of a medicament for the treatment of arthritis or cancer.

26. The use of a nucleic acid sequence as defined in any one of claims 17 to 19 or a cell as defined in claim 22 in the manufacture of a medicament for the treatment of arthritis or cancer.

27. A method for the treatment or arthritis or cancer comprising the administration to a subject of a composition comprising a fusion protein comprising a latency associated peptide (LAP) and a pharmaceutically active agent in which the LAP and

the pharmaceutically active agent are connected by an amino acid sequence comprising an aggrecanase proteolytic cleavage site.

28. A method for the treatment of arthritis or cancer comprising the administration to a subject a nucleic acid construct as claimed in any one of claims 17 to 19 or a cell as claimed in claim 22.

29. A kit of parts comprising a fusion protein as claimed in nay one claims 1 to 14, or a nucleic acid construct as claimed in any one of claims 17 to 19, or a cell as claimed in claim 22 and an administration vehicle.

30. A process for preparing the fusion protein as claimed in nay one of claims 1 to 14, comprising production of the fusion protein recombinantly by expression in a host cell, purification of the expressed fusion protein and association of the pharmaceutically active agent to the purified fusion protein by means of peptide bond linkage, hydrogen or salt bond, or chemical cross linking.

31. A process for preparing a nucleic acid construct of any one of claims 17 to 19 comprising ligating together nucleic acid sequences encoding a latency associated peptide, an aggrecanase cleavage sequence, and a pharmaceutically active agent, optionally including a linker sequence on either side of the aggrecanase cleavage site.

FIG. 1

```
                1               20                  40                  60
Hu TGF-β1 MPPSGLRLLPLLLPLLWLLV-LTPGPPAAGLSTCKTIDMELVKRKRIEAIRGQILSKLRLASPPSQGE-VP-PGP
Hu TGF-β2 MHYCVLSAFLILH    LVTVAL-------SLSTCSTLDMDQFMRKRIEAIRGQILSKLKLTSPP---EDYPEPEE
Hu TGF-β3 MKMHLQRALVVLALLHFATVSL-------SLSTCTTLDFGHIKKKRVEAIRGQILSKLRLTSPP---EPTV-MTH
Ck TGF-β4 ------------------------------------------------------------------------
Fg TGF-β5 MEV-------LWMLLVLLV-LHLSSLANSLSTCKAVDMEEVRKRRIEAIRGQILSKLKLDKIPDVDS-EK-MTV
          +         +       + +       +++*    +       + ++++++++++ +    +

                80                  100                 120
Hu TGF-β1 LPEAVLALYNSTRDRVAGESAEPE-PEP-------EADYYAKEVIRVLMV----ETHNEIYDKFKQSTHSIYMFF
Hu TGF-β2 VPPEVISIYNSTRDLL--QEKASR-RAAACERERSDEEYYAKEVYKIDMPPFFPS-ENAIPPTFYRPY-FRIVRF
Hu TGF-β3 VPYQVLALYNSTRELL--EEHGER-KEEGCTQENTESEYYAKEIHKFDMIQGLAE-HNELAVCPKGIT-SKVFRF
Ck TGF-β4 ----------------M--DPMSIGPK--SCG-----------------GSPW-RPP-GTAPWSIG-SR--RA
Fg TGF-β5 PSEAIF-LYNSTLE-VIREKATRE-EEEHVGHDQNIQDYYAKQVYRF----ESITELEDHEFKFK----------F
                 ++++                ++++

          140                 160                 180
Hu TGF-β1 NTSEL-----RE-AVPEPVLLS-RAELRLLRLKL----KV-EQHVELYQ-----KYSNNSWRYLSNRLLAPSDSPE
Hu TGF-β2 DVSA----------MEKNASNLV-KAEFRVFRLQNPK-ARVPEQRIELYQILKSKDLISPTQRYIDSKVVKTRAEGE
Hu TGF-β3 NVSS----------VEKNRTNLF-RAEFRVLRVPNPS-SKRNEQRIELFQILRP-DEHIAKQRYIGGKNLPTRGTAE
Ck TGF-β4 TASSSCSTSSRVRAEVGGRALLHRAELRHLRQKAAADSAGTEQRLELYQGYG-----NASWRYLHGRSVRATSDDE
Fg TGF-β5 NASHV-----RENVGMN-SLLH-HAELRMYK-KQTD--KNMDQRMELFW--KYQENGTTHSRYLESKYITPVTDQE
          +               +   ++ +       +   ++              ++          +
```

```
                200                 220                     240
Hu TGF-β1 WLSFDVTGVVRQWLSRGGEIEGFRLSAHCSG--------DSRDNTLQVDIN-GFTTGR------RGDLATI-----
Hu TGF-β2 WLSFDVTDAVHEWLHHKDRWLGFKISLHCPCCTFVPSNNYIIPNKSEELEARFA-GIDGTSTYTSGDQKTIKSTRK
Hu TGF-β3 WLSFDVTDTVREWLLRRESNLGLEISIHCPCHTFQP-NGDILENIHEVMEIKFK-GVDNEDDHGRGDLGRLK---K
Ck TGF-β4 WLSFDVTDAVHQWLSGSELIGVFKLSVHCPCEMGPG-HADEMRISIEGFEQQ------------RGDMQSIA---K
Fg TGF-β5 WNSFDVTKTVNEWLKRAEENEQFGLQPAGKG------------PTPQAKD----IDIEGFPAL-RGDLASL-SSK
          + +++++  +  ++              * *                                  ++

                    260                     280                 300
Hu TGF-β1 ----HGMWRPFLLLMATPLERA-QK--LQSS---RHRRALDTNYCFSST--EKNCCVRQLYIDFRKDLGWKWIHEP
Hu TGF-β2 KNSGKT---PHLLLMLLPSYRL-ESQ----QTNRRKKRALDAAYCFRNV--QDNCCLRPLYIDFKRDLGWKWIHEP
Hu TGF-β3 QKDNN--N-PHLILMMIPPHRL-DNPGQGGQ---RKKRALDTNYCFRNL--EENCCVRPLYIDFRQDLGWKWVHEP
Ck TGF-β4 -KHRR--V-PYVLAMALPAERANE---LHSA---RRRRDLDTDYCFGPGTDEKNCCVRPLYIDFRKDLQWKWIHEP
Fg TGF-β5 ENT-----KPYL--MITSMPAERIDTVT SS---RKKRGVGQEYCFGNN--GPNCCVKPLYINFRKDLGWKWIHEP
                  +    +                  +  +     +*+       +** + +++ +  ++ +++ +++

                320                 340                 360                 380         390
Hu TGF-β1 KGYHANFCLGPCPYIWSLDTQYSKVLALYNQHNPGASAAPCCVPQALEPLPIVYYVGRKPKVEQLSNMIVRSCKCS
Hu TGF-β2 KGYNANFCAGACPYLWSSDTQHSRVLSLYNTINPEASASPCCVSQDLEPLTILYYIGKIPKIEQLSNMIVKSCKCS
Hu TGF-β3 KGYYANFCSGPCPYLRSADTTHSTVLGLYNTLNPEASASPCCVPQDLEPLTILYYVGRIPKVEQLSNMVVKSCKCS
Ck TGF-β4 KGYMANFCMGPCPYIWSADTQYIKVLALYNQHNPGASAAPCCVPQILDPLPIIYYVGRNVRVEQLSNMVVRACKCS
Fg TGF-β5 KGYEANYCLGNCPYIWSMDTQYSKVLSLYNQHNPGASISPCCVPDVLEPLPIIYYVGRIAKVEQLSNMVVRSCNCS
          +++ ++ * + *++  + ++    ++ +++  ++ ++  +**+   + ++ + ++ +     ++++++ +  * *+
```

FIG. 1 CONT'D

Phage Code	Peptide Sequence	Avg. % Cleaved
D04	----MMF-KGQR-VERVLT	99.7
B05	---HNFF-RQRETYMVF--	99.6
B07	--NWQEFQ-AKRSVAY---	99.5
D03	----LEL-ESN-SVIMRWP	99.4
C02	--DYMEV-RRQMSMQM---	99.4
F03	---ALEM-RAAD-VEYHF-	99.3
D01	VEHLMEVQRKT-TW-----	99.2
C05	---GVEV-KRQLSYHYM--	99.1
B08	----QELVGANIETYML--	99.0
C06	--QQMEVSRYV-QYKW---	98.9
F05	---LQSF-RQAP-VDIWW-	98.9
H04	----QEL-RGKISIQPFK-	98.9
E03	---QQEYMSGQYDIIF---	98.8
D07	---SMEFA-ATVTSTFE--	98.7
C07	---EQQL-KGRQTHIII--	98.5
B02	----MEL-KGQ-TDMFYII	98.5
F01	---GAYAV-GRWSYVDA--	98.3
B01	---GQFATSPKITIHK---	98.2
B06	--DVQEF-RGV-TAVIR--	98.0
A05	----HEA-RTVSTTYLML-	98.0
C04	---YMEN-RGSTTVFFN--	97.9
F06	----QELIGSY-SVMPTN-	97.9
E01	-HYYMEATRDIEMV-----	97.9
A06	----NEAHSSGITIMLR--	97.8
H01	-DHPMEF-RSKITMK----	97.8
A01	--TFAEM-KGTVSYAL---	97.7
F04	-GVHMESMRRY-TVI----	97.7
E05	---FQEYTG---TYDIMDP	97.6
A04	-FQAVEASK-TLHFW----	97.6
D02	---YLETSRTY-TTVWP--	97.5
F08	-TDYLEV-RSQPIIY----	97.2
B03	-TFEQEV-RAPN-ISW---	97.1
B04	---PQEVQGIAVEWV----	96.9
F10	----AEA-KAS-TLHVYLM	96.8
D09	--DYMEVVGNKISYI----	96.7
F09	--VIMEAV-GRKTILQ---	96.5
D10	--FQAEAARAV-TYSS---	96.5
F07	-EDYVYV-KDVGTTN----	96.4
A08	----QEY-KAHHSYKLMS-	96.2
C03	---YNEY-RATPTFAVV--	95.8
H03	-----EYFHANTTRIVQS-	95.4
G02	---ALEASRFI-SWDIN--	95.3
C01	----WEAVAAP--IMHTWV	94.3
E02	---FQEL-KAAETFWM---	94.1
F02	--NTLYAV-APPVIYV---	90.4
G01	-FQPYEVQRIT-TVM----	89.9
A02	--KPMESGRRT-TVYY---	88.5
G05	----MEF-KGALQYRLQP-	82.9
H02	---PQEV-KQARKWIIE--	79.4
A07	-YRQQEVKRHIQIV-----	34.2

E-[AFVLMY]-X(0,1)-[RK]-X(2,3)-[ST]-[VYIFWMLA]

FIG. 2

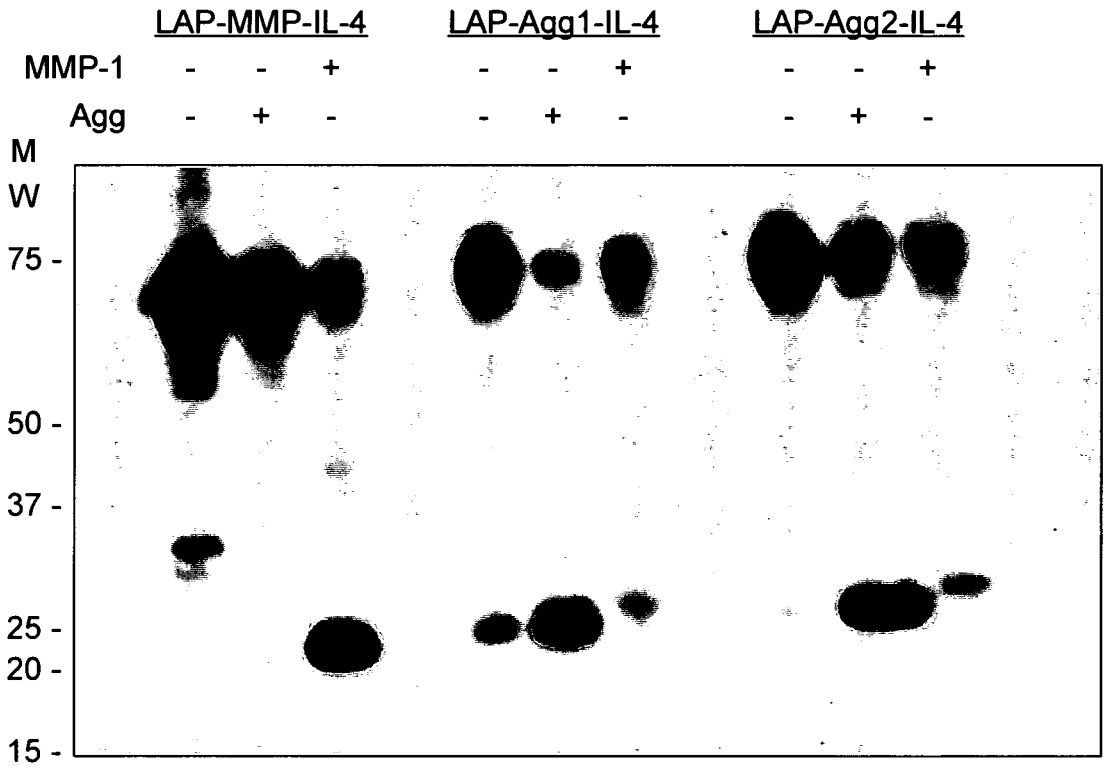
LAP-MMP-IL-4
LAP-Agg1-IL-4
LAP-Agg2-IL-4
MMP-1 - - + - - + - - +
Agg - + - - + - - + -
M
W
75 -
50 -
37 -
25 -
20 -
15 -

FIG. 3

FIG. 4

```
                9          18          27          36          45          54
5'  ATG CCG CCC TCC GGG CTG CGG CTG CTG CCG CTG CTG CTA CCG CTG CTG TGG CTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Met Pro Pro Ser Gly Leu Arg Leu Leu Pro Leu Leu Leu Pro Leu Leu Trp Leu

               63          72          81          90          99         108
    CTG GTG CTG ACG CCT GGC CCG CCG GCC GCG GGA CTA TCC ACC TGC AAG ACT ATC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Leu Val Leu Thr Pro Gly Pro Pro Ala Ala Gly Leu Ser Thr Cys Lys Thr Ile

              117         126         135         144         153         162
    GAC ATG GAG CTG GTG AAG CGG AAG CGC ATC GAG GCC ATC CGC GGC CAG ATC CTG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Asp Met Glu Leu Val Lys Arg Lys Arg Ile Glu Ala Ile Arg Gly Gln Ile Leu

              171         180         189         198         207         216
    TCC AAG CTG CGG CTC GCC AGC CCC CCG AGC CAG GGG GAG GTG CCG CCC GGC CCG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Ser Lys Leu Arg Leu Ala Ser Pro Pro Ser Gln Gly Glu Val Pro Pro Gly Pro

              225         234         243         252         261         270
    CTG CCC GAG GCC GTG CTC GCC CTG TAC AAC AGC ACC CGC GAC CGG GTG GCC GGG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Leu Pro Glu Ala Val Leu Ala Leu Tyr Asn Ser Thr Arg Asp Arg Val Ala Gly

              279         288         297         306         315         324
    GAG AGT GCA GAA CCG GAG CCC GAG CCT GAG GCC GAC TAC TAC GCC AAG GAG GTC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Glu Ser Ala Glu Pro Glu Pro Glu Pro Glu Ala Asp Tyr Tyr Ala Lys Glu Val

              333         342         351         360         369         378
    ACC CGC GTG CTA ATG GTG GAA ACC CAC AAC GAA ATC TAT GAC AAG TTC AAG CAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Thr Arg Val Leu Met Val Glu Thr His Asn Glu Ile Tyr Asp Lys Phe Lys Gln

              387         396         405         414         423         432
    AGT ACA CAC AGC ATA TAT ATG TTC TTC AAC ACA TCA GAG CTC CGA GAA GCG GTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Ser Thr His Ser Ile Tyr Met Phe Phe Asn Thr Ser Glu Leu Arg Glu Ala Val

              441         450         459         468         477         486
    CCT GAA CCC GTG TTG CTC TCC CGG GCA GAG CTG CGT CTG CTG AGG AGG CTC AAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Glu Pro Val Leu Leu Ser Arg Ala Glu Leu Arg Leu Leu Arg Arg Leu Lys

              495         504         513         522         531         540
    TTA AAA GTG GAG CAG CAC GTG GAG CTG TAC CAG AAA TAC AGC AAC AAT TCC TGG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Leu Lys Val Glu Gln His Val Glu Leu Tyr Gln Lys Tyr Ser Asn Asn Ser Trp

              549         558         567         576         585         594
    CGA TAC CTC AGC AAC CGG CTG CTG GCA CCC AGC GAC TCG CCA GAG TGG TTA TCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Arg Tyr Leu Ser Asn Arg Leu Leu Ala Pro Ser Asp Ser Pro Glu Trp Leu Ser

              603         612         621         630         639         648
    TTT GAT GTC ACC GGA GTT GTG CGG CAG TGG TTG AGC CGT GGA GGG GAA ATT GAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Phe Asp Val Thr Gly Val Val Arg Gln Trp Leu Ser Arg Gly Gly Glu Ile Glu

              657         666         675         684         693         702
    GGC TTT CGC CTT AGC GCC CAC TGC TCC TGT GAC AGC AGG GAT AAC ACA CTG CAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Gly Phe Arg Leu Ser Ala His Cys Ser Cys Asp Ser Arg Asp Asn Thr Leu Gln
```

```
        711         720         729         738         747         756
GTG GAC ATC AAC GGG TTC ACT ACC GGC CGC CGA GGT GAC CTG GCC ACC ATT CAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Val Asp Ile Asn Gly Phe Thr Thr Gly Arg Arg Gly Asp Leu Ala Thr Ile His

        765         774         783         792         801         810
GGC ATG AAC CGG CCT TTC CTG CTT CTC ATG GCC ACC CCG CTG GAG AGG GCC CAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Gly Met Asn Arg Pro Phe Leu Leu Leu Met Ala Thr Pro Leu Glu Arg Ala Gln

        819         828         837         846         855         864
CAT CTG CAA AGC GAA TTC GGA GGC GGG GGT TCA GAC GTC CAA GAA TTC CGC GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
His Leu Gln Ser Glu Phe Gly Gly Gly Gly Ser Asp Val Gln Glu Phe Arg Gly

        873         882         891         900         909         918
GTC ACA GCT GTG ATC CGT GGA GGC GGG GGT TCA GCG GCC GCA CAT ATC CAC GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Val Thr Ala Val Ile Arg Gly Gly Gly Gly Ser Ala Ala Ala His Ile His Gly

        927         936         945         954         963         972
TGC GAC AAA AAT CAC TTG AGA GAG ATC ATC GGC ATT TTG AAC GAG GTC ACA GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Cys Asp Lys Asn His Leu Arg Glu Ile Ile Gly Ile Leu Asn Glu Val Thr Gly

        981         990         999        1008        1017        1026
GAA GGG ACG CCA TGC ACG GAG ATG GAT GTG CCA AAC GTC CTC ACA GCA ACG AAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Glu Gly Thr Pro Cys Thr Glu Met Asp Val Pro Asn Val Leu Thr Ala Thr Lys

       1035        1044        1053        1062        1071        1080
AAC ACC ACA GAG AGT GAG CTC GTC TGT AGG GCT TCC AAG GTG CTT CGC ATA TTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Asn Thr Thr Glu Ser Glu Leu Val Cys Arg Ala Ser Lys Val Leu Arg Ile Phe

       1089        1098        1107        1116        1125        1134
TAT TTA AAA CAT GGG AAA ACT CCA TGC TTG AAG AAG AAC TCT AGT GTT CTC ATG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Tyr Leu Lys His Gly Lys Thr Pro Cys Leu Lys Lys Asn Ser Ser Val Leu Met

       1143        1152        1161        1170        1179        1188
GAG CTG CAG AGA CTC TTT CGG GCT TTT CGA TGC CTG GAT TCA TCG ATA AGC TGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Glu Leu Gln Arg Leu Phe Arg Ala Phe Arg Cys Leu Asp Ser Ser Ile Ser Cys

       1197        1206        1215        1224        1233        1242
ACC ATG AAT GAG TCC AAG TCC ACA TCA CTG AAA GAC TTT CTG GAA AGC CTA AAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Thr Met Asn Glu Ser Lys Ser Thr Ser Leu Lys Asp Phe Leu Glu Ser Leu Lys

       1251        1260        1269        1278        1287        1296
AGC ATC ATG CAA ATG GAT TAC TCG CAC CAT CAC CAC CAC CCA TTG AGG GCC CTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ser Ile Met Gln Met Asp Tyr Ser His His His His His Pro Leu Arg Ala Leu

       1305        1314        1323        1332        1341        1350
TTC TAT AGT GTC ACC TAA ATG CTA GAG CTC GCT GAT CAG CCT CGA CTG TGC CTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Phe Tyr Ser Val Thr *** Met Leu Glu Leu Ala Asp Gln Pro Arg Leu Cys Leu

       1359
CTA GTT GCC AGC C 3'
--- --- --- --- -
Leu Val Ala Ser
```

FIG. 4 CONT'D

FIG. 5

```
                9          18          27          36          45          54
5'  ATG CCG CCC TCC GGG CTG CGG CTG CTG CCG CTG CTG CTA CCG CTG CTG TGG CTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Met Pro Pro Ser Gly Leu Arg Leu Leu Pro Leu Leu Leu Pro Leu Leu Trp Leu

               63          72          81          90          99         108
    CTG GTG CTG ACG CCT GGC CCG CCG GCC GCG GGA CTA TCC ACC TGC AAG ACT ATC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Leu Val Leu Thr Pro Gly Pro Pro Ala Ala Gly Leu Ser Thr Cys Lys Thr Ile

              117         126         135         144         153         162
    GAC ATG GAG CTG GTG AAG CGG AAG CGC ATC GAG GCC ATC CGC GGC CAG ATC CTG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Asp Met Glu Leu Val Lys Arg Lys Arg Ile Glu Ala Ile Arg Gly Gln Ile Leu

              171         180         189         198         207         216
    TCC AAG CTG CGG CTC GCC AGC CCC CCG AGC CAG GGG GAG GTG CCG CCC GGC CCG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Ser Lys Leu Arg Leu Ala Ser Pro Pro Ser Gln Gly Glu Val Pro Pro Gly Pro

              225         234         243         252         261         270
    CTG CCC GAG GCC GTG CTC GCC CTG TAC AAC AGC ACC CGC GAC CGG GTG GCC GGG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Leu Pro Glu Ala Val Leu Ala Leu Tyr Asn Ser Thr Arg Asp Arg Val Ala Gly

              279         288         297         306         315         324
    GAG AGT GCA GAA CCG GAG CCC GAG CCT GAG GCC GAC TAC TAC GCC AAG GAG GTC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Glu Ser Ala Glu Pro Glu Pro Glu Pro Glu Ala Asp Tyr Tyr Ala Lys Glu Val

              333         342         351         360         369         378
    ACC CGC GTG CTA ATG GTG GAA ACC CAC AAC GAA ATC TAT GAC AAG TTC AAG CAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Thr Arg Val Leu Met Val Glu Thr His Asn Glu Ile Tyr Asp Lys Phe Lys Gln

              387         396         405         414         423         432
    AGT ACA CAC AGC ATA TAT ATG TTC TTC AAC ACA TCA GAG CTC CGA GAA GCG GTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Ser Thr His Ser Ile Tyr Met Phe Phe Asn Thr Ser Glu Leu Arg Glu Ala Val

              441         450         459         468         477         486
    CCT GAA CCC GTG TTG CTC TCC CGG GCA GAG CTG CGT CTG CTG AGG AGG CTC AAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Pro Glu Pro Val Leu Leu Ser Arg Ala Glu Leu Arg Leu Leu Arg Arg Leu Lys

              495         504         513         522         531         540
    TTA AAA GTG GAG CAG CAC GTG GAG CTG TAC CAG AAA TAC AGC AAC AAT TCC TGG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Leu Lys Val Glu Gln His Val Glu Leu Tyr Gln Lys Tyr Ser Asn Asn Ser Trp

              549         558         567         576         585         594
    CGA TAC CTC AGC AAC CGG CTG CTG GCA CCC AGC GAC TCG CCA GAG TGG TTA TCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Arg Tyr Leu Ser Asn Arg Leu Leu Ala Pro Ser Asp Ser Pro Glu Trp Leu Ser

              603         612         621         630         639         648
    TTT GAT GTC ACC GGA GTT GTG CGG CAG TGG TTG AGC CGT GGA GGG GAA ATT GAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Phe Asp Val Thr Gly Val Val Arg Gln Trp Leu Ser Arg Gly Gly Glu Ile Glu

              657         666         675         684         693         702
    GGC TTT CGC CTT AGC GCC CAC TGC TCC TGT GAC AGC AGG GAT AAC ACA CTG CAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
    Gly Phe Arg Leu Ser Ala His Cys Ser Cys Asp Ser Arg Asp Asn Thr Leu Gln
```

```
        711         720         729         738         747         756
GTG GAC ATC AAC GGG TTC ACT ACC GGC CGC CGA GGT GAC CTG GCC ACC ATT CAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Val Asp Ile Asn Gly Phe Thr Thr Gly Arg Arg Gly Asp Leu Ala Thr Ile His

        765         774         783         792         801         810
GGC ATG AAC CGG CCT TTC CTG CTT CTC ATG GCC ACC CCG CTG GAG AGG GCC CAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Gly Met Asn Arg Pro Phe Leu Leu Leu Met Ala Thr Pro Leu Glu Arg Ala Gln

        819         828         837         846         855         864
CAT CTG CAA AGC GAA TTC GGA GGC GGG GGA TCC CAC AAC GAG TTC CGA CAG CGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
His Leu Gln Ser Glu Phe Gly Gly Gly Gly Ser His Asn Glu Phe Arg Gln Arg

        873         882         891         900         909         918
GAG ACA TAT ATG GTC TTC GGA GGC GGG GGT TCA GCG GCC GCA CAT ATC CAC GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Glu Thr Tyr Met Val Phe Gly Gly Gly Gly Ser Ala Ala Ala His Ile His Gly

        927         936         945         954         963         972
TGC GAC AAA AAT CAC TTG AGA GAG ATC ATC GGC ATT TTG AAC GAG GTC ACA GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Cys Asp Lys Asn His Leu Arg Glu Ile Ile Gly Ile Leu Asn Glu Val Thr Gly

        981         990         999        1008        1017        1026
GAA GGG ACG CCA TGC ACG GAG ATG GAT GTG CCA AAC GTC CTC ACA GCA ACG AAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Glu Gly Thr Pro Cys Thr Glu Met Asp Val Pro Asn Val Leu Thr Ala Thr Lys

       1035        1044        1053        1062        1071        1080
AAC ACC ACA GAG AGT GAG CTC GTC TGT AGG GCT TCC AAG GTG CTT CGC ATA TTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Asn Thr Thr Glu Ser Glu Leu Val Cys Arg Ala Ser Lys Val Leu Arg Ile Phe

       1089        1098        1107        1116        1125        1134
TAT TTA AAA CAT GGG AAA ACT CCA TGC TTG AAG AAG AAC TCT AGT GTT CTC ATG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Tyr Leu Lys His Gly Lys Thr Pro Cys Leu Lys Lys Asn Ser Ser Val Leu Met

       1143        1152        1161        1170        1179        1188
GAG CTG CAG AGA CTC TTT CGG GCT TTT CGA TGC CTG GAT TCA TCG ATA AGC TGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Glu Leu Gln Arg Leu Phe Arg Ala Phe Arg Cys Leu Asp Ser Ser Ile Ser Cys

       1197        1206        1215        1224        1233        1242
ACC ATG AAT GAG TCC AAG TCC ACA TCA CTG AAA GAC TTT CTG GAA AGC CTA AAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Thr Met Asn Glu Ser Lys Ser Thr Ser Leu Lys Asp Phe Leu Glu Ser Leu Lys

       1251        1260        1269        1278        1287        1296
AGC ATC ATG CAA ATG GAT TAC TCG CAC CAT CAC CAC CAC CCA TTG AGG GCC CTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Ser Ile Met Gln Met Asp Tyr Ser His His His His His Pro Leu Arg Ala Leu

       1305        1314
TTC TAT AGT GTC ACC TAA 3'
--- --- --- --- --- ---
Phe Tyr Ser Val Thr ***
```

FIG. 5 CONT'D