

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 August 2006 (03.08.2006)

PCT

(10) International Publication Number
WO 2006/081430 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2006/002951

(22) International Filing Date: 27 January 2006 (27.01.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/647,013 27 January 2005 (27.01.2005) US

(71) Applicant (for all designated States except US): **FIVE PRIME THERAPEUTICS, INC.** [US/US]; 1650 Owens Street,, Suite 200, San Francisco, CA 94158-2261 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **HALENBECK, Robert, Forgan** [US/US]; 59 Golden Hinde Boulevard, San Rafael, CA 94903 (US). **BOSCH, Elizabeth** [US/US]; 23047 Cricket Hill Road, Cupertino, CA 95014 (US). **LIN-NEMANN, Thomas** [DE/US]; 1338 Baker Street #7, San Francisco, CA 94115 (US). **LEE, Ernestine** [US/US]; 226 Arlington Avenue, Kensington, CA 94707 (US).

(74) Agent: **GARRETT, Arthur, S.**; Finnegan, Henderson, Farabow, Garrett & Dunner LLP, 901 New York Avenue, NW, Washington, DC 20001-4413 (US).

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

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

(54) Title: LEADER SEQUENCES FOR DIRECTING SECRETION OF POLYPEPTIDES AND METHODS FOR PRODUCTION THEREOF

(57) Abstract: The present invention provides leader sequences that are useful for the production of heterologous secretable polypeptides; heterologous secreted polypeptides; nucleic acid constructs that encode such leader sequences and heterologous secreted polynucleotides; vectors that contain such nucleic acid constructs; recombinant host cells that contain such nucleic acid constructs; vectors, polypeptides, and methods of making and using such secreted polypeptides with such heterologous leader sequences.



WO 2006/081430 A2

LEADER SEQUENCES FOR DIRECTING SECRETION OF POLYPEPTIDES AND METHODS FOR PRODUCTION THEREOF

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit (pursuant to 35 U.S.C. § 119(e)) of provisional application 60/647,013, filed in the United States Patent and Trademark Office on January 27, 2005, the disclosures of which are hereby incorporated by reference.

FIELD OF THE INVENTION

[0002] The present invention relates to leader sequences that are useful for production of heterologous secretable polypeptides; heterologous secreted polypeptides; nucleic acid constructs encoding such leader sequences; nucleic acid constructs encoding such heterologous secretable polypeptides; vectors that contain such nucleic acid constructs; recombinant host cells that contain such nucleic acid constructs, vectors and polypeptides; and methods of making such secretable polypeptides with such heterologous leader sequences; and methods of using such secretable polypeptides.

BACKGROUND OF THE INVENTION

[0003] Proteins are the most prominent biomolecules in living organisms. In addition to being structural components and catalysts, they play crucial roles in regulatory processes. The cooperation of numerous cellular and extracellular proteins controls and affects the regulation of cell proliferation and metabolism. For example, many signal transduction pathways that affect physiological responses operate through proteins *via* intermolecular interactions.

[0004] Extracellular proteins, sometimes referred to as “secreted proteins,” or “secretable proteins” herein, often function as intercellular signal communicators. In this role, they act as ligands. Their counterpart, the membrane-associated receptors that have extracellular, intracellular, or cytoplasmic domains, transmit extracellular signals into the cells when ligand/receptor binding events take place on the cell surfaces.

[0005] While receptors often make potentially important therapeutic targets, secretable proteins are of particular interest as therapeutic agents. Because of their frequent involvement in signaling or hormonal pathways, secretable proteins tend to exhibit high and specific biological activities (*Schoen*, 1994). For example, secretable proteins have been reported to control or regulate physiological processes such as differentiation and

proliferation, blood clotting and thrombolysis, somatic growth and cell death, as well as various immune responses (*Id.*). Significant resources and research efforts have been expended to discovering new secretable proteins and investigating their regulatory functions. Some of these secretable proteins, including cytokines and peptide hormones, have been manufactured and used as therapeutic agents (*Zavyalov et al. 1997*), but they constitute a minority amongst the thousands of proteins that are expected to be secreted and potentially efficacious therapeutically.

[0006] Typically, a secretable protein is expressed as a full-length polypeptide, sometimes referred to as a “protein precursor,” which is then processed in the Endoplasmic Reticulum (ER) and the Golgi in the post-translational phase. During this phase, a signal peptidase cleaves off a characteristic hydrophobic amino acid sequence at the N-terminus, a sequence that is generally referred to as a “signal peptide” (SP) or a “secretory leader sequence.” A typical SP is about 16 to 30 amino acid residues in length. The resulting polypeptide sans the SP is then exported to outside the cell. The resulting polypeptide is called a “mature protein” or a “secreted polypeptide.” And compared to the original secretable protein, this mature protein lacks the signal peptide sequence. Some proteins do not have an SP at the N-terminus, such as some of the members in the fibroblast growth factor family.

[0007] Naturally-occurring secretable proteins are expressed in varying amounts depending on their physiological roles *in vivo*. Many of them, under the regulation of their natural or endogenous SP, are expressed in quantities that are too low to be used commercially. It would therefore be advantageous if nucleic acid constructs and methods are devised to enable the production of secretory proteins *in vivo* or *in vitro* to meet the manufacturing needs for therapeutic applications.

SUMMARY OF THE INVENTION

[0008] The present invention provides nucleic acid and polypeptide constructs for producing proteins in higher yields than when such proteins are produced from sequences that comprise their endogenous signal peptide. The present invention also provides vectors, host cells and methods for producing proteins in higher yields than when such proteins are produced from DNA sequences that encode the protein with its endogenous signal peptide or without an endogenous signal peptide; the higher yield being achieved either by replacing the endogenous secretory leader sequence with an heterologous secretory leader sequence of the invention, or by adding a heterologous secretory leader sequence of the invention to a

protein that would otherwise not contain a leader sequence. Accordingly, the present invention provides polypeptide and polynucleotide constructs where the polypeptides and polynucleotides are modified, such as to form a fusion molecule with a fusion partner. The fusion molecules of the invention may be prepared by any conventional technique.

[0009] Accordingly, the present invention comprises the following embodiments:

[0010] 1. A heterologous polypeptide comprising a secretory leader and a second polypeptide, wherein the secretory leader is operably linked to the N-terminal of the second polypeptide, wherein the secretory leader is not so linked to the second polypeptide in nature, and wherein the secretory leader comprises a leader sequence of a secretable protein.

[0011] 2. The heterologous polypeptide of 1, wherein the second polypeptide is a secretable protein selected from collagen type IX alpha 1 chain, long splice form, alpha-2-antiplasmin precursor (alpha-2-plasmin inhibitor), trinucleotide repeat containing 5, ARMET protein, calumenin, COL9A1 protein, NBL1, PACAP protein, alpha-1B-glycoprotein precursor (alpha-1-B glycoprotein), similar to brain-specific angiogenesis inhibitor 2 precursor, SPOCK2, protein disulfide-isomerase (EC 5341) ER60 precursor, serine (or cysteine) proteinase inhibitor, clade A (alpha-1), GM2 ganglioside activator precursor, coagulation factor X precursor, secreted phosphoprotein 1 (osteopontin, bone sialoprotein 1), Vitamin D-binding protein precursor, interleukin 6 (interferon, beta 2), orosomucoid 1 precursor, hemopexin, glycoprotein hormones, alpha polypeptide precursor, kininogen 1, prolyl 4-hydroxylase, beta subunit, proopiomelanocortin, prostaglandin D2 synthase 21 kDa, alpha-2-glycoprotein 1, zinc, chromogranin A, cystatin M precursor, clusterin isoform 1, inter-alpha (globulin) inhibitor H1, leukemia inhibitory factor (cholinergic differentiation factor), lumican, secretoglobin, family 2A, member 2, nov precursor, reticulocalbin 1 precursor, reticulocalbin 2, EF-hand calcium binding domain, gastric intrinsic factor (vitamin B synthesis), cerberus 1, lipocalin 2 (oncogene 24p3), interleukin 18 binding protein isoform C precursor, cell growth regulator with EF hand domain 1, leukocyte immunoglobulin-like receptor, subfamily A, spondin 2, extracellular matrix protein, transmembrane protein 4, sparco/osteonectin, cwcv and kazal-like domain proteoglycan, Rho GTPase activating protein 25 isoform b, dickkopf homolog 3, ameloblastin precursor, chorionic gonadotropin, beta polypeptide 8 precursor, multiple coagulation factor deficiency 2, similar to common salivary protein 1, hypothetical protein FLJ32115, oncoprotein-induced transcript 3, hypothetical protein MGC40499, interleukin

18 binding protein isoform A precursor, interleukin 1 receptor antagonist isoform 1 precursor, WFIKKN2 protein, similar to hypothetical protein 9330140G23, and SEQ ID NOs: 2-3, 9, 19, 22, 26, 28, 31, 37, 41, 47, 54, 57, 62, 68, 75, 79, 82, 86, 88, 94, 97, 102, 104, 107, 111, 116, 120, 127, 131, 137, 140, 145, 147, 153, 159, 167, 175, 177, 181, 185, 189, 191, 196, 200, 207, 209, 215, 218, 222, 227, 232, 235, 239, 241, 245, 248, and 254.

[0012] 3. The heterologous polypeptide of 1, wherein the secretory leader comprises an amino acid sequence selected from SEQ ID NOs: 20 – 21, 23-25, 27, 32-36, 38-40, 48-53, 76-78, 80-81, 83-85, 87, 95-96, 103, 108-110, 112-115, 117-119, 121-126, 128-130, 132-136, 138-139, 141-144, 154-158, 160-166, 178-180, 186-188, 197-199, 210-214, 223-226, 233-234, 240, and 246-247.

[0013] 4. The heterologous polypeptide of 1, wherein the second polypeptide is selected from a secretable polypeptide, an extracellular portion of a transmembrane protein, and a soluble receptor.

[0014] 5. The heterologous polypeptide of 4, wherein the secretable polypeptide is selected from a growth factor, a cytokine, a lymphokine, an interferon, a hormone, a stimulatory factor, an inhibitory factor, a soluble receptor, and splice variants thereof.

[0015] 6. A secretory leader comprising a leader amino acid sequence selected from the leader sequences of the secretable polypeptides of Table 1 and the secretory leaders listed in Table 2.

[0016] 7. The secretory leader of 6, the amino acid sequence of which is selected from the amino acid sequences of Appendix A, the amino acids residues of SEQ ID NOs: 1, 4-8, 10-18, 20-21, 23-25, 27, 29-30, 32-36, 38-40, 42-46, 48-53, 55-56, 58-61, 63-67, 69-74, 76-78, 80-81, 83-85, 87, 89-93, 95-96, 98-101, 103, 105-106, 109-110, 112-115, 117-119, 121-126, 128-130, 132-136, 138-139, 141-144, 146, 148-152, 154-158, 160-166, 168-174, 176, 178-180, 182-184, 186-188, 190, 192-195, 197-199, 201-206, 208, 210-214, 216-217, 219-221, 223-226, 228-231, 233-234, 236-238, 240, 242-244, 246-247, 249-253, and 255-256.

[0017] 8. The heterologous polypeptide of 1, further comprising a fusion partner.

[0018] 9. The heterologous polypeptide of 8, wherein the fusion partner is a polymer.

- [0019] 10. The heterologous polypeptide of 9, wherein the polymer is a third molecule, and wherein the third molecule is selected from polyethylene glycol and all or part of human serum albumin, fetuin A, fetuin B and Fc.
- [0020] 11. An isolated nucleic acid molecule comprising a polynucleotide sequence selected from: (1) a polynucleotide sequence encoding an amino acid sequence of a heterologous polypeptide according to any one of 1-5 and 8-10; (2) a polynucleotide encoding an amino acid sequence of a secretory leader according to any one of 6-7.
- [0021] 12. A nucleic acid molecule encoding a heterologous polypeptide, comprising a first polynucleotide that encodes a secretory leader of any one of 6-7, a second polynucleotide that encodes a second polypeptide, wherein the first polynucleotide and the second polynucleotide are operably linked to facilitate secretion of the heterologous polypeptide from a cell, and wherein the first and second polynucleotide are not so linked in nature.
- [0022] 13. The nucleic acid of claim 12, wherein the second polypeptide is selected from a secretable polypeptide, an extracellular portion of a transmembrane protein, and a soluble receptor.
- [0023] 14. The nucleic acid molecule of claim 12, further comprising a third polynucleotide, wherein the third polynucleotide is a Kozak sequence or a fragment thereof that is situated at its 5' end.
- [0024] 15. The nucleic acid molecule of 14, further comprising a fourth polynucleotide, wherein the fourth polynucleotide comprises a restriction enzyme-cleavable sequence at its 3' end.
- [0025] 16. The nucleic acid molecule of 15, further comprising a fifth polynucleotide that encodes a tag.
- [0026] 17. The nucleic acid molecule of 16, wherein the tag is a purification tag.
- [0027] 18. The nucleic acid molecule of 16, wherein the tag is selected from V5, HisX6, HisX8, an avidin molecule, and a biotin molecule.
- [0028] 19. The nucleic acid molecule of 16, further comprising a sixth polynucleotide that encodes a second enzyme-cleavable sequence that can be cleaved by a second enzyme, wherein the second cleavable sequence is situated upstream of the tag if the tag is situated at the C-terminus of the heterologous polypeptide, or downstream of the tag if the tag is situated at the N-terminus of the heterologous polypeptide.

- [0029] 20. The nucleic acid molecule of 19, wherein the second enzyme is thrombin or TEV from a tobacco virus.
- [0030] 21. A vector comprising the nucleic acid molecule of any one of claims 11-20, further comprising an origin of replication and a selectable marker.
- [0031] 22. The vector of 21, wherein the origin of replication is selected from SV40 ori, Pol ori, EBNA ori, and pMB1 ori.
- [0032] 23. The vector of 21, wherein the selectable marker is an antibiotic resistance gene.
- [0033] 24. The vector of 23, wherein the antibiotic resistance is selected from puromycin resistance, kanamycin resistance, and ampicillin resistance.
- [0034] 25. A recombinant host cell comprising a cell and the heterologous polypeptide of any of 1-4 and 8-10, the nucleic acid molecule of any of 11-20, or the vector of any one of 21-24.
- [0035] 26. The recombinant host cell of 25, wherein the cell is a eukaryotic cell.
- [0036] 27. The recombinant host cell of 26, wherein the cell is a human cell.
- [0037] 28. A method of producing a secreted polypeptide, comprising:
- [0038] (a) providing the nucleic acid molecule of any of 11-20; and
- (b) expressing the nucleic acid molecule in an expression system.
- [0039] 29. The method of 28, wherein the expression system is a cellular expression system or a cell free expression system.
- [0040] 30. The method of 28, wherein the expression system is a cellular expression system and the cell is a mammalian cell.
- [0041] 31. The method of 30, wherein the mammalian cell is selected from a 293 cell line, a PER.C6® cell line, and a CHO cell line.
- [0042] 32. The method of 31, wherein the 293 cell is a 293-T cell or a 293-6E cell.

Description of the Figures

[0043] **FIGURE 1:** is an alignment of the amino acid sequences of: (a) a leader sequence of the present invention ("collagen_leader"); (b) a cDNA clone previously designated as MGC:21955 having an annotation of an unknown protein, and designated herein as CLN00517648; and (c) a publicly accessible sequence NP_001842_NM_001851, corresponding to collagen type IX alpha I chain, long form (*Homo Sapiens*). These

sequences all start with a methionine ("M") as amino acid residue 1 at the N terminus. This clone CLN00517648_5pv1 was sequenced and found to contain 253 amino acid residues.

[0044] **FIGURE 2:** is a Western blot showing expression of several secretable polypeptides of the invention in media conditioned by cultured 293-T cells, which are transfected with cDNAs encoding proteins of the invention, subcloned into a pTT5 vector (as described in greater detail in Examples 2-4). The construct expressing the secretable protein encoded by clone CLN00517648 demonstrated the highest level of protein secretion in the conditioned media. The amount of protein secreted into the conditioned media was compared to two standards: (1) V5-Hisx6 tagged Delta-like protein 1 extracellular protein (15 µl at 16, 66, and 266 ng/ml); and (2) V5-Hisx6 tagged CSF-1 Receptor extracellular domain (15 µl at 8, 33, and 133 ng/ml). These standards were mixed and loaded into the three right hand lanes at the designated concentrations.

[0045] **FIGURE 3:** is a diagrammatic representation of a starting vector pTT5 (4398 bps) provided by Dr. Yves Durocher (Durocher, 2002).

[0046] **FIGURE 4:** shows the sequence of Vector A, which is inserted into the pTT5 vector to replace the "ccdb" region for the purpose of this invention. Vector A includes from left to right: an EcoR I site; the open reading frame (ORF), or the gene of interest encoding the mature polypeptide, which is represented by "-----," a BamH1 site; a cleavable sequence exemplified by a sequence encoding a thrombin cleavage site; a tag exemplified by V5H8; and a linker sequence followed by a stop codon.

[0047] **FIGURE 5:** shows sequences for Vector B and Vector C. Vector B includes, from left to right: a Kozak sequence, a leader sequence ("SP") such as the collagen leader sequence of the present invention, an EcoR1 site, the ORF or the gene of interest encoding the mature polypeptide as represented by "-----," a BamH1 site, a tag such as V5H8, and a linker sequence followed by a stop codon. Vector C includes, from left to right: a Kozak sequence, a leader sequence ("SP") exemplified by the collagen leader sequence of the present invention, an EcoR1 site, the ORF or the gene of interest encoding the mature polypeptide as represented by "-----," a BamH1 site, a cleavable sequence exemplified by a sequence encoding thrombin, a tag such as V5H8, and a linker sequence followed by a stop codon.

[0048] **FIGURE 6:** shows sequences for Vector D and Vector E. Vector D includes, from left to right: an EcoR1 site, the ORF or the gene of interest encoding the mature polypeptide as represented by "-----," a BamH1 site, and an Fc domain sequence followed

by a stop codon. Vector E includes, from left to right: a Kozak sequence (“GCCGCCACC”), a signal peptide/leader sequence of the invention, an EcoR1 site, the ORF or the gene of interest encoding the mature polypeptide as represented by “-----,” a BamH1 site, and an Fc domain sequence followed by a stop codon.

[0049] **FIGURE 7:** is an example of a pTT2p vector for making stable puromycin-resistant cell lines. Specifically, the pTT2p vector includes, *inter alia*, murine polyoma signals to make an episomal pTT2-gateway vector.

[0050] **FIGURE 8:** shows an SDS-PAGE analysis of protein expression in CHO SOY medium, employing 28 of the secretable proteins described herein. The top two (2) panels show SDS-PAGE developed with Coomassie stain and the bottom two (2) panels show SDS-PAGE developed with silver stain. Table 3, columns 6-11, identifies the specific leader sequence represented in each SDS-PAGE lane. In the three right-hand lanes, a bovine serum albumin (BSA) standard was run at concentrations that reflect corresponding expression levels of 8, 16, and 32 milligrams/liter (mg/L), respectively.

[0051] **FIGURE 9:** shows an SDS-PAGE analysis of protein expression in CHO SOY medium, employing the secretable proteins of 29-56 as described herein. The top two (2) panels show SDS-PAGE developed with Coomassie stain and the bottom two (2) panels show SDS-PAGE developed with silver stain. Table 3, columns 6-11, identifies the specific secretable protein represented in each SDS-PAGE lane. A bovine serum albumin (BSA) standard was run at concentrations that reflect corresponding expression levels of 8, 16, and 32 milligrams/liter (mg/L).

[0052] **Table 1:** lists information regarding the secretable proteins from which the leader sequences of the invention are derived. Column 1 lists the internal designation identification numbers; column 2 lists the reference identification numbers; column 3 lists the identities of the secretable proteins.

[0053] **Table 2:** lists information regarding the leader sequences of the invention. Column 1 lists the internal designation identification numbers; column 2 lists the SEQ ID NOs. for the leader sequences (P1); column 3 lists the reference identification numbers; column 4 lists the leader sequence types, *i.e.*, full length versus alternative leader sequences; and column 5 lists the secretable proteins from which the leader sequences are derived.

[0054] **Table 3:** summarizes the results obtained with the leader sequences of the current invention. Column 1 lists the clone designation identification numbers; column 2 lists the protein concentrations in micrograms/milliliter (μg/ml) as detected and measured

from the Coomassie-stained SDS-PAGE; column 3 ranks the expression levels as measured by Coomassie-stained SDS-PAGE, silver stained SDS-PAGE, or quantitative Western Blot using an Anti-V5 antibody relative to purified V5-tagged protein standards, of each construct on a scale of 1 to 56, from the lowest at 56 to the highest at 1; column 4 lists whether a band was detected using silver-stain developed SDS-PAGE; column 5 lists the molecular weights of the tested secretable proteins in Daltons; column 6 lists the gel numbers and lane numbers corresponding to Figures 8-9; column 7 lists the internal designations for the secretable proteins; column 8 lists protein identification numbers; column 9 lists the internal designation identification numbers; column 10 lists the source identification numbers; column 11 identifies the secretable proteins.

[0055] **Appendix A/Sequence Listing** lists the amino acid sequences of the leader sequences (P1) in Table 2.

Detailed Description of the Invention

[0056] To express and secrete the proteins of interest in larger quantities (*e.g.*, about 10% more, 20% more, 30% more, or a higher percentage more) than those obtained when the proteins are expressed and secreted from DNA sequences that encode their full-length amino acid sequence and contain their endogenous signal peptide, the inventors replaced their endogenous secretory leader sequence with that from another, *i.e.*, different or heterologous, secretable protein. The latter secretable protein of interest is typically one that is expressed and/or secreted at high levels ("high expressor protein" or "high secretor protein"), or moderately high levels ("moderate expressor protein" or "moderate secretor protein") under typical conditions for assaying protein expression and secretion, which are not limited to those described in detail in the Examples of the invention. In other words, if one were to express a panel of proteins (including but are not limited to those listed in this specification, in Appendix A, and those listed in Tables 1-3), and all were expressed under the same assay conditions, one would find that some proteins are expressed and/or secreted at higher levels than others. Accordingly, it is an aspect of the invention to recognize the differences in expression and secretion levels among the proteins of the invention, and take advantage of these recognized differences to further identify from the leader sequences those that are useful for improving the secretion and/or expression of otherwise low expressor proteins, or of proteins that are not secreted at the desirable levels.

[0057] Employing heterologous secretory leader sequences is further advantageous in that, during the secretion process, the resulting mature amino acid sequence of the

secretable polypeptide is not altered as the secretory leader sequence is removed in the endoplasmic reticulum (ER) or the Golgi. A secretory leader sequence of the invention serves to direct certain proteins to the ER. The ER separates the membrane-bounded proteins from all other types of proteins amongst those comprising the leader sequences. Each group is then separately moved to the Golgi apparatus. The Golgi apparatus then distributes the proteins to vesicles such as secretory vesicles, the cell membranes, the lysosomes, or other organelles.

[0058] Moreover, the addition of a heterologous secretory leader facilitates the expression and secretion of the extracellular domains of transmembrane proteins. An example of such a transmembrane protein is the Type II single transmembrane proteins (STM), the secretory leader of which is also the transmembrane domain, which must be removed before the protein becomes soluble and secreted.

[0059] Thus, to identify robust secretory leader sequence(s), which enhance or improve the secretion and expression of proteins relative to that achieved by the endogenous leader sequence, and which optionally can be used universally for making secretable proteins, many different secretable proteins have been cloned and expressed, as described herein. The expression and secretion levels of the cloned and expressed proteins in the supernatant of the mammalian 293 cells have also been measured, the results of which are shown in, for example, Example 1, Figures 8-9, and Table 3. Several high-expressor and high-secretor proteins were observed. The high-expressor proteins may or may not be the same as the high-secretor proteins for the purposes of this invention.

[0060] In one embodiment, a secretory leader sequence that is a part of the secretable protein collagen type IX alpha I chain, long form, has been identified. This particular leader sequence was selected to further examine its ability to promote expression and secretion when used as a heterologous and/or universal secretory leader sequence. The amino acid sequence of the secretory leader, which is part of the secretable protein collagen type IX alpha I chain, long form, is predicted to be MKTCWKIPVFFFVCSFLEPWASA (SEQ ID NO: 1). As further described herein, vectors were constructed to comprise this particular secretory leader. Using these vectors, several proteins were cloned without their own naturally-existing secretory leaders, yielding secretable proteins with a heterologous secretory leader sequence. The expression and secretion levels of these fusion proteins were found to be about 10%, about 20%, about 30%, about 40%, about 50%, about 60%,

about 70% or more higher than the expression or secretion levels as observed with their non-fusion counterparts.

[0061] The present invention may be more clearly understood in light of the following definitions. Generally, the terms used herein have their ordinary meanings and the meanings given them specifically below.

[0062] The terms “polynucleotide,” “nucleotide,” “nucleic acid,” “nucleotide molecule,” “nucleic acid molecule,” “nucleic acid sequence,” “polynucleotide sequence,” and “nucleotide sequence” are used interchangeably herein to refer to polymeric forms of nucleotides of any length. The polynucleotides can contain deoxyribonucleotides, ribonucleotides, and/or their analogs or derivatives. For example, nucleic acids can be naturally occurring DNA or RNA, or can be synthetic analogs of the naturally occurring DNA or RNA, as known in the art. The terms also encompass genomic DNA; genes; gene fragments; exons; introns; regulatory sequences or regulatory elements, such as promoters, enhancers, initiation and termination regions, other control regions, expression regulatory factors and expression controls; isolated DNA; and cDNA. In addition, the terms encompass mRNA, tRNA, rRNA, ribozymes, splice variants, antisense RNA, antisense conjugates, RNAi, siRNA and isolated RNAs. The terms also encompass recombinant polynucleotides, heterologous polynucleotides, branched polynucleotides, labeled polynucleotides, DNA/RNA hybrids, polynucleotide constructs, vectors comprising the subject nucleic acids, nucleic acid probes, primers and primer pairs. The terms comprise modified nucleic acid molecules, such as analogs of purines and pyrimidines, with alterations in the backbones; sugars; or heterocyclic bases, such as methylated nucleic acid molecules; peptide nucleic acids; and nucleic acid molecule analogs, which may be suitable as, for example, probes if they demonstrate superior stability and/or binding affinity under assay conditions. Analogous of purines and pyrimidines, including radiolabeled and fluorescent analogs, are known in the art. The polynucleotides can have any three-dimensional structure. The terms also encompass single-stranded, double-stranded and triple-helical molecules that are DNA, RNA, or hybrid DNA/RNA, and that may encode a full-length gene or a biologically active fragment thereof. Biologically active fragments of polynucleotides can encode the polypeptides herein, as well as anti-sense, ribozymes, or RNAi molecules. Thus, for example, the full length polynucleotides herein may be treated with enzymes, such as Dicer, to generate a library of short RNAi fragments, which are also within the scope of the present invention.

[0063] The terms “polypeptide,” “peptide,” and “protein,” used interchangeably herein, refer to a polymeric form of amino acids of any length. The amino acids can include naturally-occurring amino acids; coded and non-coded amino acids; chemically or biochemically modified, derivatized, or designer amino acids; amino acid analogs; peptidomimetics and depsipeptides; and polypeptides having modified, cyclic, bicyclic, depsicyclic, or depsibicyclic peptide backbones. The terms may also refer to conjugated proteins; fusion proteins, including, but not limited to, GST fusion proteins, fusion proteins with a heterologous amino acid sequence, fusion proteins with heterologous and homologous leader sequences, fusion proteins with or without N-terminal methionine residues; pegylated proteins; and immunologically tagged proteins. Also included in the terms are variations of naturally occurring proteins, where such variations are homologous or substantially similar to the naturally occurring proteins, as well as their corresponding homologs from different species. Variants of polypeptide sequences include insertions, additions, deletions, or substitutions when compared with the original polypeptides, but nonetheless retaining the same type of biological activity albeit possibly at a different level. The term also includes peptide aptamers.

[0064] A “secretory leader,” “signal peptide,” or a “leader sequence,” contains a sequence comprising amino acid residues that directs the intracellular trafficking of the polypeptide to which it is a part. Polypeptides contain secretory leaders, signal peptides or leader sequences, typically at their N-terminus. These polypeptides may also contain cleavage sites where the secretory leaders, signal peptides or leader sequences may be cleaved from the rest of the polypeptides by signal endopeptidases. Such polypeptides, after cleavage at the cleavage sites, generate mature polypeptides. Cleavage typically takes place during secretion or after the intact polypeptide has been directed to the appropriate cellular compartment.

[0065] According to the invention, a “high secretor signal peptide/secretory leader sequence” is one that (i) can be operably linked to a protein as an heterologous sequence, thereby replacing its endogenous signal peptide; and (ii) is capable of enhancing the level of secretion of the protein at least about 5 fold, when compared to the level of secretion that the protein exhibits when it carries its endogenous SP.

[0066] Also according to the invention, a “moderate secretor signal peptide/secretory leader sequence” is one that (i) can be operably linked to a protein as an heterologous sequence, thereby replacing its endogenous signal peptide; and (ii) is capable

of enhancing the level of secretion of the protein about 2 to 5 fold, when compared to the level of secretion that the protein exhibits when it carries its endogenous SP.

[0067] Further according to the invention, a "low secretor signal peptide/secretory leader" is one that (i) can be operably linked to a protein as an heterologous sequence, thereby replacing its endogenous signal peptide; and (ii) is capable of enhancing the level of secretion of the protein less than about 2 fold or does not enhance the level of secretion of the protein when compared to the level of secretion that the protein exhibits when it carries its endogenous SP.

[0068] Moreover, a secretory leader of the invention can also be added to a protein which is otherwise not predicted to be secreted via the ER-Golgi and does not have an endogenous signal peptide. In this case, the above definitions of "high/moderate/low secretor signal peptide/secretory leader sequence" are not applicable since there is no baseline secretion level for the protein that can be used for comparison purposes. In this case, the effect that the addition of the signal peptide/secretory leader sequence has on the secretion of an otherwise non-secretable protein will be compared among the resulting heterologous proteins.

[0069] For the purpose of this invention, the above definitions of "high/moderate/low secretor signal peptide/secretory leader sequence" relate only to the signal peptide (or secretory leader sequences). They do not relate to "high secretor proteins," "moderate secretor proteins" or "low secretor proteins". The proteins themselves were ranked as such on a basis of a relative scale that served to rank all the proteins of the invention (Tables 1-3 and Appendix A) relatively to each other, with regards to their own expression and secretion levels in either wheat germ extracts, or mammalian cells (see Examples 1-3 for detailed explanation).

[0070] A "secretable" protein is one capable of being directed to the ER, secretory vesicles, or the extracellular space by a secretory leader, signal peptide, or leader sequence. It may also be one that is released into the extracellular space without necessarily containing a signal sequence. If the secretable protein is one that is released into the extracellular space, it can undergo processing to produce a "mature" polypeptide." Proteins that contain transmembrane domains and typically remain inserted into the plasma membrane are considered, for the purposes of the invention, secretable proteins because they are also synthesized in the ER-Golgi, and some fragments or parts of such proteins can be released into the extracellular compartment, for example, by proteolytic cleavage. Thus, release into

the extracellular space can occur in multiple ways, including, for example, exocytosis and proteolytic cleavage.

[0071] The terms “mature protein” and “secreted protein” are used interchangeably herein, and refer to the form(s) of a secretable protein after it is secreted to the outside of the cell (for example, into the media conditioned by cells in culture). Typically, the mature protein has the amino acid sequence of the secretable protein sans the signal peptide. However, when a protein is expressed in nature or recombinantly, parts of the signal peptides are often not removed, resulting in a mature-protein mixture that may contain many forms of the mature protein, attached to varying lengths of the signal peptides. Thus, multiple “mature forms” can exist for a secretable protein depending on the specific amino acids cleaved off by the signal endopeptidase. Other proteases can also cleave off amino acids from a secretable protein, further adding to the heterogeneity of its “mature-protein.” The exact place where a signal peptide has been removed from a particular protein sample may be determined by N-terminal protein sequencing or otherwise by standard methods known to those skilled in the art.

[0072] A “biologically active” entity, or an entity having “biological activity,” is one that has the structural, regulatory, or biochemical functions of a naturally occurring molecule, or one that has the functions related to or associated with a metabolic or physiological process. A biologically active polynucleotide fragment or polypeptide fragment according to this invention is one that exhibits activities similar, but not necessarily identical, to the activities of the counterpart polynucleotide or polypeptide, to which the fragment is a part. Biological activities may include, but are not limited to, an improved desired activity and a decreased undesirable activity. For example, an entity demonstrates biological activity when it participates in molecular interactions with other molecules. An example of such an interaction is hybridization. Another example of such an interaction may be the exhibition of therapeutic effectiveness in alleviating a disease condition, or prophylactic effectiveness in inducing an immune response to the molecule. Another example of such an interaction may be the demonstration of potential uses as diagnostic tools in determining the presence of the molecule, for example, when the active fragment of a polynucleotide or a polypeptide is unique to the polynucleotide or the polypeptide, allowing the detection of the polynucleotide or the polypeptide by detecting fragment. A biologically active polypeptide or fragment thereof includes one that can participate in a biological reaction, for example, one that can serve as an epitope or immunogen to stimulate

an immune response, which includes but is not limited to the production of antibodies; or one that participates in signal transduction pathways by binding to receptors, proteins, or nucleic acids; or one that activates enzymes or substrates. Yet another example of such an interaction may be the suitability of using the polynucleotide molecule as a primer in PCR.

[0073] An "isolated" or "substantially isolated" polynucleotide or polypeptide, or a polynucleotide or polypeptide in "substantially pure form," in "substantially purified form," or as an "isolate," is one that is substantially free of the sequences with which it is associated in nature, or of other nucleic acid sequences that do not include a sequence or fragment of the subject polynucleotide or polypeptide. "Substantially free" means that less than about 10%, less than about 20%, less than about 30%, less than about 40%, or less than about 50%, of the composition is composed of the undesired materials.

[0074] "Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their desired function. Thus, a given promoter operably linked to a coding sequence is capable of effecting the expression of the coding sequence when the proper transcription factors and conditions are present. The promoter need not be contiguous with the coding sequence, so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between the promoter sequence and the coding sequence, as can translated introns, and the promoter sequence can still be considered "operably linked" to the coding sequence.

[0075] "Recombinant," when used to describe a nucleic acid molecule, means a polynucleotide of genomic, cDNA, viral, or synthetic origin which, by virtue of its origin or manipulation, is not associated with all or a portion of the polynucleotide with which it is associated in nature. The term "recombinant," when used to describe a protein or polypeptide, means a polypeptide produced by expression of a recombinant polynucleotide.

[0076] A "control element" refers to a polynucleotide sequence that aids in the expression of a coding sequence to which it is linked. The term may refer to promoters, transcription termination sequences, upstream regulatory domains, polyadenylation signals, and when appropriate, leader sequences and enhancers, which collectively provide for the transcription and translation of a coding sequence in a host cell.

[0077] A "promoter" as used herein refers to a DNA regulatory region capable of binding RNA polymerase in a mammalian cell and initiating transcription of a downstream (3' direction) coding sequence operably linked thereto. For purposes of the present

invention, a promoter sequence includes the minimum number of bases or elements required to initiate transcription of a gene of interest at a level detectable above background. Within the promoter sequence is a transcription initiation site, as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters often, but not always, contain “TATA” boxes and “CAT” boxes. Promoters further include those that are naturally contiguous to nucleic acid molecules and those that are not naturally contiguous to nucleic acid molecules. Additionally, promoters may include inducible promoters; conditionally active promoters, such as a cre-lox promoter; constitutive promoters; and tissue specific promoters.

[0078] A “selectable marker” refers to a gene that confers one or more phenotypes on a cell expressing the marker, such that the cell can be identified in appropriate conditions under which the phenotypes associated with the markers are manifested and observable. Generally, a selectable marker allows selection of transformed cells based on their ability to thrive in the presence or absence of one or more chemicals and/or other agents that inhibit an essential cell function. Suitable markers, therefore, include genes coding for proteins that confer drug resistance or sensitivity thereto, impart color to, or change the antigenic characteristics of those cells transfected with a molecule encoding the selectable marker, when the transfected cells are grown in an appropriate selective medium. For example, selectable markers include: cytotoxic markers and drug resistance markers, whereby cells are selected by their ability to grow on media containing one or more of the cytotoxins or drugs; auxotrophic markers by which cells are selected by their ability to grow on defined media with or without particular nutrients or supplements, such as thymidine and hypoxanthine; metabolic markers by which cells are selected for phenotypes such as their abilities to grow on defined media containing the appropriate sugar as the sole carbon source; or markers that confer the abilities of forming colored colonies on chromogenic substrates or the abilities to fluoresce.

[0079] “Transformation,” as used herein, refers to the insertion of a polynucleotide into a host cell, regardless of the method used for insertion, which may be, for example, transformation, transfection, infection, and the like. The introduced polynucleotide may be maintained as a nonintegrated vector, for example, an episome, or alternatively, may be integrated into the host genome.

[0080] A “gene” comprises a DNA region encoding a gene product, as well as all DNA sequence regions that regulate the production of the gene product, whether or not such

regulatory sequence regions are adjacent to coding sequences that may or may not be transcribed. Accordingly, a gene may be, for example, a promoter sequence, a terminator, a translational regulatory sequence such as a ribosome binding site or an internal ribosome entry site, an enhancer, a silencer, an insulator, a boundary element, a replication origin, a matrix attachment site, or a locus control region.

[0081] “Gene expression” refers to the conversion of the information, contained in a gene, into a gene product. A gene product can be the direct transcriptional product of a gene (*e.g.*, mRNA, tRNA, rRNA, antisense RNA, ribozyme, structural RNA or any other type of RNA) or a protein produced by translating an mRNA. A gene product can also be an RNA that is modified, by a process such as capping, polyadenylation, methylation, or editing; or a protein modified by, for example, methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, myristilation, and glycosylation.

[0082] A “coding sequence” or a sequence that “encodes” a selected polypeptide, is a nucleic acid molecule that is transcribed (in the case of a DNA) and translated (in the case of an mRNA) into a polypeptide *in vivo*, when the sequence is placed under the control of one or more appropriate regulatory sequences. The coding sequence begins at a start codon at the 5' (amino) terminus and ends at a translation stop codon at the 3' (carboxy) terminus. A coding sequence can be, for example, a cDNA from viral, prokaryotic, or eukaryotic mRNA; a genomic DNA viral sequence (*e.g.* DNA viruses and retroviruses); a prokaryotic DNA; or a synthetic DNA sequence. A transcription termination sequence may be located at a position that is 3' to the coding sequence.

[0083] A “fragment” refers to a polypeptide or polynucleotide comprising only a part of the sequence and structure of an intact full-length polypeptide or polynucleotide. The polypeptide fragment can comprise a C-terminal deletion, an N-terminal deletion, and/or an internal deletion from the intact polypeptide. The polynucleotide fragment can comprise a 5' deletion, a 3' deletion, and/or an internal deletion from the intact polynucleotide. A fragment of a protein generally comprises at least about 5-10 contiguous amino acid residues of the full-length molecule, at least about 15-25 contiguous amino acid residues of the full-length molecule, and at least about 20-50 or more contiguous amino acid residues of the full-length molecule. A fragment of a polynucleotide generally comprises at least about 15-30 contiguous nucleotides of the full-length molecule, at least about 45-75 continuous nucleotides of the full-length molecule, and at least about 60-150 or more contiguous nucleotides of the full-length molecule. In a certain embodiment, the number of

amino acid residues in the fragment may be any integer between 5 and the total number of amino acid residues in the full-length molecule. In another embodiment, the number of nucleotides in the polynucleotide fragment may be any integer between 15 and the total number of nucleotides in the full-length molecule.

[0084] The term “host cell” or “recombinant host cell” refers to an individual cell, cell line, cell culture, or a cell *in vivo*, which can be or has been a recipient of one or more polynucleotides or polypeptides of the invention, which may be, for example, a recombinant vector, an isolated polynucleotide, an antibody, or a fusion protein. Host cells may be progeny of a single host cell, and the progeny may not necessarily be identical in morphology, physiology, in total DNA, RNA, or in polypeptide complement to the original recipient cell, as a result of natural, accidental, or deliberate mutations and/or changes. Host cells can be prokaryotic or eukaryotic, including but are not limited to, mammalian, insect, amphibian, reptile, crustacean, avian, fish, plant and fungal cells. A host cell may be a cell that is transformed, transfected, transduced, or infected *in vivo* or *in vitro* with a polynucleotide of the invention such as a recombinant vector. A host cell that comprises a recombinant vector of the invention may be called a “recombinant host cell.”

[0085] The term “receptor” refers to a polypeptide that binds to a specific extracellular molecule and this binding may initiate a cellular response.

[0086] The term “ligand” refers to a molecule that binds to a specific site on another molecule.

[0087] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention as claimed. Moreover, it must be understood that the invention is not limited to the particular embodiments described, as the embodiments may, of course, vary. Further, the terminology used to describe particular embodiments is not intended to be limiting, since the scope of the present invention will be limited only by its claim.

[0088] Unless defined otherwise, the meanings of all technical and scientific terms used herein are those commonly understood by one of ordinary skill in the art to which this invention belongs. One of ordinary skill in the art will appreciate that any methods and materials similar or equivalent to those described herein can also be used to practice or test the invention.

[0089] It must be noted that, as used herein and in the appended claims, the singular forms “a,” “or,” and “the” include plural referents unless the context clearly dictates

otherwise. Thus, for example, reference to “a subject polypeptide” includes a plurality of such polypeptides and reference to “the agent” includes reference to one or more agents as well as equivalents thereof known to those skilled in the art.

[0090] Further, all numbers expressing quantities of ingredients, reaction conditions, % purity, polypeptide and polynucleotide lengths, and so forth, used in the specification and the claims, are modified by the term “about,” unless otherwise indicated. Accordingly, the numerical parameters set forth in the specification and claims are approximations that may vary depending upon the desired properties of the present invention. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should be construed in light of the number of reported significant digits, applying customary rounding techniques.

[0091] Nonetheless, the numerical values set forth in the specific examples are reported as precisely as possible. Any numerical value, however, inherently contains certain errors from the standard deviation of its experimental measurement.

[0092] All publications cited are incorporated by reference herein in their entireties, including references cited in such publications are also incorporated by reference in their entireties.

Leader Sequences

[0093] As described herein, secretory leader sequences, which are identified from secretable proteins, are demonstrated to be useful for producing proteins at an amount that is about 5% higher, about 10% higher, about 20% higher, about 30% higher, about 40% higher, or about 50% or more higher, than when such proteins are produced under the same conditions from DNA sequences that contain the protein's endogenous secretory leader sequence. Secretory leader sequences identified and described herein include, for example, those from the following secretable proteins: interleukin-9 precursor, T cell growth factor P40, P40 cytokine, triacylglycerol lipase, pancreatic precursor, somatoliberin precursor, vasopressin-neurophysin 2-copeptin precursor, beta-endoendorphin-dynorphin precursor, complement C2 precursor, small inducible cytokine A14 precursor, elastase 2A precursor, plasma serine protease inhibitor precursor, granulocyte-macrophage colony-stimulating factor precursor, interleukin-2 precursor, interleukin-3 precursor, alpha-fetoprotein precursor, alpha-2-HS-glycoprotein precursor, serum albumin precursor, inter-alpha-trypsin inhibitor light chain, serum amyloid P-component precursor, apolipoprotein A-II precursor, apolipoprotein D precursor, colipase precursor, carboxypeptidase A1 precursor, alpha-s1

casein precursor, beta casein precursor, cystatin SA precursor, follitropin beta chain precursor, glucagon precursor, complement factor H precursor, histidine-rich glycoprotein precursor, interleukin-5 precursor, alpha-lactalbumin precursor, Von Ebner's gland protein precursor, matrix Gla-protein precursor, alpha-1-acid glycoprotein 2 precursor, phospholipase A2 precursor, dendritic cell chemokine 1, statherin precursor, transthyretin precursor, apolipoprotein A-1 precursor, apolipoprotein C-III precursor, apolipoprotein E precursor, complement component C8 gamma chain precursor, serotransferrin precursor, beta-2-microglobulin precursor, neutrophils defensins 1 precursor, triacylglycerol lipase gastric precursor, haptoglobin precursor, neutrophils defensins 3 precursor, neuroblastoma suppressor of tumorigenicity 1 precursor, small inducible cytokine A13 precursor, CD5 antigen-like precursor, phospholipids transfer protein precursor, dickkopf related protein-4 precursor, elastase 2B precursor, alpha-1-acid glycoprotein 1 precursor, beta-2-glycoprotein 1 precursor, neutrophils gelatinase-associated lipocalin precursor, C-reactive protein precursor, interferon gamma precursor, kappa casein precursor, plasma retinol-binding protein precursor, interleukin-13 precursor, and any of the secretable proteins listed in Tables 1-3.

[0094] The above-identified secretory leader sequences, together with the vectors and methods of the invention, are useful in expressing a wide variety of polypeptides, including, for example, secretable polypeptides, extracellular proteins, transmembrane proteins, and receptors, such as a soluble receptor. Examples of such polypeptides include cytokines and growth factors, such as Interleukins 1 through 18; the interferons; the lymphokines; hormones; RANTES; lymphotoxin- β ; Fas ligand; flt-3 ligand; ligand for receptor activator of NF-kappa B (RANKL); TNF-related apoptosis-inducing ligand (TRAIL); CD40 ligand; Ox40 ligand; 4-1BB ligand and other members of the TNF family; thymic stroma-derived lymphopoietin; stimulatory factors such as, for example, granulocyte colony stimulating factor and granulocyte-macrophage colony stimulating factor; inhibitory factors; mast cell growth factor, stem cell growth factor, epidermal growth factor, growth hormone, tumor necrosis factor; leukemia inhibitory factor; oncostatin-M; splice variants; and hematopoietic factors such as erythropoietin and thrombopoietin.

[0095] Descriptions of some of the proteins that can be expressed according to the invention may be found, for example, in HUMAN CYTOKINES: HANDBOOK FOR BASIC AND CLINICAL RESEARCH, Vol. II (Aggarwal and Gutterman, eds., Blackwell Sciences, Cambridge, MA 1998); in GROWTH FACTORS: A PRACTICAL APPROACH (McKay and Leigh,

eds., Oxford University Press Inc., New York, NY 1993); and in THE CYTOKINE HANDBOOK (A.W. Thompson, ed., Academic Press, San Diego, CA 1991).

[0096] Receptors for any of the aforementioned proteins may also be expressed using secretory leader sequences, vectors and methods described herein. The receptors may include, for example, both forms of tumor necrosis factor receptor (referred to as p55 and p75), Interleukin-1 receptors (types 1 and 2), Interleukin-4 receptor, Interleukin-15 receptor, Interleukin-17 receptor, Interleukin-18 receptor, granulocyte-macrophage colony stimulating factor receptor, granulocyte colony stimulating factor receptor, receptors for oncostatin-M and leukemia inhibitory factor, receptor activator of NF-kappa B (RANK), receptors for TRAIL, and receptors that comprise death domains, such as Fas or Apoptosis-Inducing Receptor (AIR).

[0097] Other proteins can also be expressed using the secretory leader sequences, vectors and methods described herein. These proteins include, for example, cluster of differentiation antigens (referred to as "CD proteins" or "CD molecules") such as those disclosed in LEUKOCYTE TYPING VI (Proceedings of the VIth International Workshop and Conference; Kishimoto et al. eds.; Kobe, Japan 1996), or in the proceedings of subsequent workshops. Examples of CD molecules include CD27, CD30, CD39, CD40, and ligands thereto, such as the CD27 ligand, the CD30 ligand and the CD40 ligand. Several of these are members of the TNF receptor (TNFR) family, which includes 4-1BB and OX40; the ligands, including the 4-1BB ligand and the OX40 ligand, are often members of the TNF family. Accordingly, members of the TNF and TNFR families can be expressed using the secretory leader sequences, vectors and methods of the present invention.

[0098] Proteins that are enzymes may also be expressed employing the herein described secretory leader sequences, vectors and methods. These enzymes may include, for example, members of the metalloproteinase-disintegrin family; various kinases such as streptokinase, tissue plasminogen activator, Death Associated Kinase Containing Ankyrin Repeats, IKR 1, or IKR 2; TNF-alpha Converting Enzyme; and numerous other enzymes. Ligands for enzymes can also be expressed by applying the secretory leader sequences, vectors and methods of the instant invention.

[0099] The secretory leader sequences, vectors and methods described herein, are also useful for the expression of other types of recombinant proteins. These recombinant proteins may include, for example, immunoglobulin molecules or portions thereof, as well as chimeric antibodies (*e.g.*, antibodies that have human constant regions coupled to murine

antigen-binding regions) or fragments thereof. Numerous techniques are known by which DNAs encoding immunoglobulin molecules can be manipulated to yield DNAs capable of encoding recombinant proteins such as single chain antibodies, antibodies with enhanced affinity, or other antibody-based polypeptides (*see, e.g., Larrick et al. 1989; Reichmann et al. 1988; Roberts et al. 1987; Verhoeyen et al. 1988; Chaudhary et al. 1989*).

Vectors, Host Cells, and Protein Production

[00100] The present invention provides recombinant vectors that contain, for example, nucleic acid constructs that encode one or more secretory leader sequences of interest or selected heterologous polypeptides of interest that are not necessarily secretory leader sequences, and host cells that are genetically engineered to incorporate the recombinant vectors.

[00101] The vector of the invention may be one that contains a selectable marker for propagation in a host and a secretory leader sequence such as one of those listed in Table 1. Such selectable markers may be, for example, dihydrofolate reductase; G418; neomycin-, puromycin-resistance for eukaryotic cell cultures; or tetracycline-, kanamycin-, puromycin-, or ampicillin-resistance for *E. coli* and other bacterial cultures.

[00102] The vector of the invention may be, for example, a phage, plasmid, viral, or retroviral vector. Generally, a plasmid vector is introduced in a precipitate form, such as a calcium phosphate precipitate, or in a complex comprising a charged lipid. If the vector is a virus, it may be packaged *in vitro* using an appropriate packaging cell line, and then incorporated into host cells by transduction. A retroviral vectors may be replication competent or replication defective. And when it is replication defective, viral propagation generally occurs only in complementing host cells.

[00103] Among vectors useful in the present invention are the herein described vectors employing a pTT vector backbone (*see, e.g., Figures 3-7*) (*Durocher et al. 2002*). Briefly, the pTT vector backbone may be prepared with the following method: (1) obtain the pIRESpuro/EGFP (pEGFP) basic vector and the pSEAP basic vector from CLONETECH® (Palo Alto, CA); (2) obtain the pcDNA3.1, pcDNA3.1/Myc-(His)₆, and pCEP4 vectors from INVITROGEN®; (3) obtain SUPERGLO™ GFP variant ("sgGFP") from Q-BIOGENE® (Carlsbad, CA); (4) prepare a pCEP5 by the following steps: (a) remove the CMV promoter and the polyadenylation signal of pCEP4 by sequential digestion and self-ligation, using *Sal* I and *Xba* I restriction enzymes, resulting in a pCEP4Δ plasmid; (b) ligating a *Bgl* II fragment from pAdCMV5 (*Massie et al. 1998*),

which encodes the CMV5-poly(A) expression cassette, into a *Bgl* II-linearized pCEP4Δ, resulting in a pCEP5 vector; (5) generate the pTT vector by deleting the hygromycin and EBNA1 expression cassettes, the deletion of the former being accomplished by *Bsm* I and *Sal* I excision and subsequent fill-in and ligation, while the deletion of the latter being accomplished by *Cla* I and *Nsi* I excision and subsequent fill-in and ligation; (6) replacing the ColEI origin, which comprises the *Fsp* I-*Sal* I fragment that includes the 3' end of β-lactamase ORF, with a *Fsp* I-*Sal* I fragment containing the pMBI origin and the same 3' end of β-lactamase ORF from pcDNA3.1; (7) ligating in-frame into the pcDNA3.1/Myc-His digested with *Hind* III and *Eco*R V; and (8) add a Myc-(His)₆ C-terminal fusion tag to SEAP, which is a *Hind* III-*Hpa* I fragment from pSEAP-basic. Plasmids are then amplified in *E. coli* (DH5α) grown in LB medium. They are purified from the medium using MAXI-PREP™ columns from QIAGEN™ (Mississauga, ON, Canada). The quantity of the plasmids thus made is measured by diluting the plasmids in 50 mM Tris-HCl, pH 7.4, and measuring the absorbencies at 260 nm and 280 nm. For the purpose of the invention, plasmid preparations with A₂₆₀/A₂₈₀ ratios between about 1.75 and about 2.00 are used.

[00104] The nucleic acid constructs of interest may be a DNA that is operatively linked to an appropriate promoter. The appropriate promoter may be, for example, the phage lambda PL promoter, the *E. coli* lac, trp, phoA and tac promoters, the SV40 early and late promoters, or one of the promoters from retroviral LTRs. The promoters may also be, for example, the metallothionein promoters derived from the genome of mammalian cells. Alternatively, the promoters may be the adenovirus late promoters or the vaccinia virus 7.5K promoters derived from mammalian viruses. Other suitable promoters are known to the person skilled in the art.

[00105] The expression constructs further contain sites for transcription initiation, termination and, in the transcribed region, a ribosome binding site. The coding portion of the transcripts expressed by the constructs will preferably include an appropriately-positioned translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) at the end of the polypeptide to be translated. The heterologous polypeptides the polynucleotides encode may include, for example, extracellular fragments of secretable proteins, type I membrane proteins, type II membrane proteins, multi-membrane proteins, and soluble receptors.

[00106] A construct can be introduced into a host cell by calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or any other methods known to the person skilled in the art. Such methods are described in many standard laboratory manuals, such as by *Davis et al.*, in *BASIC METHODS IN MOLECULAR BIOLOGY* (1986). Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera Sf9* cells; animal cells such as CHO, COS, 293 (including 293-6E and 293-T) and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for growing these representative host cells are known in the art.

[00107] A variety of host-expression vector systems may be used to express the polypeptides of the invention. Such host-expression systems are vehicles by which the coding sequences of interest may be produced and subsequently purified. These systems can also be cells that, when transformed or transfected with the appropriate nucleotide coding sequences, express the polypeptides of the invention. These systems may include, for example, microorganisms, such as bacteria like *E. coli* or *B. subtilis*, transformed with recombinant bacteriophage DNA, plasmid DNA, or cosmid DNA expression vectors that contain the polypeptide coding sequences. These systems may also include, for example, yeast such as *Saccharomyces* or *Pichia*, transformed with recombinant yeast expression vectors that contain the polypeptide coding sequences. They may also be insect cells infected with recombinant virus expression vectors such as baculovirus, which contain the polypeptide coding sequences. They may also be plant cells infected with recombinant virus expression vectors such as cauliflower mosaic viruses ("CaMV") or tobacco mosaic viruses ("TMV"), or transformed with recombinant plasmid expression vectors such as Ti plasmids, which contain the polypeptide coding sequences. They may further include mammalian cells such as COS, CHO, BHK, 293, 293-6E, PER.C6®, 293T, or 3T3, which harbor recombinant expression constructs that contain promoters.

[00108] After the host cells are transfected with the vectors or DNA constructs encoding the polypeptides of interest, the cells are then grown on proper mediums and under proper conditions to produce the polypeptides of the present invention.

[00109] Typically, a heterologous polypeptide may be expressed as a fusion protein. It may further include not only one or more of the secretion signals, but also one or more of

the secretory leader sequences as exemplified in Table 1. The expression of such fusion proteins according to the invention is described in detail below.

[00110] Additionally, peptide moieties and/or purification tags may be added to the polypeptide to facilitate purification, improve stability, and engender secretion or excretion. The moieties and/or tags may be removed prior to the final steps of purification. The techniques are familiar and routine to one skilled in the art. In certain embodiments, such a tag may be a hexa-histidine peptide, such the one provided in a pQE™ vector (QIAGEN™, Inc., Chatsworth, CA). Another peptide tag, the “HA” tag that is an epitope derived from the influenza hemagglutinin protein may also be fused with the polypeptide of the present invention. (*See Wilson et al.* 1984). Other suitable purification tags may be, for example, V5, HISX8, avidin, or biotin.

[00111] In a certain embodiment, the fusion protein comprises a heterologous region from immunoglobulin, the presence of which may facilitate purification and may help to stabilize the purified protein. For example, EP-A-O 464 533 and its Canadian counterpart 2045869 describe fusion proteins comprising various parts of the immunoglobulin constant region (Fc) and a human protein or parts thereof. According to EP-A 0232 262, the Fc regions in a fusion protein is thought to be advantageous for use in therapy and diagnosis because they tend to lead to improved pharmacokinetic properties. But for some other uses, it might be desirable to delete the Fc regions after the fusion protein has been expressed, detected and purified, especially when the Fc regions hinder the use of the polypeptide to which the regions are fused in therapy and diagnosis. For example, the deletion of Fc regions might be necessary when the fusion protein is used as an antigen for immunization.

[00112] The purification tags may also be used in drug discovery. For example, a human protein hIL-5 was fused with the Fc regions to facilitate the identification of hIL-5 antagonists using high-throughput screening assays. (*Bennett et al.* 1995; *Johanson et al.* 1995).

[00113] A heterologous polypeptide of the invention can be purified from a recombinant cell culture by well-known methods, which include, for example, ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. In a particular embodiment, high performance liquid chromatography (“HPLC”) is employed for purification. Polypeptides of the present invention may include, for example, products

purified from directly-isolated or cultured natural sources such as bodily fluids, tissues and cells; products of chemical synthetic procedures; products produced by recombinant techniques from prokaryotic or eukaryotic hosts such as bacterial cells, yeast, higher plant cells, insect cells, mammalian cells; or products produced by recombinant techniques from cell-free expression systems.

Modifications

[00114] The invention encompasses polypeptides that are differentially modified during or after translation, for example, by glycosylation, acetylation, methylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, or linkage to an antibody molecule or other cellular ligand. Any of these chemical modifications may be carried out by known techniques, including specific chemical cleavage by cyanogen bromide; digestion by trypsin, chymotrypsin, papain, or V8 protease; treatment by NABH_4 ; acetylation; formylation; oxidation; reduction; and metabolic synthesis in the presence of tunicamycin.

[00115] Depending upon the hosts employed in the recombinant production procedures, the polypeptides of the present invention may be glycosylated or non-glycosylated. A polypeptide of the invention may also include an initial modified methionine residue at the N-terminus, usually as the result of host-mediated processes. It is known in the art that the N-terminal methionine encoded by the translation initiation codon can generally be removed with high efficiency after translation in eukaryotic cells. While the N-terminal methionines can be efficiently removed from most prokaryotic proteins, the removal processes are not always efficient in prokaryotes. The efficiency depends on the nature and identity of the amino acids to which the N-terminal methionines are covalently linked.

[00116] Additional post-translational modifications according to by the invention include, for example, N-linked or O-linked carbohydrate chains, processing of N-terminal or C-terminal ends, attaching the chemical moieties to the amino acid backbone, chemical modifications of N-linked or O-linked carbohydrate chains, or addition or deletion of an N-terminal methionine as the result of prokaryotic host cell expression. To facilitate detection and isolation of the protein, the polypeptide may also be modified with one or more detectable labels, which may be, for example, an enzymatic, fluorescent, isotopic, or affinity label.

[00117] Additional embodiments of the invention may be chemically modified derivatives of the polypeptides of the invention, which may provide additional advantages such as increased solubility, stability and circulating time for the polypeptides, or decreased immunogenicity in biological systems (U.S. Pat. No. 4,179,337). The chemical moieties used in derivitization may be selected from water soluble polymers such as polyethylene glycol, ethylene glycol/propylene glycol copolymers, carboxymethylcellulose, dextran, polyvinyl alcohol and the like. The polypeptides may be modified at random positions, or at predetermined positions within the molecule, and may include one, two, three, or more attached chemical moieties. Reaction conditions may be selected from any of those known in the art and those subsequently developed, but should be selected so that the protein to be modified is not exposed or will suffer only limited loss of activity due to harsh temperature, solvent, and pH conditions. In general, the larger the ratio of polymer to polypeptide conjugate, the greater the percentage of conjugated product. The optimum ratio, measured by the efficiency of the reaction, may be determined by factors such as the desired degree of derivatization (*e.g.*, mono-, di-, tri- etc.), the molecular weight of the polymer selected, the degree of branching, and the reaction conditions. The ratio of polymer to polypeptide generally ranges from 1:1 to 100:1. One or more purified conjugates may be prepared from each mixture by standard purification techniques, which includes, for example, dialysis, salting-out, ultrafiltration, ion-exchange chromatography, gel filtration chromatography, and electrophoresis.

[00118] A polymer may be of any molecular weight, and may be branched or unbranched. In certain embodiments, where the polypeptides of the invention are modified by polyethylene glycol, the molecular weight of the polyethylene glycol is from about 1 kDa to about 100 kDa. The term "about," when used in the description of polyethylene glycol, is intended to suggest that, during the preparations of polyethylene glycol, some molecules will weigh more, some less, than the stated molecular weight. The size of the polyethylene glycol used in the modification may depend on the desired therapeutic profile, such as, for example, the desired duration of sustained release; the effects, if any, on biological activities; the ease of handling; the degree or the lack of antigenicity; as well as other known effects of the polyethylene glycol on a therapeutic protein or an analog.

[00119] There are a number of attachment methods available to those skilled in the art. For example, EP 0 401 384 describes the coupling of PEG to G-CSF. *Malik et al.* reported pegylation of GM-CSF using tresyl chloride (*Malik et al.* 1992). Polyethylene

glycol may be covalently bound to a reactive group of an amino acid residue, which may be, for example, a free amino group, a carboxyl group, or a sulfhydryl group. In the context of pegylation, which means attaching polyethylene glycol moieties to a molecule, reactive groups are defined as groups to which an activated polyethylene glycol molecule may be bound.

[00120] One may specifically desire proteins chemically modified at the N-terminus. This may be accomplished by reductive alkylation, which exploits the different reactivities of primary amino groups such as the internal lysines and the N-terminal amino acid, which are available for derivatization. For example, one may selectively attach a polymer to the N-terminus of a protein by performing the reaction at a pH where only the α -amino group of an N-terminal residue, and not the ϵ -amino residue, would be susceptible to the reaction, taking advantage of the pKa differences between these types of amino groups. The polymer used in reductive alkylation typically have a single reactive aldehyde. The N-terminally chemically modified protein may be separated from other monoderivatized moieties, if necessary, by purifying the N-terminally modified protein from a population of protein molecules that are modified elsewhere.

Fusion Molecules of the Invention

[00121] In a further embodiment of the invention, the heterologous polypeptides of may be combined with one or more fusion partners to form fusion molecules. Such fusion molecules may advantageously provide improved pharmacokinetic properties when compared to their unmodified non-fused counterparts. These fusion molecules comprising the heterologous polypeptides of the invention may be prepared by a person skilled in the art who is apprised with the disclosures herein. Suitable chemical moieties for derivatization of a heterologous polypeptide in this regard may be, for example, polymers such as water soluble polymers; all or part of human serum albumin; fetuin A; fetuin B; or Fc regions.

[00122] Specifically, a modified heterologous polypeptide of the invention may be prepared by attaching one or more polyaminoacids, peptide moieties, or branch-point amino acids to the polypeptide. Polyaminoacids are commercially available and widely used in drug delivery technology and other emerging technologies such as gene therapy. In addition to the advantages one may achieve with a fusion molecule as described above, the polyaminoacid may be a carrier that serves to increase the polypeptide's circulation half-life. For the therapeutic purpose of the present invention, such a polyaminoacid should

ideally be one that does not generate neutralizing antigenic or other adverse responses. As described herein, the position at which the polyaminoacid is attached to the polypeptide or fusion polypeptide may be located at the N-terminus, C-terminus, or any other positions in between. The polyaminoacid may also be connected by a chemical "linker" moiety to either end of the selected polypeptide or fusion polypeptide.

[00123] A method for preparing a fusion molecule conjugated with one or more polymers, such as water-soluble polymers is described above.

[00124] Additionally, heterologous polypeptides of the present invention and the epitope-bearing fragments thereof can be combined with parts of the immunoglobulin constant domain, resulting in chimeric polypeptides. These particular fusion molecules facilitate purification and tend to show an increased half-life *in vivo* when compared to their pre-fusion counterparts. Examples of these chimeric polypeptides include, for example, the chimeric proteins comprising the first two domains of the human CD4-polypeptide and various domains of the mammalian immunoglobulin constant regions (EP A 394,827; Traunecker et al. 1988). A fusion molecule having a disulfide-linked dimeric structure tends to be more efficient in binding and neutralizing other molecules than, for example, a monomeric polypeptide or fragment (Fountoulakis et al. 1995).

[00125] In another embodiment, a human serum albumin fusion molecule may also be prepared as described herein and as further described in U.S. Patent No. 6,686,179, which is hereby incorporated by reference in its entirety.

[00126] Moreover, the polypeptides of the present invention can also be fused to a purification tag, which is a peptide region that would facilitate the purification of the polypeptides to which they are a part. The method of fusing the tag to the polypeptide of interest is described herein.

[00127] It will be clear to those skilled in the art that the invention may be practiced in ways other than those particularly described in the foregoing descriptions and the examples herein. Many modifications and variations of the present invention are possible in light of the teachings herein and, therefore, are within the scope of the appended claims.

Examples

Example 1: Expression of Biologically Active Mature Secreted Proteins using a Cell-free System.

[00128] Recombinant technologies allow for expression of proteins *in vitro* or *in vivo*. Examples of *in vitro* systems for protein expression include cell-free systems such as

rabbit reticulocyte lysates and wheat germ extracts, and cell-based systems such as bacteria, insect cells, yeast cells and mammalian cells (for example, CHO cells, 293 cells, and human embryonic retinal cells PER.C6[®] cells (Crucell, Netherlands)). *In vivo* expression of recombinant proteins is useful, for example, in the generation of transgenic animals in which the transgene(s) encodes protein(s) tagged with markers such as, for example, Green Fluorescent Proteins and its variants or β -galactosidase. Such tags allow for easier visualization, tracking and/or isolation of the cells in which the tagged protein is expressed. Another example of *in vivo* expression of recombinant proteins is the use of transgenic mice, or of cells implanted into mice, that have been genetically modified for the expression of secretable proteins. The latter can be proteins that, for example, are thought to promote tumor development, work as hormones, as growth factors, and/or as survival factors. In that setting, it can be important to obtain various levels of protein secretion (low, moderate, high) in order to obtain a specific result (*e.g.* tumor promotion). Many proteins are not efficiently secreted when expressed in recombinant settings. In that case, it is useful to be able to replace, via recombinant methods, its endogenous leader sequence with a leader sequence that is capable of driving its efficient secretion.

[00129] It is often useful to confirm that a given isolated cDNA is capable of supporting the expression of the protein which its nucleotide sequence encodes *in vitro*, before the cDNA is used to express that protein *in vivo*. This process may also serve, for example, to obtain further information regarding the post-translational modifications that the protein undergoes in a specific host cell (*e.g.* CHO cells versus PER.C6[®] cells), and the activity of the protein. In the case of a secretable protein, the cDNA sequence may either encode its full-length form, its mature form (*i.e.*, the protein without the leader sequence), or any other parts of the protein, such as a particular domain.

Preparation of Plasmid Templates for Recombinant Protein Expression in Cell-free Systems.

[00130] To recombinantly express a cDNA encoding the mature form of any protein of interest, it is often useful that the cDNA be modified in order to include, in addition to the coding sequence, a translational initiation site/translational enhancer (*e.g.* KOZAK sequence, Omega sequence, Non-Omega sequences). In this example, the mature form refers to the most typical product of secretion, which is the protein without the signal peptide. Furthermore, if no antibody exists for the protein of interest, a tag may also be added which facilitates both the detection and the purification processes. Examples of such

tags are Glutathione-S-Transferase (GST), and the epitopes V5, HisX6, and HisX8(H8). The addition of these features to a cDNA encoding a protein of interest can be done by a variety of cloning methods. If no appropriate restriction enzyme sites are present in the cDNA of interest. PCR amplification methods such as those described below can also be used during the cloning process. A cloning process that involves three PCR steps and results in a mature ORF tagged with Glutathione-S-Transferase is exemplified below.

[00131] To begin, a first plasmid containing the cDNA sequence encoding the mature open reading frame (mature ORF) of interest was provided for the first PCR. To add the translational initiation site/translational enhancer to the 5' region of the coding sequence for the mature ORF, a nucleotide primer (forward primer FP1) was designed and synthesized, which contained 5'GTTCTGTTCCAGGGGCCC 3' followed by the first nineteen nucleotides predicted to encode the amino terminus of the mature secretable protein of interest. A second primer (reverse primer RP1) was designed and synthesized, based on a region of the plasmid approximately 1000 nucleotides downstream from the coding sequence (mature ORF) of the cDNA to be expressed. In fact, the RP1 primer was designed as the reverse complement of the vector sequence in this region such that RP1 could be used with FP1 in a PCR to amplify the mature ORF. The exact sequence of RP1 would vary depending on the starting plasmid, but it was typically 17-23 nucleotides long with a T_m of approximately 55-65°C.

[00132] The purified starting plasmid containing the cDNA to be expressed as a mature ORF, or *E coli* cells containing the purified plasmid, was then added as a template to a standard PCR, which included the two primers (FP1 and RP1), as described above, standard PCR reagents, and a DNA polymerase. The reaction mixture was then subjected to 15-30 cycles of PCR amplification. The product of this first PCR is called the "PCR1 coding templates" for the purpose of this application.

[00133] A separate PCR was performed to prepare a "GST-Mega primer," whose purpose was to provide the GST portion of the final GST-mature ORF expression template in the second PCR step. To this end, a different starting plasmid template was used, for example, one containing a GST coding sequence downstream from the Non-Omega translation initiation sequence, and which is herein referred to as "template 2." It is often useful that the GST fusion protein is linked to the mature ORF via a cleavable bridge. To this end, the template might have a GST protein modified to include a protease-cleavable sequence, such as one sensitive to thrombin, or to the commercially available PreScission™

Protease (Amersham, NJ). This allows for the two proteins, mature ORF and GST, to be separated at the end of the purification procedure by protease-mediated cleavage. Thus, a PCR was prepared to amplify "template 2" using two primers: FP2, of sequence 5' GGTGACACTATAGAACTCACCTATCTCCCAACA 3'; and RP2, of sequence 5' GGGCCCCTGGAACAGAACTTC 3'. The amplification took place for 15 to 30 cycles in a standard PCR mixture that included template 2, the two primers described above (FP2 and RP2), standard PCR reagents, and a DNA polymerase. After the PCR was complete, the amplification product was treated with exonuclease I for 30 minutes at 37°C, and then heat-inactivated at 80°C for 30 minutes. The product was then purified by agarose gel electrophoresis and extracted using a gel purification kit (Amersham, NJ), producing the "GST-Mega primer." The "GST-Mega primer" was, in fact, one of the two templates used in the second PCR that yields a GST-fusion expression template. The other template of the final reaction was the "PCR1 coding template," prepared as described above.

[00134] The final construct, which was the mature ORF/GST fusion expression template, was prepared as follows. The two templates "GST-Mega Primer" and "PCR1 coding template" were combined via the second PCR involving the mature ORF. This PCR reaction mix included: (i) standard PCR reagents; (ii) a DNA polymerase; (iii) an aliquot of the "PCR1 coding template" (e.g., 0.5 µl); (iv) an aliquot of the "GST-Mega primer" (e.g., 1 µl); (v) a fifth primer, FP3, of sequence 5' GCGTAGCATTTAGGTGACACT 3', which comprised part of the SP6 promoter sequence, and was annealed to the 5' end of the "GST Mega primer" via its common 3' end (compare underlined sequences); and (vi) a sixth primer, RP3, which was designed as the reverse complement of the vector sequence in the same region of the vector as RP1 but starting three nucleotides upstream of RP1 to specifically anneal only on the full-length PCR1 coding template; RP3 is typically 17-23 nucleotides long with a T_m of approximately 55-65°C, and can be used in amplifying the "PCR1 coding template." After 15-30 cycles of PCR amplification, the "Mature ORF/GST-fusion expression template" was thus generated.

Expression of GST-fusion Expression Templates in Wheat Germ Extracts

[00135] In order to express a mature protein of interest in a cell-free system, the mRNA can be both transcribed and translated from the "Mature ORF/GST-fusion expression template" in the same reaction, or in separate reactions. A separate *in vitro* transcription reaction (50 µl) can be prepared with 5 µl of the "GST-fusion expression template" in the following buffer: 80 mM HEPES KOH pH 7.8, 16 mM Mg(OAc)₂, 2 mM

spermidine, 10 mM DTT, 1 unit of SP6 (Promega, WI) and 1 unit of RNasin (Promega, WI). The reaction mixture is incubated for 3 hours at 37°C. The resulting mRNA is subjected to ethanol precipitation in a solution containing 200 µl of RNase-free water, 37.5 µl of 5 M ammonium acetate, and 862 µl of 99% ethanol. The ethanol precipitation comprises the steps of mixing by vortexing and pelleting by centrifugation at 15,000 x g for 10 minutes at 4°C. The mRNA pellet is then washed in 70% ethanol and again pelleted by centrifugation at 15,000 x g for 5 minutes at 4°C, after which steps the pelleted mRNA is ready for *in vitro* translation.

[00136] Wheat germ extracts can be used for *in vitro* translation of the mRNA, prepared separately as described above. First, a stock solution of 2x Dialysis Buffer was prepared from mixing two separate stocks of amino acids. The first stock contained 20 mM HEPES KOH buffer pH 7.8, 200 mM KOAc, 5.4 mM Mg(OAc)₂, 0.8 mM Spermidine, 100 µM DTT, 2.4 mM ATP, 0.5 mM GTP, 32 mM creatine phosphate, 0.02 % NaN₃, and 0.6 mM of an amino acid mix that did not contain aspartic acid, tryptophan, glutamic acid, isoleucine, leucine, phenylalanine and tyrosine. The second stock contained a 80 mM mix of the amino acids aspartic acid, tryptophan, glutamic acid, isoleucine, phenylalanine and tyrosine in 1 N HCl. After all the amino acids in the second stock were dissolved, the two stocks were mixed, so that the final concentration of the second-stock of amino acids was 0.6 mM. The 2x Dialysis Buffer stock was then adjusted to pH 7.6 using 5 N KOH, filter sterilized, and stored frozen in aliquots at -80°C.

[00137] To resuspend the *in vitro* transcribed mRNA (prepared separately as described above), a 50 µl “translation mixture” was prepared that includes Wheat Germ Reagent (Promega, WI) at a final OD_{260 nm} of 60 prepared in 1 x Dialysis buffer containing 2 mM dithiothreitol (DTT). After removing the supernatant (ethanol) from the precipitated mRNA, the 50 µl “translation mixture” was added to the precipitate and allowed to sit for 5-10 minutes before the mRNA was resuspended into the translation mixture. The complete translation mixture containing the resuspended mRNA was then layered under 250 µl of 1x Dialysis Buffer that had already been added to one well of 96 well round bottom microtiter plate to setup a Bilayer Reaction. The plate was then sealed manually with a plate seal and the *in vitro* translation reaction allowed to incubate for 20 hours at 26°C.

[00138] At the end of the *in vitro* translation reaction period, and to recover the recombinant mature ORF protein expressed as a GST fusion, the translation mixture was transferred to a tube and diluted five-fold with phosphate buffer-saline containing 0.25 M

sucrose and 2 mM DTT. Ten microliters of glutathione(GSH)-sepharose beads (Amersham-Pharmacia Biotech, NJ), to which the Glutathione-S-Transferase (GST) protein binds, were then added to the mixture, which was then incubated at 4°C for 3 hours, with constant agitation to ensure mixing. The GSH-sepharose beads, containing the bound GST-fusion protein, were then washed three times in PBS containing 0.25 M sucrose and 2 mM DTT. If the mature ORF and the GST were recombinantly engineered to be fused via a protease cleavable bridge, a fourth wash was then performed in a protease-cleavage buffer containing 50 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 2 mM DTT, and 0.25 M sucrose. This wash buffer was also called the “final wash buffer.” After the wash buffer was carefully removed from the bead mixture, 10 µl of the final wash buffer collected from the last step was mixed with the beads, and 0.4 µl of the appropriate protease such as PreScission™ Protease (Amersham, NJ) was added to the mixture. A pipette was then used to gently suspend the beads. This bead mixture/suspension was then allowed to sit overnight at 4°C. To recover the cleaved mature ORF protein product, 20 µl of the final wash buffer was added and entire liquid fraction (without the beads) recovered by pipetting (after allowing the beads to settle), or by filtering through a sintered frit.

[00139] Aliquots of the recovered liquid fraction (containing the purified mature protein) were analysed by ELISA and/or Coomassie/Silver Staining of SDS-PAGE gels, in order to quantify the level of expression of the mature protein.

[00140] To stabilize the recovered mature ORF protein, a solution of 10 mg/ml purified BSA in PBS was added to the purified protein solution so that the final concentration of BSA became about 1 mg/ml. The protein sample was then dialyzed in PBS and filter-sterilized for storage. Western blot analysis can be done from aliquots recovered throughout various steps along the purification procedure to assess, for example, the level of protein expression, and to determine whether or not the protein translated corresponds to the protein expected to be encoded by the cDNA of interest, both in terms of its length and its sequence. The protein can also be used in future characterization studies, such as biological activity measurements, mass-spectrometry, and post-translational modification assays. To produce additional protein from the same mRNA template, the single Bilayer Reaction can be repeated multiple times, and the purification and formulation can be scaled accordingly.

[00141] Typically, sixteen Bilayer reactions (set up as described above) will produce sufficient biologically active protein for testing in most typical assays such as biological

activity assays. Since these reactions are done in 96 well plates, this expression system is suitable for high-throughput assays in which multiple cDNAs of interest can be translated simultaneously in separate wells. Once a cDNA is shown capable to encode a specific protein of interest in wheat germ extracts, it can be desirable to express larger amounts of protein than those typically obtainable with this expression system. It can also be desirable to compare the post-translational modifications that a given protein undergoes in different cell systems, for example those that occur in a plant-based system such as the wheat germ lysates, with those that occur in a mammalian system (*e.g.* CHO cells, 293 cells, PER.C6[®] cells).

Evaluation of the Expression Levels of Various Signal-Peptide-Less Mature Proteins

[00142] Column 3 (“Highest Expressors”) of Table 3 summarizes the results of a high-throughput expression experiment aimed at comparing the expression levels of various proteins of the invention, without their endogenous signal peptide and under standardized conditions. Starting with a set of cDNAs that included those encoding the full-length proteins listed in Table 1 and Appendix A, mature ORF templates were prepared as described in detail in the previous paragraphs, to express the mature version of each protein (*i.e.*, the protein without its endogenous signal peptide). After purification, the expression levels were quantified by Coomassie-stained SDS-PAGE, silver stained SDS-PAGE, or quantitative Western Blot using an Anti-V5 antibody relative to purified V5-tagged protein standards, and 56 of the “highest expressors” were ranked from 1 (high) to 56 (low) based on their expression levels, relatively to each other. Under these standardized conditions, among the “highest expressors” of column 3/Table 3, the very highest expressor (ranked 1) was the mature version of the beta-subunit of prolyl 4-hydroxylase, (CLN00517790); a moderate expressor (ranked 20) was the mature version of the long form of alpha I collagen type IX (CLN 00517648); and the lowest expressor (ranked 56) was WFIKKN-related protein (CLN 00463474).

Example 2. Identification of Leader Sequence-Containing Proteins that are Secreted at High Levels from Mammalian Cells.

[00143] The next set of assays aimed at comparing proteins on the basis of the amounts that could be recovered from the conditioned media (*i.e.*, on the basis of “secretion”). The cDNAs used for Example 1, table 1 and table 2, were subcloned into modified versions of the pTT mammalian expression vector, and the proteins were

expressed with their endogenous signal peptides/leader sequences, in mammalian cells. After quantifying the levels of the resulting protein present in the conditioned media, proteins were ranked again, this time from “high secretors” to “moderate secretors” to “low secretors.”

[00144] Later on, this information served as the baseline to assess whether one could improve secretion of a protein by re-engineering its signal peptide/leader sequence. This “re-engineering” was done by replacing the endogenous signal peptide of a “low secretor” protein, with that of a “moderate” or “high secretor.”

[00145] In order to proceed with the above re-engineering, the amino acid sequence corresponding to the signal peptide/leader sequence of each of the proteins of the invention had first to be identified (Appendix A, Table 1 and 2). Based on a defined set of attributes, cDNAs from an existing library can be predicted to encode secretable proteins bioinformatically. For example, a signal peptide is typically encoded by the first 6-27 amino acid codons (18-81 nucleotides) of the ORF, and it usually begins with 1-4 polar amino acids, followed by a stretch of hydrophobic amino acids, and then followed by a short region of charged amino acids just before the site where the secretion-related cleavage takes place. Using these attributes, together with other physical characteristics, cDNAs can be predicted to encode secretable proteins while the identities of the proteins may remain unknown. The results of one of such analysis done on our complete cDNA library are summarized on Appendix A, and Tables 1 and 2. A current limitation still is that one can not predict whether or not the presence of a putative signal peptide/leader sequence allows a protein containing said leader sequence to be secreted *in vitro* or *in vivo*, and what the efficiency of this process will be.

Preparation of the Expression Vectors for High-Throughput Screening of Leader Sequences

[00146] In order to identify signal peptides or secretory leader sequences that yield high secretion levels in proteins, a set of cDNAs predicted to encode secretable proteins (using a cDNA library existing in house and the methods described above) were subcloned into one of several modified version of the pTT5 expression vector (FIG. 3) using subcloning techniques similar to those described in detail in Example 1. Some of the modified vectors contained cleavable tags (Vectors A and C, FIGS. 4 and 5) in frame with a C-terminal V5 and HisX8 epitope tag (Vector A, B and C, FIGS 4 and 5), or in frame with an Fc domain sequence (Vector D and E, FIG. 6). The presence of a HisX8 tag (which

consists of a group of eight His residues), allows for purification of the recombinantly-expressed proteins using standard Nickel column-based technologies familiar to those skilled in the art, and commercially available (e.g. Qiagen Inc., CA). When long-term selection for stable transfectants was necessary, the proteins were also expressed in vectors such as the pTT2p vector shown in FIG. 7.

[00147] The plasmid DNAs for each cDNA clone inserted into pTT5 were purified using the QIAGEN™ TURBO™ DNA system in 96-well plates. The DNA concentration for each clone was determined by absorbance at 260 nm, and subsequently adjusted, for example, to a concentration of 50 µg/ml in a suitable buffer. The expression/secretion assays were done after the resulting pTT5-based constructs were transiently transfected into 293T cells (ATCC®, VA) using a high-throughput 96-well system. These steps are described next.

High-throughput Transfection in 96-well Plates

[00148] For transient transfection of ten 96-well plates, 10 µl of each cDNA plasmid were combined with 50 µl of GIBCO® OPTI-MEM I™ (GIBCO, Gaithersburg, MD, Cat#:319-85-070) in separate wells (one for each cDNA) of a round-bottom 96-well polystyrene plate. This plate was named the “master transfection plate.” Then, 37.5 µl of each OPTI-MEM I™/cDNA mix were pre-incubated for 5 minutes with 2.5 µl of FUGENE™ 6 transfection reagent (Roche Applied Science, Palo Alto, CA, cat#1988387) in separate wells (one for each cDNA) of another round-bottom 96-well polystyrene plate. The mixture was then incubated at room temperature for about 30 minutes, resulting in one “transfection complex” per cDNA.

[00149] Each transfection complex was subsequently diluted by the addition of 100 µl of OPTI-MEM I™, mixed several times by repeated pipetting, and then transferred 20 µl at a time into ten separate wells. Each well was on a separate 96 well flat bottom polylysine-coated plate (Becton Dickinson, Rockville, MD, cat# 356461) to facilitate collection of samples for up to 10 different assays after transfection. Each plate contained up to 96 different cDNAs.

[00150] Two hundred microliters of a suspension of 2×10^5 cell/ml of 293T cells in DMEM medium (containing 10% FBS, penicillin and streptomycin) were then added to each well. The different mixtures of cells and diluted transfection complex were allowed to incubate at 37°C in 5 % CO₂. After approximately 40 hours, the medium was removed

from the wells by aspiration, the cells were briefly washed with 150 μ l PBS, and new pre-warmed medium was added.

[00151] To prepare the set of transfected cells used for the purpose of assaying the expression and secretion levels of each protein, 150 μ l of fresh HYQ-PF™ CHO Liquid Soy medium (Hyclone, Logan, UT, Cat# SH30359.02) were added to each well,

[00152] To prepare the set of transfected cells used for the purpose of assaying the activity of the secreted protein, 150 μ l fresh DMEM medium containing 5% FBS, penicillin and streptomycin were added to the wells instead of the HYQ-PF™ CHO Liquid Soy, and the resulting mixtures were incubated at 37°C in 5 % CO₂.

[00153] After an additional 48 hours, during which the various cDNA expressed their respective secretable proteins, the culture supernatants from all ten 96-well plates were harvested and, when appropriate, combined into a single sterile deep-well plate, covered with a sterile lid. The deep-well plates were centrifuged at 1,400 RPM for 10 minutes to pellet any loose cells or cell debris. The supernatants were then transferred to new sterile deep-well plates so that the level of protein released into the conditioned media (i.e. secreted protein) could be measured. This was achieved by Western blot using anti-V5-HRP antibody and sandwich ELISA using the anti-penta-HIS antibody as a capture step and anti-V5-HRP to detect expression and measure expression levels relative to purified V5His standard. The layer of cells, which remained attached to the plates, was solubilized with 0.2% SDS, 0.5% NP-40 in PBS; the resulting cell lysates were used to assay the levels of protein in the cell lysates by ELISA.

[00154] In the first set of screening assays, a subset of leader sequences were identified that were shown to correlate with high secretion levels of the proteins they belonged to. The results of a high-throughput secretion assay, done following the steps just described are shown in Figures 8 and 9. Using high-throughput expression of cDNAs in the 293T cells , several cDNAs were identified that lead to high secretion levels. A total of 56 cDNA (previously ranked as the highest expressors among our complete library of cDNAs for secretable proteins) were screened in this assay. Their identity, respective position in each lane, and relative expression level in the conditioned media, are all summarized in Table 3, columns 2 and 4. Column 2 quantified the concentration of each protein that was secreted into the conditioned media, relatively to the concentration of one or more standards that were separated in adjacent lanes (BSA for FIGS. 8 and 9). Typically, secretion levels do correlate to expression levels (compare column 2 to column 2), but not always. For

example, the “highest expressor” (ranked 1 on column 3) is also the “highest secretor” (secreted protein concentration of 32 µg/mL) according to Table 3. These results correspond to the full length protein beta subunitprolyl 4-hydroxylase (column 11). On the other hand, several proteins were secreted at the level of about 4 µg/mL (column 2), but had ranked between 16 and 21 on expression (column 3). The long form of alpha I collagen type IX (encoded by CLN00517648) is among the latter.

Example 3. A Set of Leader Sequences from High-Secretors is Useful for Converting Low-Secretors into High-Secretors .

[00155] The high-throughput assay described in detail in Example 2 provided a panel of cDNAs from the “highest expressor” proteins with levels of secretion which varied from “low secretor proteins” to “high secretor proteins.” For a summary of their identity and properties, see Tables 1, 2 and 3. The next question was whether the signal peptide/leader sequences of the high-secretors were transferable into other proteins. More importantly, we asked whether the secretion of “low secretor proteins” could be improved by replacing their endogenous leader sequence with one taken from any one of “high secretor proteins” of the invention. To this end, a series of experiments were conducted, using standard subcloning techniques, transfection and expression methods essentially as described in detail in Examples 1 and 2. One of these experiments is exemplified next.

[00156] The signal peptide/leader sequence from CLN00517648 was used to replace the signal peptide/leader sequence of a panel of proteins, which in the initial sets of high-throughput expression and high-throughput secretion assays had been shown to be low expressing proteins, low secretor proteins, or both. The proteins encoded by the resulting re-engineered cDNAs, which carried the heterologous leader sequence of the high secretor clone CLN00517648 instead of their own endogenous leader sequence, were found to have become high secretor proteins from what otherwise had been low expressor/low secretor proteins. Indeed, the signal peptide/leader sequence of CLN00517648 is capable of enhancing the secretion of type I TM proteins and type II proteins. Some specific examples of proteins whose secretion was improved by this process include cDNA constructs encoding the following proteins: human CD30 Ligand, SCDFR1, Ox40 Ligand, all of which were engineered to replace their endogenous signal peptide/leader sequence with that of CLN00517648 according to the process described in Examples 1 and 2. Moreover, the total level of expression of the modified proteins was also increased by this substitution. This was determined both by quantified the total levels of protein in both cell lysates and

conditioned media. Thus, the signal peptide/leader sequence from CLN00517648 can enhance both the expression and the secretion of low expressor proteins.

[00157] The high-throughput results described above, showing improvements in secretion and/or expression levels of low secretors and/or low expressors by replacing their endogenous leader sequence with that of either CLN00517648 or of another protein (heterologous leader sequence) selected from the list of "highest expressors" (see Table 3, column 3), were further confirmed using the scale-up procedures described in Example 4.

Example 4. Scale-Up Process for Expression of Leader Sequence-Containing Proteins in 293-6E Cells

[00158] An alternative to the 96-well high-throughput transfection-expression assay is one in which both the transfection and the expression are done in larger scale protocols. These can use, for example, 293-6E cells provided by Y. Durocher grown in shaker flasks rather than 96-well plates. For the high-throughput process, the 293-6E cells can be treated with the same reagents and subject to the same conditions as the ones used for the 293T cells except that PEI is used for DNA transfection in shake flasks instead of Fugene 6.

[00159] For the scale-up process, the 293-6E cells were grown in polycarbonate Erlenmeyer flasks fitted with a vented screw cap and rotated on a table top shaker at 100 RPM in FREESTYLE™ Medium (INVITROGEN®, Carlsbad, CA) at 37°C in 5% CO₂. The cell densities in those flasks were maintained in a range from 0.5 to 3 x 10⁶ cells/ml. Typically 50 ml cultures were grown in 250 ml flasks. One day before transfection, 293-6E cells were diluted into fresh FREESTYLE™ Medium to a cell density of about 0.6 x 10⁶ cells/ml. On the day of transfection, the cells were predicted to be in the log phase, which is characterized by a cell density range of 0.8 to 1.5 x 10⁶ cells/ml. The volumes of the log-phase cell cultures were adjusted so that their cell densities were about 10⁶ cells/ml.

[00160] For each cDNA, a different transfection mix was prepared. To prepare each transfection mix, 2.5 ml sterile PBS were added to two 15-ml tubes. The first tube also contained 50 µg DNA. The second tube also contained 100 µl PEI solution, which includes 1 mg/ml sterile stock solution of linear 25 kDa Polyethylenimine, pH 7.0 (from Polysciences, Warrington, WI). The solutions in the two tubes were then combined and allowed to incubate together for 15 minutes at room temperature, yielding the transfection complex. The transfection complex was then transferred to a 293-6E suspension culture and allowed to grow for 4 -6 days at 37°C in 5% CO₂; this process was repeated for each cDNA.

[00161] To determine protein secretion levels, culture supernatants were analyzed by Western blot. Samples (15 μ l per cDNA) were resolved by SDS-PAGE on 26-lane CRITERION™ gels (Bio-Rad, Inc., Hercules, CA) and transferred to nitrocellulose membranes. The membranes were blocked, and probed with an antibody against the specific epitope introduced at the cloning step. For example, for proteins tagged with a V5 and/or a HisX8 epitope, either an anti-V5 or an anti-HisX8 epitope antibody, conjugated to HRP (INVITROGEN®, Carlsbad, CA), was used. The HRP signal was developed using standard HRP chemiluminescence substrates (ECL Detection Kit, Amersham).

[00162] Secretion levels were determined by comparing the intensity of signal obtained for each secreted protein to that of one of three purified mass standards (for example, 15 μ l of standards at 8, 33, and 133 ng/ml) that were loaded into separate lanes of the same gels. The comparison involved determining the area of the bands present on either the Coomassie-stained gel, the silver-stained gel, or the Western blot; this process was done with a image scanner and NIH *Image* freeware, which can be downloaded from Scion Corporation website. Various protein standards were used. Examples include a V5-HisX6-tagged Delta-like protein 1 extracellular protein, a V5-HisX6-tagged CSF-1 Receptor extracellular domain, and/or a POSITOPE™ (INVITROGEN®, Carlsbad, CA, cat#: R900-50) containing a V5-HisX6 tag. These standards can be expressed separately using, for example, a baculovirus expression system, and purified to >90% purity.

[00163] Figure 2 exemplifies the results of a large-scale expression experiment in which the cDNAs (including the V5H8 epitopes) of twenty clones were subcloned into a pTT5 vector (FIG. 5). The resulting clones were transfected into 293T cells, using the methods herein described. The levels of secreted protein in 15- μ l samples of conditioned media were assessed by a Western Blot. Two V5 His standards were mixed in each standards and loaded into the right-hand lanes according to the following concentrations: (1) the higher molecular weight, V5-Hisx5 tagged Delta-like protein 1 extracellular protein, loaded at 16, 66 and 266 ng/ml; and (2) the lower molecular weight, V5-Hisx6 tagged CSF-1 extracellular domain, loaded at 8,33, and 133 ng/ml. An anti-V5 antibody (Invitrogen, CA) was used for the Western Blot. From this Western Blot experiment, the clone expressing a protein encoded by CLN00717648 produced the highest level of secreted protein in the conditioned media. These results were confirmed by large-scale expression in 293-6E cells.

Example 5. Classification of the Signal Peptides/Leader Sequences of the Invention on the Basis of their Ability to Enhance Secretion and/or Expression of Heterologous Proteins

[00164] The combined results from the experiments described in Examples 1-4, suggest a classification of the leader sequences of the invention according to their ability to, in their role as heterologous leader sequences, improve secretion and/or expression of the proteins they are inserted into. The leader sequences are accordingly classified under categories such as “high secretor signal peptide/secretory leaders,” “moderate secretory signal peptide/secretory leaders,” or “low secretory signal peptide/secretory leader sequences.”

[00165] Because the secretion levels and the increases in secretion caused by the heterologous polypeptide of the invention is separate and distinct from the expression levels of the resulting polypeptides, the resulting polypeptides were also ranked on the basis of their expression levels on a relative scale that served to rank all the proteins of the invention (Tables 1-3 and Appendix A) relatively to each other. These rankings were made for expression and secretion levels in either wheat germ extracts, or mammalian cells (*see* Examples 1-3).

[00166] Moreover, whereas the above classification is based on the results obtained from using *in vitro* assays, the classification extends to results that can be obtained while expressing the proteins of the invention *in vivo*. As already discussed in Example 1, the signal peptides/leaders sequences of the invention can be assayed for their ability to be used to improve the *in vivo* expression of heterologous proteins they are attached to. For example, any of the leader sequences described in Table 2 can be operatively linked to an heterologous protein using cloning methods essentially as described in Examples 1 and 2. The resulting cDNA construct can then be electroporated or microinjected into embryonic stem (ES) cells (for example, mouse or pig ES cells), which are then used, according to standard methods known to those skilled in the art, for generating transgenic animals (*e.g.* mice or pigs). Depending on the protein, and on other properties of the cDNA construct (for example, the specific promoter used to drive expression of the recombinant protein), the secreted recombinant protein can be assayed from bodily fluids such as, for example, blood, milk, saliva, and its expression levels quantified. The assay can be done such that two recombinant proteins are expressed that vary only by their signal peptide (*i.e.*, comparing endogenous signal peptide and heterologous signal peptide of the invention).

[00167] It is possible that the signal peptide/leader sequences of the invention do not fall into the same categories when, instead of being used for protein expression *in vitro* they are used for protein expression *in vivo*. However, the results from the *in vitro* assays described herein should serve as guidelines for choosing which particular signal peptide/leader sequence one can use in order to achieve the desired levels of protein expression both *in vitro* and *in vivo*.

[00168] The specification is most thoroughly understood in light of the following references, all of which are hereby incorporated by reference in their entireties. The disclosures of the patents and other references cited above are also hereby incorporated by reference.

1. Agrawal, S. et al. eds. (1998) Antisense Research and Application (Handbook of Experimental Pharmacology, v. 131). Springer-Verlag NY, Inc.
2. Andreeff, M. et al. eds. (1999) Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications. John Wiley & Sons, Inc., New York, NY.
3. Ansel, H.C. et al. eds. (1999) Pharmaceutical Dosage Forms and Drug Delivery Systems. 7th ed. Lippencott Williams & Wilkins Publishers.
4. Beigelman, L. et al. (1995) *Nucleic Acids Res.* 23:4434-4442.
5. Chen, S.Y. et al. (1994) *Hum. Gene Ther.* 5:595-601.
6. Cheng, W.F. et al (2001) *J. Clin. Invest.* 108:669-678.
7. Chien, C. et al (1991) *Proc. Natl. Acad. Sci.* 88:9578-9581.
8. Coligan, J.E. et al. eds. (2002) Current Protocols in Immunology, vols. 1-4, including quarterly suppl.) John Wiley & Sons, Inc., New York, NY.
9. Deutscher, M.P. et al. eds. (1990) Guide to Protein Purification: Methods in Enzymology. (Methods in Enzymology Series, v. 182). Academic Press.
10. Dieffenbach, C.W. et al. eds. (1995) PCR Primer: A Laboratory Manual. Cold Spring Harbor Laboratory Press.
11. Durocher, Y. et al. (2002) *Nucleic Acids Res.* 30(2) e9.
12. Fields, S. et al. (1989) *Nature* 340:245-246.
13. Fukuhara, A. et al. (2004) Scienceexpress @ [www.sciencexpress.org/16 December 2004/Page 1/10.1126/science.1097243](http://www.sciencexpress.org/16%20December%202004/Page%201/10.1126/science.1097243).
14. Furth, P.A. et al. (1992) *Anal. Biochem.* 205:365-368.
15. Gaudilliere, B. et al. (2002) *J. Biol. Chem.* 277:46442-46446.

16. Gennaro, A., ed. (2000) Remington: The Science and Practice of Pharmacy. 20th ed. Lippincott, Williams, & Wilkins.
17. Gorman, C.M. et al. (1982) *Proc. Natl. Acad. Sci.* 79:6777-6781.
18. Grosschedl, R. et al. (1985) *Cell* 41:885-897.
19. Grosveld, F. et al. eds. (1992) Transgenic Animals. 1st ed. Academic Press.
20. Harlow, E. et al. eds. (1988) Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory.
21. Harlow, E. et al. eds. (1998) Using Antibodies: A Laboratory Manual: Portable Protocol No. I. Cold Spring Harbor Laboratory.
22. Hartmann, G. et al. eds. (1999) Manual of Antisense Methodology (Perspectives in Antisense Science). 1st ed. Kluwer Law International.
23. Hassanzadeh, G.H.G. et al. (1998) *FEBS Lett.* 437:75-80.
24. Heiser, A. et al. (2002) *J. Clin. Invest.* 109:409-417.
25. Hirschberg, C. (1987) *Annu. Rev. Biochem.* 56:63-87.
26. Hoogenboom, H.R. et al. (1998) *Immunotechnology* 4:1-20.
27. Howard, G.C. et al. (2000) Basic Methods in Antibody Production and Characterization. CRC Press.
28. Jameson, D.M. et al. (1995) *Methods Enzymol.* 246:283-300.
29. Jia, S.H. et al. (2004) *J. Clin. Investigation* 113(9): 1318-1327.
30. Jones, P. ed. (1998a) Vectors: Cloning Applications: Essential Techniques, John Wiley & Sons, Inc., New York, NY.
31. Jones, P. ed. (1998b) Vectors: Expression Systems: Essential Techniques, John Wiley & Sons, Inc., New York, NY.
32. Jost, C.R. et al. (1994) *J. Biol. Chem.* 269:26,267-26,273.
33. Kabat, E.A. et al. (1991) *J. Immunol.* 147:1709-1719.
34. Kibbe, A.H., ed. (2000) Handbook of Pharmaceutical Excipients. 3rd ed. Pharmaceutical Press.
35. Kirkpatrick, K.L. et al. (2001) *Eur. J. Surg. Oncol.* 27:754-760.
36. Knutson, K.L. et al. (2001) *J. Clin. Invest.* 107:477-484.
37. Kolonin, M.G. et al. (1998) *Proc. Natl. Acad. Sci.* 95:14,266-14,271.
38. Liu, A.Y. et al. (1987) *Proc. Natl. Acad. Sci.* 84:3439-3443.
39. Liu, A.Y. et al. (1987) *J. Immunol.* 139:3521-3526.
40. Machiels, J.P. et al. (2002) *Semin. Oncol.* 29:494-502.

41. Massie, B. et al. (1998) *J. Virol.*, 72, 2289-2296).
42. Matz, M.V. et al. (1999) *Nat. Biotechnol.* 17:969-973.
43. Mayer, B.J. (2001) *J. Cell Sci.* 114:1253-1263.
44. Milligan, J.F. et al. (1993) *J. Med. Chem.* 36:1923-1937.
45. Mitchell, D.A. et al. (2000) *J. Clin. Invest.* 106:1065-1069.
46. Mitsumoto, Y. et al. (1991) *Biochem. Biophys. Res. Commun.* 175: 652-9.
47. Mitsumoto, Y. et al. (1992) *J. Biol. Chem.* 267: 4957-4962.
48. Okayama, H. et al. (1983) *Mol. Cell. Biol.* 3:280-289.
49. O'Neil, N.J. et al. (2001) *Am. J. Pharmacogenomics* 1:45-53.
50. Peelle, B. et al. (2001) *J. Protein Chem.* 20:507-519.
51. Pertl, U. et al. (2003) *Blood* 101:649-654.
52. Phillips, M.I., ed. (1999) Antisense Technology, Part A. Methods in Enzymology
Vol. 313. Academic Press, Inc.
53. Phillips, M.I., ed. (1999) Antisense Technology, Part B. Methods in Enzymology
Vol. 314. Academic Press, Inc.
54. Pinkert, C.A., ed. (1994) Transgenic Animal Technology: A Laboratory
Handbook. Academic Press.
55. Remington, J.P. (1985) Remington's Pharmaceutical Sciences. 17th ed. Mack
Publishing Co.
56. Samal, B. et al. (1994). *Mol. Cell. Biol.* 14(2): 1431 – 1437.
57. Sambrook, J. et al. eds. (1989) Molecular Cloning, A Laboratory Manual. 2nd
ed. Cold Spring Harbor Laboratory Press.
58. Schoen, F. J. (1994) Robbins Pathologic Basis of Disease. W.B. Saunders Co.,
Philadelphia, PA.
59. Stein, C.A. et al. eds. (1998) Applied Antisense Oligonucleotide
Technology. Wiley-Liss.
60. Tang, D.C. et al. (1992) *Nature* 356:152-154.
61. Wagner, R.W. et al. (1996) *Nat. Biotechnol.* 14:840-844.
62. Wagner, R.W. et al. (1993) *Science* 260:1510-1513.
63. Xu, C.W. et al. (1997) *Proc. Natl. Acad. Sci. (USA)* 94:12473-12478.
64. Xu, Y. et al. (1999) *Proc. Natl. Acad. Sci.* 96:151-156.
65. Yu, Z. et al. (2002) *J. Clin. Invest.* 110:289-294.
66. Zallipsky, S. (1995) *Bioconjugate Chem.*, 6:150-165.

67. Zhu, J. et al. (1997) *Proc. Natl. Acad. Sci.* 94:13,063-13,068.

68. Zavyalov, et al. (1997) AP 105(3):161-186.

Sequence Listing

[00169] Applicants include a Sequence Listing provided in both electronic and paper format as Appendix A.

Industrial Applicability

[00170] The leader sequences, heterologous secreted polypeptides, nucleic acids, vectors, host cells and methods of making these find use in a number of investigative, diagnostic, and therapeutic applications.

Claims

[00171] 1. A heterologous polypeptide comprising a secretory leader and a second polypeptide, wherein the secretory leader is operably linked to the N-terminal of the second polypeptide, wherein the secretory leader is not so linked to the second polypeptide in nature, and wherein the secretory leader comprises a leader sequence of a secretable protein.

[00172] 2. The heterologous polypeptide of claim 1, wherein the second polypeptide is a secretable protein selected from collagen type IX alpha 1 chain, long splice form, alpha-2-antiplasmin precursor (alpha-2-plasmin inhibitor), trinucleotide repeat containing 5, ARMET protein, calumenin, COL9A1 protein, NBL1, PACAP protein, alpha-1B-glycoprotein precursor (alpha-1-B glycoprotein), similar to brain-specific angiogenesis inhibitor 2 precursor, SPOCK2, protein disulfide-isomerase (EC 5341) ER60 precursor, serine (or cysteine) proteinase inhibitor, clade A (alpha-1), GM2 ganglioside activator precursor, coagulation factor X precursor, secreted phosphoprotein 1 (osteopontin, bone sialoprotein 1), Vitamin D-binding protein precursor, interleukin 6 (interferon, beta 2), orosomucoid 1 precursor, hemopexin, glycoprotein hormones, alpha polypeptide precursor, kininogen 1, prolyl 4-hydroxylase, beta subunit, proopiomelanocortin, prostaglandin D2 synthase 21 kDa, alpha-2-glycoprotein 1, zinc, chromogranin A, cystatin M precursor, clusterin isoform 1, inter-alpha (globulin) inhibitor H1, leukemia inhibitory factor (cholinergic differentiation factor), lumican, secretoglobin, family 2A, member 2, nov precursor, reticulocalbin 1 precursor, reticulocalbin 2, EF-hand calcium binding domain, gastric intrinsic factor (vitamin B synthesis), cerberus 1, lipocalin 2 (oncogene 24p3), interleukin 18 binding protein isoform C precursor, cell growth regulator with EF hand domain 1, leukocyte immunoglobulin-like receptor, subfamily A, spondin 2, extracellular matrix protein, transmembrane protein 4, spar/osteonectin, cwcv and kazal-like domain proteoglycan, Rho GTPase activating protein 25 isoform b, dickkopf homolog 3, ameloblastin precursor, chorionic gonadotropin, beta polypeptide 8 precursor, multiple coagulation factor deficiency 2, similar to common salivary protein 1, hypothetical protein FLJ32115, oncoprotein-induced transcript 3, hypothetical protein MGC40499, interleukin 18 binding protein isoform A precursor, interleukin 1 receptor antagonist isoform 1 precursor, WFIKK2 protein, similar to hypothetical protein 9330140G23, and SEQ ID NOs: 2-3, 9, 19, 22, 26, 28, 31, 37, 41, 47, 54, 57, 62, 68, 75, 79, 82, 86, 88, 94, 97, 102,

104, 107, 111, 116, 120, 127, 131, 137, 140, 145, 147, 153, 159, 167, 175, 177, 181, 185, 189, 191, 196, 200, 207, 209, 215, 218, 222, 227, 232, 235, 239, 241, 245, 248, and 254.

[00173] 3. The heterologous polypeptide of claim 1, wherein the secretory leader comprises an amino acid sequence selected from SEQ ID NOs: 20 – 21, 23-25, 27, 32-36, 38-40, 48-53, 76-78, 80-81, 83-85, 87, 95-96, 103, 108-110, 112-115, 117-119, 121-126, 128-130, 132-136, 138-139, 141-144, 154-158, 160-166, 178-180, 186-188, 197-199, 210-214, 223-226, 233-234, 240, and 246-247.

[00174] 4. The heterologous polypeptide of claim 1, wherein the second polypeptide is selected from a secretable polypeptide, an extracellular portion of a transmembrane protein, and a soluble receptor.

[00175] 5. The heterologous polypeptide of claim 4, wherein the secretable polypeptide is selected from a growth factor, a cytokine, a lymphokine, an interferon, a hormone, a stimulatory factor, an inhibitory factor, a soluble receptor, and splice variants thereof.

[00176] 6. A secretory leader comprising a leader amino acid sequence selected from the leader sequences of the secretable polypeptides of Table 1 and the secretory leaders listed in Table 2.

[00177] 7. The secretory leader of claim 6, the amino acid sequence of which is selected from the amino acid sequences of Appendix A, the amino acids residues of SEQ ID NOs: 1, 4-8, 10-18, 20-21, 23-25, 27, 29-30, 32-36, 38-40, 42-46, 48-53, 55-56, 58-61, 63-67, 69-74, 76-78, 80-81, 83-85, 87, 89-93, 95-96, 98-101, 103, 105-106, 109-110, 112-115, 117-119, 121-126, 128-130, 132-136, 138-139, 141-144, 146, 148-152, 154-158, 160-166, 168-174, 176, 178-180, 182-184, 186-188, 190, 192-195, 197-199, 201-206, 208, 210-214, 216-217, 219-221, 223-226, 228-231, 233-234, 236-238, 240, 242-244, 246-247, 249-253, and 255-256.

[00178] 8. The heterologous polypeptide of claim 1, further comprising a fusion partner.

[00179] 9. The heterologous polypeptide of claim 8, wherein the fusion partner is a polymer.

[00180] 10. The heterologous polypeptide of claim 9, wherein the polymer is a third molecule, and wherein the third molecule is selected from polyethylene glycol and all or part of human serum albumin, fetuin A, fetuin B and Fc.

[00181] 11. An isolated nucleic acid molecule comprising a polynucleotide sequence selected from: (1) a polynucleotide sequence encoding an amino acid sequence of a heterologous polypeptide according to any one of claims 1-5 and claims 8-10; (2) a polynucleotide encoding an amino acid sequence of a secretory leader according to any one of claims 6-7.

[00182] 12. A nucleic acid molecule encoding a heterologous polypeptide, comprising a first polynucleotide that encodes a secretory leader of any one of claims 6-7, a second polynucleotide that encodes a second polypeptide, wherein the first polynucleotide and the second polynucleotide are operably linked to facilitate secretion of the heterologous polypeptide from a cell, and wherein the first and second polynucleotide are not so linked in nature.

[00183] 13. The nucleic acid of claim 12, wherein the second polypeptide is selected from a secretable polypeptide, an extracellular portion of a transmembrane protein, and a soluble receptor.

[00184] 14. The nucleic acid molecule of claim 12, further comprising a third polynucleotide, wherein the third polynucleotide is a Kozak sequence or a fragment thereof that is situated at its 5' end.

[00185] 15. The nucleic acid molecule of claim 14, further comprising a fourth polynucleotide, wherein the fourth polynucleotide comprises a restriction enzyme-cleavable sequence at its 3' end.

[00186] 16. The nucleic acid molecule of claim 15, further comprising a fifth polynucleotide that encodes a tag.

[00187] 17. The nucleic acid molecule of claim 16, wherein the tag is a purification tag.

[00188] 18. The nucleic acid molecule of claim 16, wherein the tag is selected from V5, HisX6, HisX8, an avidin molecule, and a biotin molecule.

[00189] 19. The nucleic acid molecule of claim 16, further comprising a sixth polynucleotide that encodes a second enzyme-cleavable sequence that can be cleaved by a second enzyme, wherein the second cleavable sequence is situated upstream of the tag if the tag is situated at the C-terminus of the heterologous polypeptide, or downstream of the tag if the tag is situated at the N-terminus of the heterologous polypeptide.

[00190] 20. The nucleic acid molecule of claim 19, wherein the second enzyme is thrombin or TEV from a tobacco virus.

- [00191] 21. A vector comprising the nucleic acid molecule of any one of claims 11-20, further comprising an origin of replication and a selectable marker.
- [00192] 22. The vector of claim 21, wherein the origin of replication is selected from SV40 ori, Pol ori, EBNA ori, and pMB1 ori.
- [00193] 23. The vector of claim 21, wherein the selectable marker is an antibiotic resistance gene.
- [00194] 24. The vector of claim 23, wherein the antibiotic resistance is selected from puromycin resistance, kanamycin resistance, and ampicillin resistance.
- [00195] 25. A recombinant host cell comprising a cell and the heterologous polypeptide of any of claims 1-4 and claims 8-10, the nucleic acid molecule of any of claims 11-20, or the vector of any one of claims 21-24.
- [00196] 26. The recombinant host cell of claim 25, wherein the cell is a eukaryotic cell.
- [00197] 27. The recombinant host cell of claim 26, wherein the cell is a human cell.
- [00198] 28. A method of producing a secreted polypeptide, comprising:
- [00199] (a) providing the nucleic acid molecule of any of claims 11-20; and
- (b) expressing the nucleic acid molecule in an expression system.
- [00200] 29. The method of claim 28, wherein the expression system is a cellular expression system or a cell free expression system.
- [00201] 30. The method of claim 28, wherein the expression system is a cellular expression system and the cell is a mammalian cell.
- [00202] 31. The method of claim 30, wherein the mammalian cell is selected from a 293 cell line, a PER.C6® cell line, and a CHO cell line.
- [00203] 32. The method of claim 31, wherein the 293 cell is a 293-T cell or a 293-6E cell.

```

CLUSTAL FORMAT for T-COFFEE Version_1.37, CPU=0.00 sec, SCORE=61, Nseq=3, Len=921

b.CLN00517648_5pv1      MKTCWKIDVFFVFCSEFLEPWASAAVKRRPRFPVNSNSNGNELCPKIRIGQDDLPGLDI 60
a.collagen_leader      MKTCWKIDVFFVFCSEFLEPWASA----- 23
c.NP_001842_NM_001851  MKTCWKIDVFFVFCSEFLEPWASAAVKRRPRFPVNSNSNGNELCPKIRIGQDDLPGLDI 60
*****

b.CLN00517648_5pv1      SQFQVDKAASRRRAIQRVVGSATLQVAYKLGNNVDRIPTRNLYPSGLPEEYSFLITFRMT 120
a.collagen_leader      SQFQVDKAASRRRAIQRVVGSATLQVAYKLGNNVDRIPTRNLYPSGLPEEYSFLITFRMT 23
c.NP_001842_NM_001851  SQFQVDKAASRRRAIQRVVGSATLQVAYKLGNNVDRIPTRNLYPSGLPEEYSFLITFRMT 120

b.CLN00517648_5pv1      GSTLKKWNINWQIDSSGKEQVGKINGQTQSVVFSYKGLDGLSLQTAAFSNLSSLFDSQW 180
a.collagen_leader      GSTLKKWNINWQIDSSGKEQVGKINGQTQSVVFSYKGLDGLSLQTAAFSNLSSLFDSQW 23
c.NP_001842_NM_001851  GSTLKKWNINWQIDSSGKEQVGKINGQTQSVVFSYKGLDGLSLQTAAFSNLSSLFDSQW 180

b.CLN00517648_5pv1      HKIMIGVERSSATLFDVDCNRIESLPIKPRGPIDIDGFAVLGKLADNPQVSVPFELQWMLI 240
a.collagen_leader      HKIMIGVERSSATLFDVDCNRIESLPIKPRGPIDIDGFAVLGKLADNPQVSVPFELQWMLI 23
c.NP_001842_NM_001851  HKIMIGVERSSATLFDVDCNRIESLPIKPRGPIDIDGFAVLGKLADNPQVSVPFELQWMLI 240

b.CLN00517648_5pv1      HCDPLRPXRETCH----- 253
a.collagen_leader      HCDPLRPXRETCH----- 23
c.NP_001842_NM_001851  HCDPLRPXRETCHLPAITPSQTDERGPPGQGPAGSGPPGPGIDGIDGDRGKGP 300

b.CLN00517648_5pv1      ----- 253
a.collagen_leader      ----- 23
c.NP_001842_NM_001851  PGPFGPGEPPGKPGAPGKFTPGADGLTGPDGSPGSGIGSKGKGKGPVPGSGFPGRGIP 360

b.CLN00517648_5pv1      ----- 253
a.collagen_leader      ----- 23
c.NP_001842_NM_001851  GPPGPGGTAGLPQELGRVGVGDPGCRGPPGPPGPGPRGTGTFHDDGPLCPNACPPGRS 420

b.CLN00517648_5pv1      ----- 253
a.collagen_leader      ----- 23
c.NP_001842_NM_001851  GYPGLPGMRGHKAKEIGEPGRQGHKGEGDQGLGEVGAQGPQGAQGLAGITGIVGDK 480

b.CLN00517648_5pv1      ----- 253
a.collagen_leader      ----- 23

```

Figure 1

c.NP_001842_NM_001851	CEKGARGLDGEPGPGQCLPAPFDQGGQGFGEAGPKGDRGAEGARGIGPLPGPKGDTGLP	540
b.CLN00517648_5pv1	-----	253
a.collagen_leader	-----	23
c.NP_001842_NM_001851	GVDGRDGIPLGMPGKGEPPGKPPGDAQLQGLPGVPGIPGAKGVAGEKSTGAPGKPGQM	600
b.CLN00517648_5pv1	-----	253
a.collagen_leader	-----	23
c.NP_001842_NM_001851	GNSGKPGQQGPPGGEVGPGRPGQLPGSRBELPVGSPGLPGKLSLGSFGLPGLPGPPGLP	660
b.CLN00517648_5pv1	-----	253
a.collagen_leader	-----	23
c.NP_001842_NM_001851	GMKGRGVVGPBPKEGQASGEAGEGERGELGDIGLPGPKSAGNPGEPLGRGPEGR	720
b.CLN00517648_5pv1	-----	253
a.collagen_leader	-----	23
c.NP_001842_NM_001851	GLPGVEGPRGPPGPRGVQGEQATGLPGVQPPGCRAPTDQHIKQVCMRVIOEHFAEMAAS	780
b.CLN00517648_5pv1	-----	253
a.collagen_leader	-----	23
c.NP_001842_NM_001851	LKRPDSGATGLPKRPGPPGPPPPGNGFPQNGIRGLPGIKGPPGALGLRPGKGLGEK	840
b.CLN00517648_5pv1	-----	253
a.collagen_leader	-----	23
c.NP_001842_NM_001851	GERGPPGRGPNGLPGAIGLRPGDPPASYGKNGRDGERGPPGLAGIPGVGPPGPPGLPGF	900
b.CLN00517648_5pv1	-----	253
a.collagen_leader	-----	23
c.NP_001842_NM_001851	CEPASCTMQAGQRAFNKGPPDP	921

Figure 1

Identification of Highly Expressed Human cDNAs in 293T cells using pTT5 Expression Vector

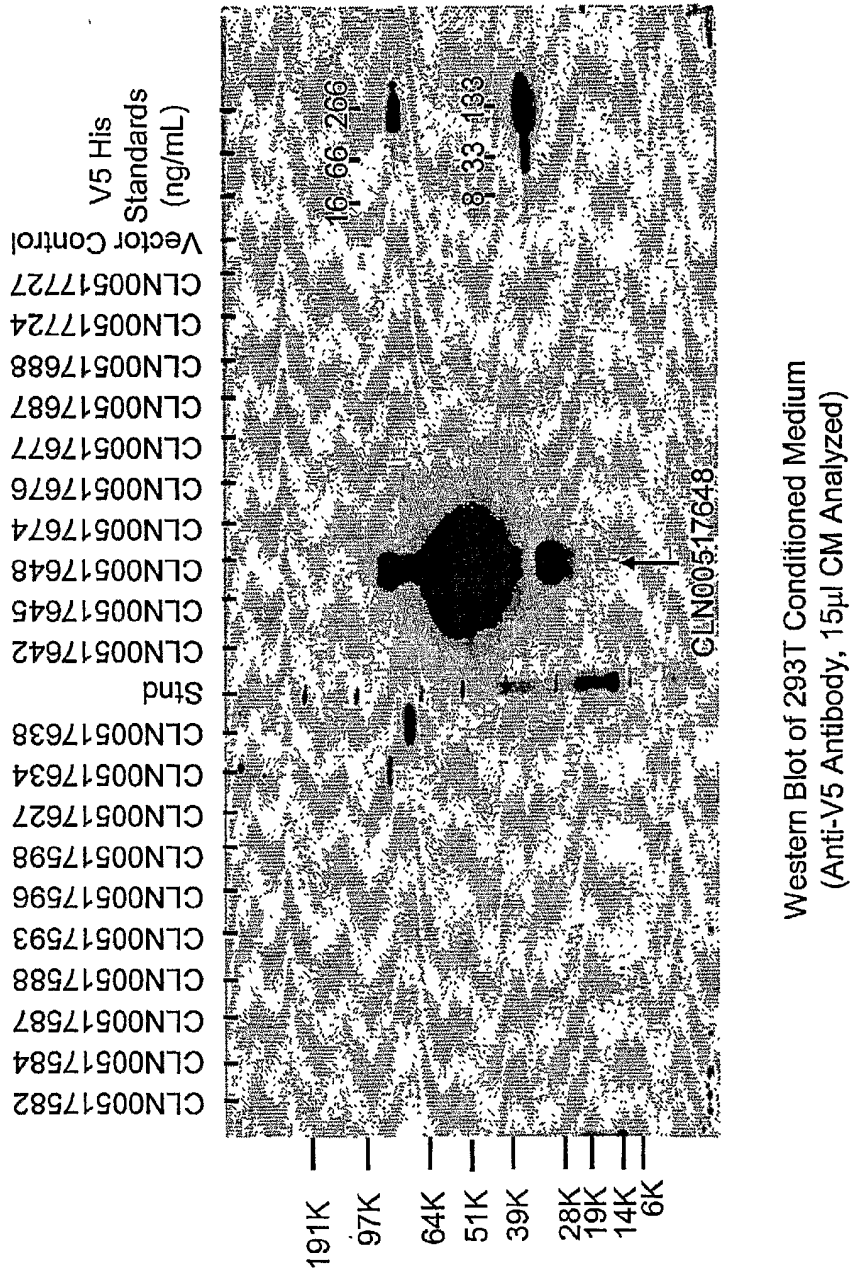


Figure 2

4/10

Vector for Type II cloning the Soluble Portion of the Proteins: Back Bone Modifications

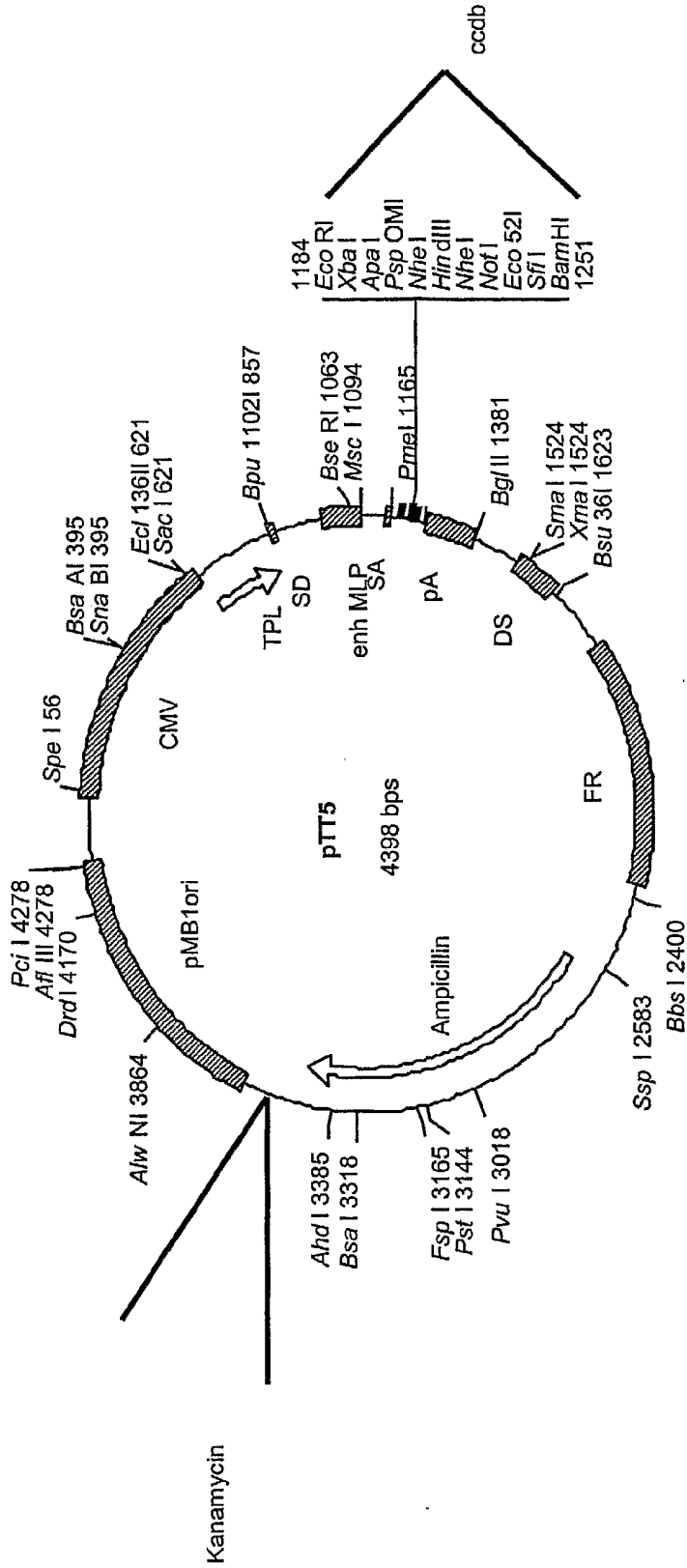


Figure 3

Vector for Secreted Protein with a Cleavable Tag

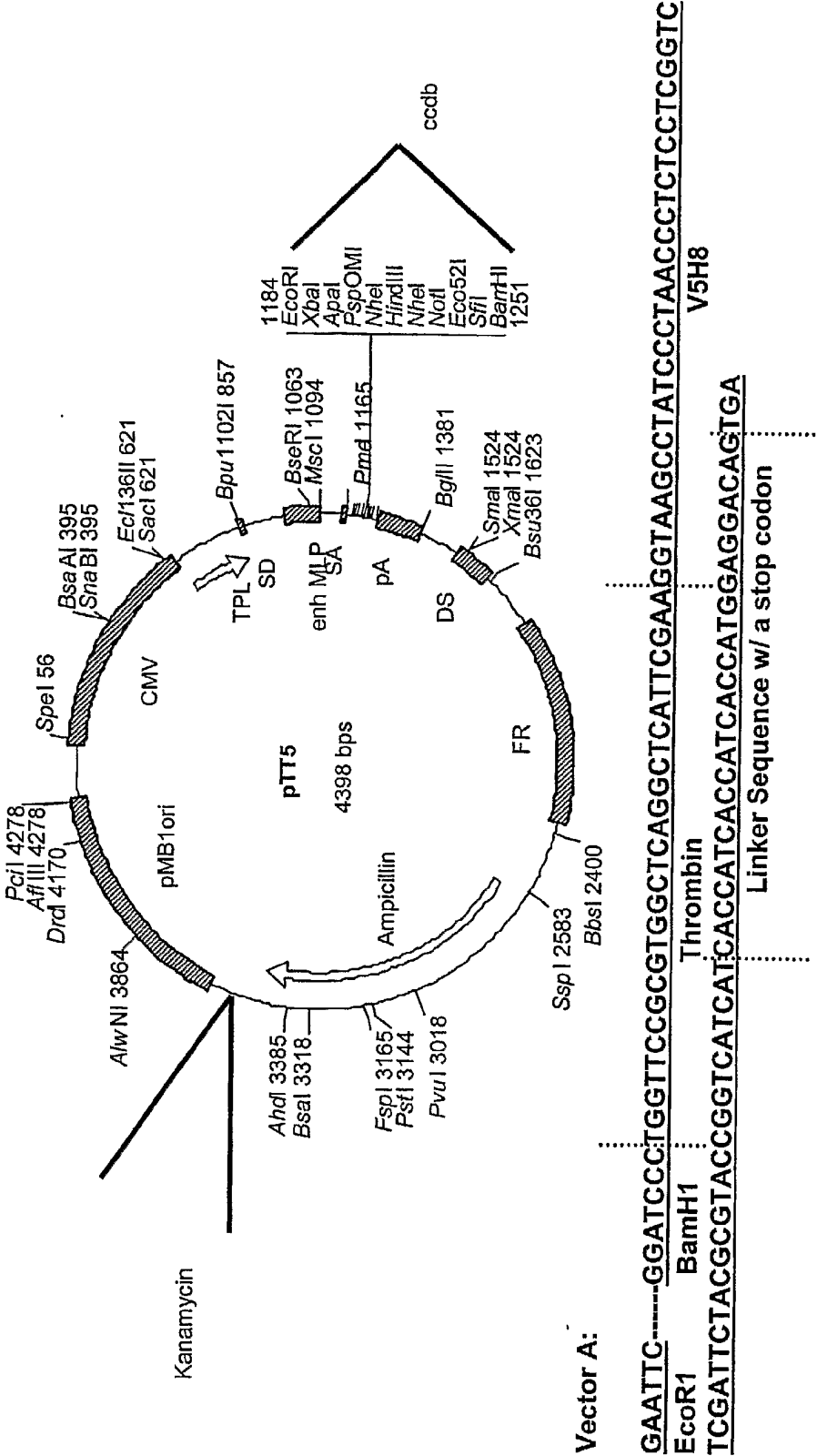


Figure 4

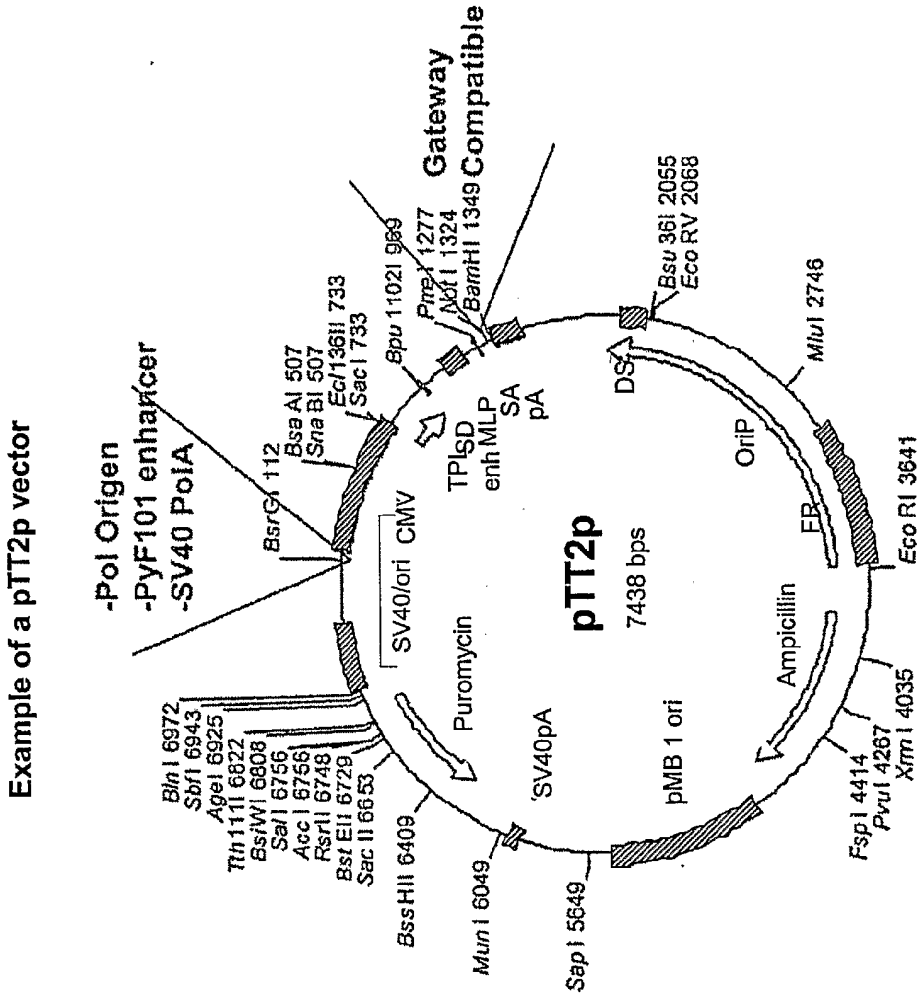
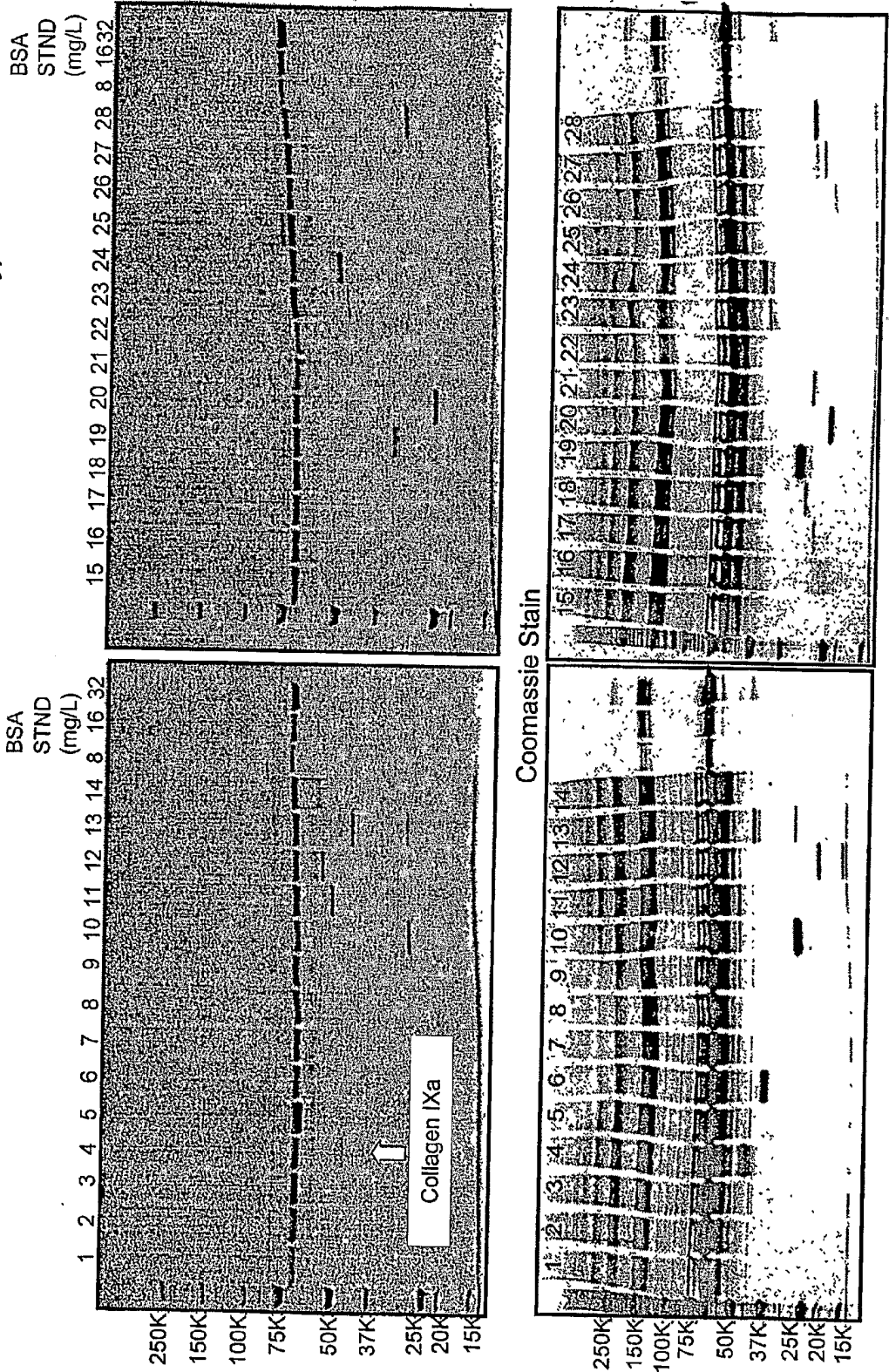


Figure 7

SDS-PAGE Analysis of High Expressor Clone (48H Sup in CHO Soy)



Silver Stain
Figure 8

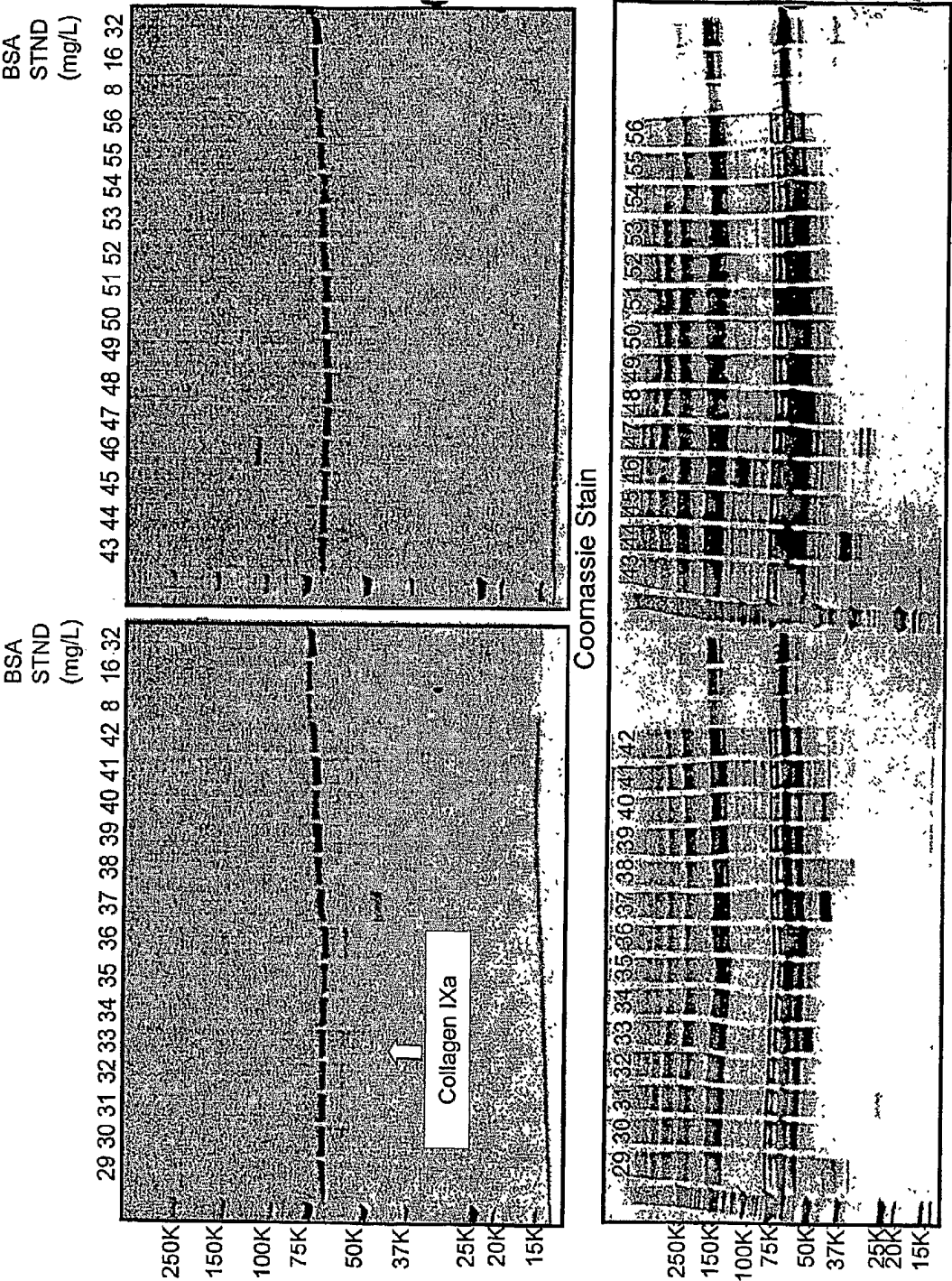


Figure 9

SEQ.ID.NO. 1 HG1018265P1 collagen_leader_seq
MKTCKWIPVFFVCSFLEPWASA

SEQ.ID.NO. 2 HG1018266P1 CLN00517648
MKTCKWIPVFFVCSFLEPWASAAVKRRPRFPVNSNSNGNELCPKIRIGQDDLPGFLLI
SQFQVDAKASRAIQRVVGSATLQVAYKLGNNVDRIPTRNLYPSGLPEEYSFLTFRMT
GSTLKNWNIWQIQDSSGKEQVGIKINGQTSVVSFYKGLDGLQTAAFSNLSLFDQW
HKIMIGVERSSATLFDVDCNRIESLPKPRGPDIDGFAVLGKLADNPQVSPFELQWMLI
HCDPLRRRETCHLPPARITPSQTTDERGPPGEQGGPPGPPGPDIDGDRGPKGP
PGPPGPGAGEPGKPGAPKPGTGPADTSPYPAFLYKVVIKVPINPPLLGLDSTRTGH
HHHHGGGQ

SEQ.ID.NO. 3 HG1018267P1 112907:21594845
MALLWGLLVLSWCLQGPCSVFSPVSAMEPLGRQLTSGPNQEQVSPPLTLKLGNOEPGGQ
TALKSPPGVCSRDPTPEQTHRLARAMMAFTADLFSLVAQSTCTCNLILSPLSVALASHL
ALGAQNHTLQRLQQLVHAGSGPCPLPHLLSRLCQDLGPGAFRLAARMYLQKGFPIKEDFLB
QSEQLFGAKPVSLTGQEDDLANINQWKEATEGKIQEFSLGPELDTVLILLNAINHFQGF
WRNKFDPSLTQRDSFHLDEQFTVPVEMMQARTYPLRWFLLEQPEIQVAHFPPKNNMSFVV
LVPTHEWNVSVLANLSWDTLHPPLVWERPTKVRLPKLYLKHQMDIVATLSQLGLQELF
QAPDLRIGISEQSLVVGVSQVHQSTLELSEVGVEAAAATSIAMSRMSLSFSVNRPFLLFFIF
EDTTGLPLFVGSVNRPNPSPARELKEQQDPSGNKDFLQSLKGFPRGDKLFGPDLKLVP
BEDYPQFGSPK

SEQ.ID.NO. 4 HG1018268P1 112907:21594845_1-17
MALLWGLLVLSWCLQG

SEQ.ID.NO. 5 HG1018269P1 112907:21594845_1-13
MALLWGLLVLSWS

SEQ.ID.NO. 6 HG1018270P1 112907:21594845_1-19
MALLWGLLVLSWCLQGPC

SEQ.ID.NO. 7 HG1018271P1 112907:21594845_1-16
MALLWGLLVLSWCLQ

SEQ.ID.NO. 8 HG1018272P1 112907:21594845_1-15
MALLWGLLVLSWCL

SEQ.ID.NO. 9 HG1018273P1 13325208:13325207
MDSMPAPASRCLLLPLLLLLLLPAPELGPSQAGAEENDWVRLPSKCEVCKYVAVELK
SAFETGKTKEVIGTYGILDQKASGVKTKSDLRLEIETETICKRLLDYSLHKERTGSN
RFAKGMSETFETLHNLVHKGVKVMVDIPYELWNETSAAEVADLKKQCDLVVEEFEEVIEDW
YRNHQEDLTTEFLCANHVLKGDTSCLAEQWSGKGKGTAAALGGKSKKSSRAKAAGRS
SSSKQRKELGGLEGDPSPEDDEGIQKASPLTHSPDEL

SEQ.ID.NO. 10 HG1018274P1 13325208:13325207_1-30
MDSMPEPASRCLLLLPLLLLLLLLLLPAPEL

SEQ.ID.NO. 11 HG1018275P1 13325208:13325207_1-25
MDSMPEPASRCLLLLPLLLLLLLLLL

SEQ.ID.NO. 12 HG1018276P1 13325208:13325207_1-33
MDSMPEPASRCLLLLPLLLLLLLLLLPAPELGPS

SEQ.ID.NO. 13 HG1018277P1 13325208:13325207_1-24
MDSMPEPASRCLLLLPLLLLLLLLLL

SEQ.ID.NO. 14 HG1018278P1 13325208:13325207_1-26
MDSMPEPASRCLLLLPLLLLLLLLLL

SEQ.ID.NO. 15 HG1018279P1 13325208:13325207_1-32
MDSMPEPASRCLLLLPLLLLLLLLLLPAPELGP

SEQ.ID.NO. 16 HG1018280P1 13325208:13325207_1-27
MDSMPEPASRCLLLLPLLLLLLLLLLPA

SEQ.ID.NO. 17 HG1018281P1 13325208:13325207_1-23
MDSMPEPASRCLLLLPLLLLLLLLLL

SEQ.ID.NO. 18 HG1018282P1 13325208:13325207_1-35
MDSMPEPASRCLLLLPLLLLLLLLLLPAPELGPSQA

SEQ.ID.NO. 19 HG1018283P1 13938307:13938306
MRRMWATQGLAVALALSVLPGSRALRPGDCEVCISYLGRFYQDLKDRDVTFSPATIENEL
IKFCREARGKENRLCYIGATDDAATKIINEVSKPLAHIPVEKICEKLLKKKQSQICELK
YDKQIDLSTVDLKKLRVKELKKILDWGETCKGCAEKSDYIRKINELMPKYAPKAASART
DL

SEQ.ID.NO. 20 HG1018284P1 13938307:13938306_1-24
MRRMWATQGLAVALALSVLPGSRA

SEQ.ID.NO. 21 HG1018285P1 13938307:13938306_1-21
MRRMWATQGLAVALALSVLPG

SEQ.ID.NO. 22 HG1018286P1 14718453:14718452
MDLRQFLMCLSLCTAFALSKPTKKORVHHEPQLSDKVHNDQAQSFYDHDHDAFLGAEEAKT
FDQLTPEESKERLGMIVDKIDADKDGFTGELKSWIKHAQKKYIYDNVENQWQEFDMNQ
DGLISWDEYRNVITYGTLYDDPDDGGFNKQMMVDRERRFKMADKDGDLIATKEEFTAFI

HPPEYDMKDIVVQETMEDIDKNADGFIDLEEYIGDMYSHDGNTEPEWVKTEREQFVEF
RDKNRDGMKDKETKDWILPSDYDHAEAEARHLVYESDQNKDGKLTKEEIVDKYDLFVGS
QATDFGEALVRHDEF

SEQ.ID.NO. 23 HG1018287P1 14718453:14718452_1-19
MDLRQFLMCLSLCTAFALS

SEQ.ID.NO. 24 HG1018288P1 14718453:14718452_1-15
MDLRQFLMCLSLCTA

SEQ.ID.NO. 25 HG1018289P1 14718453:14718452_1-17
MDLRQFLMCLSLCTAFA

SEQ.ID.NO. 26 HG1018290P1 15929966:15929965
MKTCKIPVFFFVCSFLEPWASAAVKRRPRFPVNSNSNGNELCPKIRIGQDDLPGFDLI
SQFQVDKAASRRAIQRVVGSA TLQVAYKLGNNVDFRIPTRNLVPSGLPEEYSFLTFRMT
GSTLKKNNWIQIDSSGKEQVGIKINGQTSVVFYSYKLDGSLQTAAFSNLSSLFDSQW
HKIMIGVERSSATLFDNCNRIESLPIKPRGPIIDGFVGLADNPQVSVPFELQWMLI
HCDPLRPRRETCHHELPARITPSQTDERGPPGEQPPGPPGPGIDGIDGDRGPKGP
PGPPGAGEPGKPGAPGKPGTGGADTSP

SEQ.ID.NO. 27 HG1018291P1 15929966:15929965_1-23
MKTCKIPVFFFVCSFLEPWASA

SEQ.ID.NO. 28 HG1018292P1 16356651:16356650
MMLRVLVGAVLPAMLLAAPPINKIALFPDKSAWCEAKNITQIVHSGCEAKSIQNRACL
GQCFYSVPNTFPQSTESLVHCDSCMPAQSMWEIVTLECPGHEEVPRVDKLVKILHCSC
QACGKEPSHEGLSVYVQGEDGPGSQPGTHPHPHPHPGGQTPEPEDPPGAPHTEEEGAE
D

SEQ.ID.NO. 29 HG1018293P1 16356651:16356650_1-21
MMLRVLVGAVLPAMLLAAPP

SEQ.ID.NO. 30 HG1018294P1 16356651:16356650_1-17
MMLRVLVGAVLPAMLLA

SEQ.ID.NO. 31 HG1018295P1 18204192:18204191
MRLSLPLLLLLGAWAIPGGLDRAPILTATAPQLDDEEMYSAHMPAHLRCDACRAVAYQM
WQNLAKAETKHLHTSNSGGRRELSLVYTDVLDRCSCRNWQDYGVRVDQVKRLTGPGLSE
GPEPSISVMVTGGPWPTRLSRCLHYLGEFGEDQIYEAHQQRGALEALLCGGPGGACSE
KVSATREEL

SEQ.ID.NO. 32 HG1018296P1 18204192:18204191_1-19
MRLSLPLLLLLGAWAIPG

SEQ. ID. NO. 33 HG1018297P1 18204192:18204191_1-22
MRLSLPLLLLLLGAWAIPGGLG

SEQ. ID. NO. 34 HG1018298P1 18204192:18204191_1-18
MRLSLPLLLLLLGAWAIP

SEQ. ID. NO. 35 HG1018299P1 18204192:18204191_1-16
MRLSLPLLLLLLGAWA

SEQ. ID. NO. 36 HG1018300P1 18204192:18204191_1-14
MRLSLPLLLLLLGA

SEQ. ID. NO. 37 HG1018301P1 23503038:15778555
MSMLVVFLLLLWGVTPVTEAIFYETQPSLWAESESLKPLANVTLTQCARLETPDFQL
FKNGVAQEPVHLDSPAIKHQFLLTGDTQGRYCRSGLSTGWTLGKLELTGPKSLPAPW
LSMAPVWITPGLKTTAVCGVLRGETFLLRREGDHEFLEVPEAQEDVEATFPVHQPGNY
SCSYRTDGEALSEPSATVTIEELAAPPVLMHHGESSQVLHPGNKVTLTCAVPLSGVD
FQLRGKELLVPRSSTPDRIFFHNAVALGDGGHYTCRYRLHDNQNGSGDSAPVELI
LSDETLPAPEFSPEPESGRALRLCLAPLEGARFALVREDRGRVHRFQSPAGTEALFE
LHNISVADSANSYCVYVDLKPFGGSAPSERLELHVDPGPPRPQRLRATWSGAALAGRDV
LRCEGPIDVTFFELLREGETKAVKTIPTPGAAANLELIFVGPQHAGNYRCRYRSWVPHTF
ESELSDPVVELLVAES

SEQ. ID. NO. 38 HG1018302P1 23503038:15778555_1-20
MSMLVVFLLLLWGVTPVTE

SEQ. ID. NO. 39 HG1018303P1 23503038:15778555_1-16
MSMLVVFLLLLWGVTWG

SEQ. ID. NO. 40 HG1018304P1 23503038:15778555_1-21
MSMLVVFLLLLWGVTPVTEA

SEQ. ID. NO. 41 HG1018305P1 27479535:27479534
MRALRDRAGLLLCVLLLAALLEAALGLPVKKPRLRGPGRGSLTRIAEVSASDPDRPLKEE
EEAPLLPRTHLQAEPHQGCWTVTEPAAMTPGNATPPRTPEVTPRLRLKLPGLANTTL
STPNPDQASASDPDRPLREEEARLLPRTHLQAEHLHQGCWTVTEPAALTTPGNATPPRT
QEVTPLLLEQLKLPVHATLSTPNPDNQVTIKVEDPQAEVSDILLAEPSNPPQDTLS
WLPALWSFLWGDYKGEKDRAPGKEGKEEDEDYPSIEDGEDQEDKEEDEEEQALWFN
GTTDNWDQGLAPGDWVFKOSVSYDYPQKEWSPWSPSCNGSTGKQQRTRPCGYCTAT
ETRCDLPSCEGTEDKDTLGLPSEWKLARNATMDHQDVDSCEKWLNCKSDFLIKYLS
QMLRDLPCPCPCLPAMDSPVSLQDEHQGRSFRWRDASGPRERLDIYQPTARFCLRSML
SGESSTLAAQHCYDEDSRLLTRGKGAMPNLISTDFSPLHFKFDTTTPWILCKGDWSRL
HAVLPNNNGRACTDNPLEEYLAQLQEAKEY

SEQ.ID.NO. 42 HG1018306P1 27479535:27479534_1-24
MRALRRAGLLLCVLLLAALLEAA

SEQ.ID.NO. 43 HG1018307P1 27479535:27479534_1-20
MRALRRAGLLLCVLLLAAL

SEQ.ID.NO. 44 HG1018308P1 27479535:27479534_1-26
MRALRRAGLLLCVLLLAALLEAALG

SEQ.ID.NO. 45 HG1018309P1 27479535:27479534_1-21
MRALRRAGLLLCVLLLAALL

SEQ.ID.NO. 46 HG1018310P1 27479535:27479534_1-23
MRALRRAGLLLCVLLLAALLEA

SEQ.ID.NO. 47 HG1018311P1 37182960:37182959
MRAPGCGRLVLP1LL1LAALAAEGDAKGLKEGETPGNFMEDEQWLSSISQYSGKIKHWNRR
FRDEVEDDY1KSWEDNQDEALD1TKPCQKVCSRHKVCIAQGYQYRAMCISRKLEHHR
IKQPTVKLHGNKDSICKPCMAQLASVCGSDGHTYSVCKLEQQACLSSKQLAVRCGPGC
PCPTEQAATSTADGKPETCTGQDLADLGDRLRDWFQLLHENSQKNGSSAVGAPSGILDK
SLGASCKDSIGWMFSKLDTSADLFLDQTELAALINLDKYEVCIRPFFNSCDTYKDGVRVSTA
EWCFCFWREKPPCLAELERIQIEAAKKPGIFIPSCDEGYYRKMQCDOSSGDCWRVDQ
LGLELTGTRTHGSPDCDDIVGFSGDFGSGVWEDEEEKETEEAGEAEAEAGEADDG
GYTW

SEQ.ID.NO. 48 HG1018312P1 37182960:37182959_1-24
MRAPGCGRLVLP1LL1LAALAAEG

SEQ.ID.NO. 49 HG1018313P1 37182960:37182959_1-19
MRAPGCGRLVLP1LL1LAAA

SEQ.ID.NO. 50 HG1018314P1 37182960:37182959_1-22
MRAPGCGRLVLP1LL1LAALAA

SEQ.ID.NO. 51 HG1018315P1 37182960:37182959_1-20
MRAPGCGRLVLP1LL1LAAAA

SEQ.ID.NO. 52 HG1018316P1 37182960:37182959_1-26
MRAPGCGRLVLP1LL1LAALAAEGDA

SEQ.ID.NO. 53 HG1018317P1 37182960:37182959_1-21
MRAPGCGRLVLP1LL1LAAAAAL

SEQ. ID. NO. 54 HG1018318P1 7437388:1208426
MRLRLALFPGVALLLAAGRLVAASDVLELTDNFEISRISDTGSAGLMLVEFFAPWCCHC
KRLAPEVEAAATRLKGIVPLAKVDCTANTCNKYGVSGYPTLKIFRDGEEAGAYDGPRT
ADGIVSHLKKQAGPASVPLRTEEEFKFISDKOASIVGFFDDSFSEAHSEFLKAASNLRD
NYRFAHTNVEISLVNEYDDNGEGIILFRPSHLTNKFEDKTVAYTEQKMTSGKIKKFTQENI
FGICPHMTENKDLIQGKDLLIAAYVDVYEKNAKGSNYWRNRMVMVAKKFLDAGHKLNFA
VASRKTFSHELSDFGLESTAGEIPVVAIRTAKEKFVMQEEFSRDGKALERFLQDYFDGN
LKRYLKTSEPIPSNDGPVKVVVAENFDEIVNNEKDVLIIEFYAPWCCHCKNLEPKYKELG
EKLKSDPNIVIAKMDATANDVSPYEVVRGFTIYFSPANKKLNPKKYEGGRELSDFISYL
QREATNPPVIQEEKPKKKKAQEDL

SEQ. ID. NO. 55 HG1018319P1 7437388:1208426_1-24
MRLRLALFPGVALLLAAGRLVAA

SEQ. ID. NO. 56 HG1018320P1 7437388:1208426_1-23
MRLRLALFPGVALLLAAGRLVA

SEQ. ID. NO. 57 HG1018321P1 NP_000286:NM_000295
MPSSVSWGILLIAGLCCLVPVSLAEDPQGDAAQKTDTSHHQDHPFTFNKITPNLAEFAPS
LYRQLAHQSNSTNIFSPVSIATAFAMLSLGTAKADTHDEILEGLNFNLTETPEAQIHGEF
QELLRTLNPQDSQLQLTGNGLFLSEGLKLVDFLEKLVKLYHSEAFVNFGEDEEAKKQ
INDYVEKGTQKIVDLVKELDRDTVFALVNYIFFKGKWERPFVEKDEEEDFHVDQVTVV
KVPMMKRLGMFNIQHCKKLSWVLLMKYLGNAIAIFFLPDEGLQLHLENELTHDIIITKFL
ENEDRSASLHPLKLSITGTYDLKSVLQGLGITKVFSNGADLSGVTEEAPLKLKSAVHKA
VLTIDEKGTAAAGAMFLEAIPMSIPPEVKFNKPFVFLMIDQNTKSPLFMGKVNPPTQK

SEQ. ID. NO. 58 HG1018322P1 NP_000286:NM_000295_1-24
MPSSVSWGILLIAGLCCLVPVSLA

SEQ. ID. NO. 59 HG1018323P1 NP_000286:NM_000295_1-18
MPSSVSWGILLIAGLCCL

SEQ. ID. NO. 60 HG1018324P1 NP_000286:NM_000295_1-23
MPSSVSWGILLIAGLCCLVPVSL

SEQ. ID. NO. 61 HG1018325P1 NP_000286:NM_000295_1-17
MPSSVSWGILLIAGLCC

SEQ. ID. NO. 62 HG1018326P1 NP_000396:NM_000405
MQSLMQAPILLIAGLLIATPAQAHKKPSQLSSFSWDCNCDGKDPVIRSLTEPDPIVV
PGNVTLSSVVGSTSVPLSSPLKVDLVLEKEVAGLWIKIPCTDYIGSCTFEHFCVDLMDLIP
TGEPCEPLRTYGLPCHCFKEGTYSLPKSEFVVDPDLPLPSWLITGNRYIESVLSSGKR
LGCIKIAASLKGI

SEQ.ID.NO. 63 HG1018327P1 NP_000396:NM_000405_1-23
 MQSLMQAPLLIALGLLITPAQA

SEQ.ID.NO. 64 HG1018328P1 NP_000396:NM_000405_1-18
 MQSLMQAPLLIALGLLIIA

SEQ.ID.NO. 65 HG1018329P1 NP_000396:NM_000405_1-25
 MQSLMQAPLLIALGLLITPAQAHL

SEQ.ID.NO. 66 HG1018330P1 NP_000396:NM_000405_1-20
 MQSLMQAPLLIALGLLITATP

SEQ.ID.NO. 67 HG1018331P1 NP_000396:NM_000405_1-21
 MQSLMQAPLLIALGLLITPA

SEQ.ID.NO. 68 HG1018332P1 NP_000495:NM_000504
 MGRPLHLVLLSASLAGLLLGESLFIHQANNILARVTRANSFLEEMKKGHLERECMEE
 TCSYEEAREVEFSDKTNEFWNKYKQDQCEFTSPCQKQKQKGLGHEYTCTCLEGFEKGN
 CELFTRKLCSLDNGDCDFCHEEQNSVVCSCARGYTIADNGKACITGYPYCGKQTLERR
 KRSVAQATSSGEAPDSITWKPYDAADLDPTENPFDLLDFNQTPQERGDNNLTRIVGGQE
 CKDGECPWQALLINENEGFCGGTILSEFYILTAACHLYQAKRFKRVGDRNTEQEEGGE
 AVHEVEVVIKHNRTKETDYDFDIAVLRKTPITFERMNVAPACLPERDWAESTLMTQKTGI
 VSGFGRTHEKGRQSTRKMLVPEYVDRNSCKLSSSFIITQNMFCAGYDTKQEDACQGDGSG
 GPHVTRFKDITYFTVTGIVSWGEGCARKGYGIYTKVTAFLKWIDRSNMKTRGLPKAKSHAPE
 VITSSPIK

SEQ.ID.NO. 69 HG1018333P1 NP_000495:NM_000504_1-23
 MGRPLHLVLLSASLAGLLLGES

SEQ.ID.NO. 70 HG1018334P1 NP_000495:NM_000504_1-19
 MGRPLHLVLLSASLAGLLI

SEQ.ID.NO. 71 HG1018335P1 NP_000495:NM_000504_1-20
 MGRPLHLVLLSASLAGLLI

SEQ.ID.NO. 72 HG1018336P1 NP_000495:NM_000504_1-15
 MGRPLHLVLLSASLA

SEQ.ID.NO. 73 HG1018337P1 NP_000495:NM_000504_1-21
 MGRPLHLVLLSASLAGLLIIG

SEQ.ID.NO. 74 HG1018338P1 NP_000495:NM_000504_1-17
 MGRPLHLVLLSASLAGL

SEQ.ID.NO. 84 HG1018348P1 NP_000591:NM_000600_1-24
MNSFSTSAFGPVAFSLGLLLVLP

SEQ. ID. NO. 85 HG1018349P1 NP_000591:NM_000600_1-27
MNSFSAFGPVAFSGLLLVLPAAFP

SEQ. ID. NO. 86 HG1018350P1 NP_000598:NM_000607
MALSIVLTSLPLLEAQIPLCANLVPVITNATLDQITGKWFYIASAFNEEYKSVQ
EIQATFFYTPNKTEDTIFREYQTRQDQCIYNTYINVRQENGTSRYVGGQEHFAHLL
ILRDTKYMLAFDVNDEKNWGLSVYADKPETTKQLGEFYEALDCLRIPKSDVVYTDWKK
DKCEPLEKQHEKEREKQEEGES

SEQ. ID. NO. 87 HG1018351P1 NP_000598:NM_000607_1-18
MALSIVLTSLPLLEA

SEQ. ID. NO. 88 HG1018352P1 NP_000604:NM_000613
MARVLGAPVALGLWSLCSLAIATPLPPTSAHGNVAEGETKPDVTERCSDGWSFDATT
LDDNGTMLFFKGEFVWKSXKWDRELISERWKNFSPVDAAFRQGHNSVFLIKGDKVWVYP
PEKKEGYPKLLQDEFFGIPSPDLAAVECHRGECQAEGVLPFGQDREWFWDLATCTMKER
SWPAVGNCSALRWLGRYYCFQGNQFLRFDPPVRGEVPPRYPRDVRDYFMPGCRGHGRN
GTGHGNSTHHGPEYMRCSPLVLSALTSNDHGATYAFSGTHYWRDLTSRDGWSWPIAHQ
WPQGPSAVDAAFSWEKLYLVQGTQVYVFLTKGGYTLVSGYPKRLEKEVGTPHGILLDSV
DAAFICPGSSRIHIMAGRRLLWLDLKSQAQTWTELPWPHKVDGALCMEKSLGNSCSA
NGPGLYLHGNPLYCYSVDVEKLNAAKALPQPQNVTSLLGCTH

SEQ. ID. NO. 89 HG1018353P1 NP_000604:NM_000613_1-19
MARVLGAPVALGLWSLCS

SEQ. ID. NO. 90 HG1018354P1 NP_000604:NM_000613_1-25
MARVLGAPVALGLWSLCSLAIATP

SEQ. ID. NO. 91 HG1018355P1 NP_000604:NM_000613_1-21
MARVLGAPVALGLWSLCSLAI

SEQ. ID. NO. 92 HG1018356P1 NP_000604:NM_000613_1-23
MARVLGAPVALGLWSLCSLAI

SEQ. ID. NO. 93 HG1018357P1 NP_000604:NM_000613_1-31
MARVLGAPVALGLWSLCSLAIATPLPPTSA

SEQ. ID. NO. 94 HG1018358P1 NP_000726:NM_000735
MDYYRKYAAIFLVTLSVFLVHLSAPDVQDCPECTIQENPFPSQPGAPILQCMGCCFSRA
YPTFLRSKKTMLVQKNVTSESTCCVAKSYNRVTVMGGFKVENHTACHCSTCYHKS

SEQ. ID. NO. 95 HG1018359P1 NP_000726:NM_000735_1-26
MDYYRKYAAIFLVTLSVFLVHLSAP

SEQ.ID.NO. 96 HG1018360P1 NP_000726:NM_000735_1-24
MDYRKYAAIFLVTLVFLHVLHS

SEQ.ID.NO. 97 HG1018361P1 NP_000884:NM_000893
MKLITILFLCSRLLLSLTQESQSEEDCNDKDLFKAVDAALKKYNQSNQNFVLYRIT
EATKTVGSDTFYFKEIKEGDCPVQSGKTWQDCEYKDAKAATGECTATVGRSSTKFS
VATQTCITPAEGPVVTAQYDCLGCVHPISTQSPDLEPILRHGIQYFNNNTQHSLSFMLN
EVKRAQRQVAGLNFRTYSIVQTNCSKENFLFTPDCKSLWNGDTGECTDNAYIDIQLR
IASFSQNCDIYPGKDFVQPPTKICVGCPRDIPNSPELEETLTHITIKLNAENNAFYFK
IDNVKKARQVVAGKKYFIDFVARETTCSKESNEELTESCETKLGQSLDCNAEVVVPW
EKKIYTVNCQPLGMISLMKRPPGFPFRSSRIGEIKEETTSHLRSCYKGRPPKAGABP
ASEREVS

SEQ.ID.NO. 98 HG1018362P1 NP_000884:NM_000893_1-18
MKLITILFLCSRLLLSLT

SEQ.ID.NO. 99 HG1018363P1 NP_000884:NM_000893_1-19
MKLITILFLCSRLLLSLTQ

SEQ.ID.NO. 100 HG1018364P1 NP_000884:NM_000893_1-16
MKLITILFLCSRLLLS

SEQ.ID.NO. 101 HG1018365P1 NP_000884:NM_000893_1-23
MKLITILFLCSRLLLSLTQESQS

SEQ.ID.NO. 102 HG1018366P1 NP_000909:NM_000918
MLRRALLCLAAALVRADAPEDHVLVLRKSNFAEALAAHKYLLVEFYAPWCCHCKALA
PEYAKAAGKKAEGSEIRLAKVDATEESDLAQYGVRGYPTIKFFRNGDTASPKYTAGR
EADDIVNWLKRTGPAATLPDGAASLSESVAVIGFFKDVESDSAKQFLQAAEAD
DIPFGITNSDVFYSKYQLDKGVVLFKKFDEGRNFEGETKENLDFIKHNQLPLVIEF
TEQTAPKIFGGEIKTHILFLPKSVSDYDGLSNFKTAAESFKGKILFIFIDSHTDNQR
ILEFFGLKKECPAVRLITLEEEMTKYKPESEELTAERITEFCHRFLEGIKPHLMSQEL
PEDWDQPVKVLVGKNFEDVAFDEKKNVFEFYAPWCCHCKQLAPIWIDKLGETYKDHENI
VIAKMDSTANEVEAVKVSFPTLKFFPASADRTVIDYNGERTLDGFKFLESQGDGAGD
DDDLDEEAEEPMEEDDDQKAVKDEL

SEQ.ID.NO. 103 HG1018367P1 NP_000909:NM_000918_1-17
MLRRALLCLAAALVRA

SEQ.ID.NO. 104 HG1018368P1 NP_000930:NM_000939
MPRSCCRSGALLLALLLQASMEVRCWGLESSQCQDLTTESNLEECIRACKPDLSAETPM
FPGNGDEQPLTENPRKYVMGHFRWDRFGRNRSSSSGSGAGQKREDVSAGEDCGPLPEGG
PEPRSDGAKPGPREGKRSYSMEHFRWGKPVGKKRRPVKYPNGAEDESAEAPLEFKREL

TGQRLREGDPDGDGAGAAQADLEHSLVAAEKKDEGYPYRMEHFRWGSPPKDKRYGGF
MTSEKSQTPLVTLFKNAIIKNAYKKG

SEQ.ID.NO. 105 HG1018369P1 NP_000930:NM_000939_1-23
MPRSCSRSGALLALLQASME

SEQ.ID.NO. 106 HG1018370P1 NP_000930:NM_000939_1-26
MPRSCSRSGALLALLQASMEVRG

SEQ.ID.NO. 107 HG1018371P1 NP_000945:NM_000954
MATHHTLWMGLALLGVLGDLQAAPEAQSVQPNFQDKFLGRWFSAGLASNSSWLREKKA
ALSMCKSVVAPATDGGNLITSTFLRKNQCETRTMLLQAPAGSLGSYRSRPHWGSTYSVSV
VETDYDQYALLYSQSGKPGEDFRMATLYSRTQTTPRAELKEKFTAFCKAQGFTEDTIVFL
PQTDKCMTEQ

SEQ.ID.NO. 108 HG1018372P1 NP_000945:NM_000954_1-23
MATHHTLWMGLALLGVLGDLQAA

SEQ.ID.NO. 109 HG1018373P1 NP_000945:NM_000954_1-22
MATHHTLWMGLALLGVLGDLQAA

SEQ.ID.NO. 110 HG1018374P1 NP_000945:NM_000954_1-18
MATHHTLWMGLALLGVLG

SEQ.ID.NO. 111 HG1018375P1 NP_001176:NM_001185
MVRMVPVLLSLLLLGPAVPQENQDGRYSLTFTYGLSKHVEDVPAPQALGSLNDLQFFR
YNSKDRKSQPMGLWRQVEGMEDWKQDSQLQKAREDIFMETLKDIVEYNDNSNGSHVLQGR
FGCEIENNRSSGAFWKVYDGDYIEFNKEIPAWVPFDPAAQITKQWEAEPVYVQRAKA
YLEECPATLRKYLKYSKNILDRQDPPSVVTSHQAPGEKKLKLAYDFYPGKIDVHWT
RAGEVQEPELRGDLVHNGNGTYQSWVVAVPPQDTAPYSCHVQHSLSLAQPLVVPWEAS

SEQ.ID.NO. 112 HG1018376P1 NP_001176:NM_001185_1-18
MVRMVPVLLSLLLLGPA

SEQ.ID.NO. 113 HG1018377P1 NP_001176:NM_001185_1-20
MVRMVPVLLSLLLLGPAVP

SEQ.ID.NO. 114 HG1018378P1 NP_001176:NM_001185_1-21
MVRMVPVLLSLLLLGPAVPQ

SEQ.ID.NO. 115 HG1018379P1 NP_001176:NM_001185_1-17
MVRMVPVLLSLLLLGP

SEQ.ID.NO. 116 HG1018380P1 NP_001266:NM_001275

MRSAAVLALLLCAGQVTALPVNSPMNKDTEVMKCIVEVISDTLSKPSMPVSVQECFETL
 RGDIERILSILRHQNLKELQDLALQAKERAHQKKHSGFEDELSEVLENOSSQAEELKEA
 VEEPSKDVMEKREDSKEAKSGEATDGARQALPEPMQESKAEGNNQAPGEEEEEEEA
 TINTHPASLPQKYPQAEGLDSEGLSQGLVDREKGLSAEPGWAKRKEEEEEEEAEAG
 REAVPEEGPTVVNLPHPSLGYKEIRKGSERSEALAVDAGKPGAEEAQDPGKGEQEH
 QQKEEEEMAVVPGLFRGKSGELEQEEERLSKEWEDSKRWSKMDQLAKELTAEKRLEG
 QEEEDNRDSSMKLSFRARAYGFRGPGQLRRGRWRPSSREDSLEAGLPLQVRGYPEEKKE
 EEGSANRRPEDQEELESLSAIEAELEKVAHQALRRG

SEQ. ID. NO. 117 HG1018381P1 NP_001266:NM_001275_1-18
 MRSAAVLALLLCAGQVTA

SEQ. ID. NO. 118 HG1018382P1 NP_001266:NM_001275_1-15
 MRSAAVLALLLCAGQ

SEQ. ID. NO. 119 HG1018383P1 NP_001266:NM_001275_1-14
 MRSAAVLALLLCAG

SEQ. ID. NO. 120 HG1018384P1 NP_001314:NM_001323
 MARSNPLAIGLALVAFCLALPRDARARQERMVGEIRDLSFDDPQVQKAAQAAVASYN
 MGSNSIYFRDTHIIKAQSQLVAGIKYFLTMEMGSTDCRKTTRVTGDHVDLTTCPLAAGAQ
 QEKLRCDFEVLVVPWQNSSQLLKHNCVQM

SEQ. ID. NO. 121 HG1018385P1 NP_001314:NM_001323_1-26
 MARSNPLAIGLALVAFCLALPRDA

SEQ. ID. NO. 122 HG1018386P1 NP_001314:NM_001323_1-18
 MARSNPLAIGLALVAFCL

SEQ. ID. NO. 123 HG1018387P1 NP_001314:NM_001323_1-20
 MARSNPLAIGLALVAFCLL

SEQ. ID. NO. 124 HG1018388P1 NP_001314:NM_001323_1-28
 MARSNPLAIGLALVAFCLALPRDARA

SEQ. ID. NO. 125 HG1018389P1 NP_001314:NM_001323_1-21
 MARSNPLAIGLALVAFCLLA

SEQ. ID. NO. 126 HG1018390P1 NP_001314:NM_001323_1-23
 MARSNPLAIGLALVAFCLIALP

SEQ. ID. NO. 127 HG1018391P1 NP_001822:NM_001831
 MMKTLILLFVGLLITWESGVLGDQTVSDNELQEMSNQGSKYVNKEIQNAVNGVKQIKTLI
 EKTNEERKTLILSLEAKKKEDALNETRESETKLKELPGVCNETMMALWEECKPCLKQT

CMKFYARVCRSGSLVGRQLEEFNQSSPFYFMWNGDRIDSLLENDROQQTTHMLDVMQDHF
SRASSIIDELFQDRFFTRPQDTYHYLPFSLPHRRPHFFPKSRIVRSMLMPFSPEPLNF
HAMFQPFLEMIHEAQAAMD IHFSPAQHPPTTEFIREGDDDRVTVCREIRHNSTGCLRMKD
QCDKCREILSVDCSTNNPSQAKLRDELDESQVAERLTRKYNELLKSYQWKMLNTSSLLE
QLNEQFNWVSRLANLTQGEDQYILRVTVASHTSDSDVPSGVTEVVVKLFDSDPITVTVP
VEVSRKNPKFMTVAEKALQEYRKKHREE

SEQ. ID. NO. 128 HG1018392P1 NP_001822:NM_001831_1-22
MMKTLILLFVGLLLTWESGVLG

SEQ. ID. NO. 129 HG1018393P1 NP_001822:NM_001831_1-18
MMKTLILLFVGLLLTWESG

SEQ. ID. NO. 130 HG1018394P1 NP_001822:NM_001831_1-14
MMKTLILLFVGLLLT

SEQ. ID. NO. 131 HG1018395P1 NP_002206:NM_002215
MDGAMGPRGLLLCMYLVSLILQAMPALGSATGRSKSSEKQAVDTAVDGVFIRSLKVC
KVTSRFAHYVVTQVNTANEAREVAFLPTAFISDFAVTADGNAFIGDIKOKVTAW
KQYRKAASIGENAGLVRASGRTEQFTIHLTVNPQSKVTQFLTYEEVLKRNMQYEIVIK
VKPKQLVHHFEIDVFEPQGISKLDAQASFLPKELAAQTIKKSGKKGHVLFRTVSQ
QQSCTCSTLLNGHFVTVTRDEICDLLVANNHFAFFAPQNLNMMKNVVFVIDIS
GSMRGQVKQTKEALKILGDMQPDYFDLVLFGTRVQSWKGSVLVQASEANLQAAQDFVR
GFSIDEATNLNGGLRGTEILNQVESLPELSNHASILIMLTDGDPTEGVTDRSQILKNV
RNAIRGRFPLYNLGFHNVDFNFLEVMSENNGRAQRIYEDHDATQQLQGFSYQVAKPLL
VDVLOQYQDQAVLALTQNHKKQYEGSEIVVAGRIADNKQSSFADVDQAHGEGQEFSTC
LVDEEMKKLLRERGHMLENHVERLWAYLTIQELLAKRMKVDREVRANLSSQALRMSLDY
GFVTPLTSMISIRGMADQGLKPTIDKPSSESPPLEMLGPRRTFVLSALQPSPTHSSNTQ
RLPDRVTGVDTPHFIHVPQKEDTLCFNINEEPGVILSLVQDPNTGFSVNGQLIGNKAR
SPGQHDGTYFGRGLIANPATDFQLEVTPQNTILNPGFGPVFSWRDQAVLRQDGVVVTIN
KKRNLVSVDDGGTFEVVLRVWKGSSVHQDFLGVYLD SHRMSARTHGLLGQFFHPPIGF
EVSDIHPGSDPTKPDATMVVRNRRLTVTRGLQKDYSKDPWHGAEVSCWFTHNNGAGLIDG
AYTDYIVPDI

SEQ. ID. NO. 132 HG1018396P1 NP_002206:NM_002215_1-24
MDGAMGPRGLLLCMYLVSLILQ

SEQ. ID. NO. 133 HG1018397P1 NP_002206:NM_002215_1-29
MDGAMGPRGLLLCMYLVSLILQAMPALG

SEQ. ID. NO. 134 HG1018398P1 NP_002206:NM_002215_1-30
MDGAMGPRGLLLCMYLVSLILQAMPALGS

SEQ. ID. NO. 135 HG1018399P1 NP_002206:NM_002215_1-23

MDGAMGPRGLLLCMYLVSLILQ

SEQ.ID.NO. 136 HG1018400P1 NP_002206:NM_002215_1-31
MDGAMGPRGLLLCMYLVSLILQAMPALGSA

SEQ.ID.NO. 137 HG1018401P1 NP_002300:NM_002309
MKVLAAGVVPVLLVHLHWKHGAGSPLPITPVNATCAIRHPCHNNLMNQIRSQLAQLNGSAN
ALFIIYYTAQGEPPFNIDKLCGPNVTDFFPHANGTEKAKVELYRIVVYLGTSIGNIT
RDQKILNPALSLSHSLNATADILRGLLSNVLCRLCSKYHVGHDVTYGPDTSKDVFK
KKLGCCQLLGKYKQIIAVLAQAF

SEQ.ID.NO. 138 HG1018402P1 NP_002300:NM_002309_1-22
MKVLAAGVVPVLLVHLHWKHGAG

SEQ.ID.NO. 139 HG1018403P1 NP_002300:NM_002309_1-23
MKVLAAGVVPVLLVHLHWKHGAGS

SEQ.ID.NO. 140 HG1018404P1 NP_002336:NM_002345
MSLSAFTLFLALIGTSGQYYDYDFPLSIYQSSPNCAPESYPSAMYCDELKLS
VPMVPPGIKYLVRNNQIDHIDEKAFENVTDLQWLILDHNLLENSKIKGRVFSKLKQLKK
LHINHNLTESVGPLPKSLEDLQLTNKTITKLGSEGLVNLTFIHLQHNRLKEDAVSAAF
KGLKSLEYLDLSFNQIARLPGLPVSLTYLDNNKISNIPDEYFRFNALQYLRLSHNE
LADSGIPGNSFNVSIVELDLSYNKLNPTVNNENLENYLEVNLQLEKFDIKSFCKILGP
LSYSKIKHLRLDGNRISETSLPPDMYECLRVANEVTLN

SEQ.ID.NO. 141 HG1018405P1 NP_002336:NM_002345_1-18
MSLSAFTLFLALIGTSG

SEQ.ID.NO. 142 HG1018406P1 NP_002336:NM_002345_1-15
MSLSAFTLFLALIG

SEQ.ID.NO. 143 HG1018407P1 NP_002336:NM_002345_1-17
MSLSAFTLFLALIGTIS

SEQ.ID.NO. 144 HG1018408P1 NP_002336:NM_002345_1-14
MSLSAFTLFLALIG

SEQ.ID.NO. 145 HG1018409P1 NP_002402:NM_002411
MKLLMVLMLAALSQHCYAGSGCPLEENVISKTINPQVSKTEYKELLQEFIDDNATTNAID
ELKECFLNQTDETLSNVVFQMQLIYDSSLCDLF

SEQ.ID.NO. 146 HG1018410P1 NP_002402:NM_002411_1-18
MKLLMVLMLAALSQHCYA

SEQ.ID.NO. 147 HG1018411P1 NP_002505:NM_002514
MQSVQSTSFCLRKQCLCLTFLLLHLLGVAATQCPQPCRPATPPTCAPGVRAVLGDG
CSCCLVCAQRGESCSDLEPCDESSGLYCDRSADPSNQTGICTAVEGDNCFVDGVIYRSG
EKFPSCKFTCRDGIQCVPRCQLDVLPEPNCAPARKVEVPGECEKWCIGPDEEDS
LGGLTLAAYRPBATICGVEVSDSSVNCIEQTTEWTACSKSCGMGFSTRVTNRNRQCEMLKQ
TRLCMVRPCEQEPQPTDKGKKCLRTKSLKAIHLQFNCTSLHTYKPRFCGVCSDGRC
CTPHNTKTIQAEFQCSPGQIVKKPVMTGCTCTCHTNCPKNNEAFLQELELKTTRGKM

SEQ.ID.NO. 148 HG1018412P1 NP_002505:NM_002514_1-30
MQSVQSTSFCLRKQCLCLTFLLLHLLGQVA

SEQ.ID.NO. 149 HG1018413P1 NP_002505:NM_002514_1-32
MQSVQSTSFCLRKQCLCLTFLLLHLLGQVAAT

SEQ.ID.NO. 150 HG1018414P1 NP_002505:NM_002514_1-28
MQSVQSTSFCLRKQCLCLTFLLLHLLGQ

SEQ.ID.NO. 151 HG1018415P1 NP_002505:NM_002514_1-27
MQSVQSTSFCLRKQCLCLTFLLLHLLG

SEQ.ID.NO. 152 HG1018416P1 NP_002505:NM_002514_1-31
MQSVQSTSFCLRKQCLCLTFLLLHLLGQVAA

SEQ.ID.NO. 153 HG1018417P1 NP_002892:NM_002901
MARGGRRLGLALGLLALVAPRVLRAPTVRKERVVRPDSSELGERPPEDNQSQYDH
EAPLGKEDSKTFDQLTDPDESKERLGIIVDRIDNDGDGFTTEELKTIWKRVQKRYIFDNV
AKWKDYDRDKDKISWEYKQATYGYLGNPAEFHDSSDHTFKMLPRDERRFKADL
NGDLTATREEFTAFHPPEEFHMKIIVVLETFLEDIDKNGDGVQDEYIADMFSEENG
EPDWVLSEREQFNEFRDLNKDGLDKDEIRHWILPDYDHAQAEARHLVYESDKNKDEKL
TKEEILENWNMFVGSQATNYGEDITKNHDEL

SEQ.ID.NO. 154 HG1018418P1 NP_002892:NM_002901_1-26
MARGGRRLGLALGLLALVAPRV

SEQ.ID.NO. 155 HG1018419P1 NP_002892:NM_002901_1-22
MARGGRRLGLALGLLALV

SEQ.ID.NO. 156 HG1018420P1 NP_002892:NM_002901_1-29
MARGGRRLGLALGLLALVAPRVLA

SEQ.ID.NO. 157 HG1018421P1 NP_002892:NM_002901_1-24
MARGGRRLGLALGLLALVAP

SEQ.ID.NO. 158 HG1018422P1 NP_002892:NM_002901_1-23

MARGGRRLGLALGLLLALVLA

SEQ.ID.NO. 159 HG1018423P1 NP_002893:NM_002902
 MRLGPRTAALGILLCAAGAGAKAEHYPGLGERRSDYDRREALLGVEQEDVEXYKLGHE
 EQQRRLAIKKIDLDSDGFLTESELSWIQMSFKHYAMQAKQKFVEYDKNSDDTVTWD
 EYNIQMYDRVIDFDENTALDDAEESFRKHLKDKRFEKANQDSGPGLSLEEFIAFEHP
 EEVDYMTFVIQEALEEHDKNGDFVSLEEFGLGDRWDPTANEDPEWILVEKDRFVNDYD
 KNDGRLDPQELLPWVPNNQIAQEEALHLIDEMDLNGDKKLSEEEILENPDPLTSEA
 TDYGRQLHDDYFYHDEL

SEQ.ID.NO. 160 HG1018424P1 NP_002893:NM_002902_1-25
 MRLGPRTAALGILLCAAGAGAKA

SEQ.ID.NO. 161 HG1018425P1 NP_002893:NM_002902_1-19
 MRLGPRTAALGILLCAAA

SEQ.ID.NO. 162 HG1018426P1 NP_002893:NM_002902_1-22
 MRLGPRTAALGILLCAAGAGA

SEQ.ID.NO. 163 HG1018427P1 NP_002893:NM_002902_1-18
 MRLGPRTAALGILLCAAA

SEQ.ID.NO. 164 HG1018428P1 NP_002893:NM_002902_1-20
 MRLGPRTAALGILLCAAAA

SEQ.ID.NO. 165 HG1018429P1 NP_002893:NM_002902_1-21
 MRLGPRTAALGILLCAAG

SEQ.ID.NO. 166 HG1018430P1 NP_002893:NM_002902_1-23
 MRLGPRTAALGILLCAAGAG

SEQ.ID.NO. 167 HG1018431P1 NP_005133:NM_005142
 MAWFALYLLSLWATAGTSTQSSCSVPQAEPVNGIQLMENSVTSSAYPNPSILIA
 MNLGAYNLKAQKLLTYQLMSSDNDLTIGHLGLTITMALTSSCRDPGDKVSIQRQMENW
 APSSPNARASAFYGPSSAILALCQKNSEATLPIAVRFAKTLLANSSPFNVDTGAMATIAL
 TCWYKIPVGBEGYRSLFGQVKDIVEKISMKIDNGIIGDIYSTGLAMQALSVTPEPS
 KKEWNCCKTTDMILNEIKQGFHNPMIAQILPSLKGKTYLDVPQVTCSPDHEVQPTLPS
 NPGPGPTSASNTVIYTIINNQLRGVELLFNETINVSVSGSVLLVLEEAQRKNPMFKFE
 TTTMTSWGLVVSSINNIAENVNHKTYWQFLSGVTPLNEGVDYIPFNHEHITANFTQY

SEQ.ID.NO. 168 HG1018432P1 NP_005133:NM_005142_1-19
 MAWFALYLLSLWATAGTS

SEQ.ID.NO. 169 HG1018433P1 NP_005133:NM_005142_1-18

MAWFALYLLSLLWATAGT

SEQ.ID.NO. 170 HG1018434P1 NP_005133:NM_005142_1-20
MAWFALYLLSLLWATAGTST

SEQ.ID.NO. 171 HG1018435P1 NP_005133:NM_005142_1-24
MAWFALYLLSLLWATAGTSTQTQS

SEQ.ID.NO. 172 HG1018436P1 NP_005133:NM_005142_1-16
MAWFALYLLSLLWATA

SEQ.ID.NO. 173 HG1018437P1 NP_005133:NM_005142_1-17
MAWFALYLLSLLWATAG

SEQ.ID.NO. 174 HG1018438P1 NP_005133:NM_005142_1-14
MAWFALYLLSLLWA

SEQ.ID.NO. 175 HG1018439P1 NP_005445:NM_005454
MHLLLFQLLVLLPLGKTRHQDGRQNOSSLSPLPRNQRELPFGNHEEAEEKPDLFVAV
PHLVATSPAGEQQRQREKMLSRFRWKKPEREMHPSRDSSEPPPGTQSLIQPIDGMK
MEKSPLEAKKFWHHFMFRKTPASQGVILPIKSHVHWETCTVPPFSQTITHEGCEKVV
VQNNLCFGKCGSVHFPGAQAQHSHTSCSHCLPAKFTTMHLPLNCTELSSVIKVVMLVEECQ
CKVTEHEDGHILHAGSQDSFIPGUSA

SEQ.ID.NO. 176 HG1018440P1 NP_005445:NM_005454_1-17
MHLLLFQLLVLLPLGKT

SEQ.ID.NO. 177 HG1018441P1 NP_005555:NM_005564
MPLGLLWLGALLGALHAQAQDSTDLIPAPPLSKVPLQQNFQDNQFQGWYVVGAGNA
ILREDKDPQMYATIYELKEDKSYNVTSLFRKKCDYWIRTFVGCQPGFTLGNIKSY
PGLTSYIVRVVSTNYNQHAMVFFKKVSNREYFKITLYGRTKELTSELKENFIRFSKYL
LPENHIVFPVPIDQCIDG

SEQ.ID.NO. 178 HG1018442P1 NP_005555:NM_005564_1-18
MPLGLLWLGALLGALHA

SEQ.ID.NO. 179 HG1018443P1 NP_005555:NM_005564_1-20
MPLGLLWLGALLGALHAQA

SEQ.ID.NO. 180 HG1018444P1 NP_005555:NM_005564_1-15
MPLGLLWLGALLGA

SEQ.ID.NO. 181 HG1018445P1 NP_005690:NM_005699
MRHNWTPDLSPLWVLLCAHVVTLLVRATPVSTTTAATASVRSTKDCPSQPPVFPAAK

QCPALEVTWPEVEVPLNGTSLSCVACSRFPNFSILYWLNGSGFIEHLPRLWEGSTSR
RGSTGTQLCKALVLEQLTPALHSTNFCVLVDEQVVQRHVVLVLAQLWVRSPPRGLQEQQE
LCFHMWGGKGGLCQSSL

SEQ. ID. NO. 182 HG1018446P1 NP_005690:NM_005699_1-29
MRHNWTPDLSPLWVLLCAHVVTLLVRAT

SEQ. ID. NO. 183 HG1018447P1 NP_005690:NM_005699_1-24
MRHNWTPDLSPLWVLLCAHVVTLL

SEQ. ID. NO. 184 HG1018448P1 NP_005690:NM_005699_1-28
MRHNWTPDLSPLWVLLCAHVVTLLVRA

SEQ. ID. NO. 185 HG1018449P1 NP_006560:NM_006569
MLPLTMTVLIILLPTGQAAPKGVTRPDSEVQHQLLPNPFQPGQEQGLLQSYLKGLGR
TEVQLEHLSREQVLLYLFALHDYDQSGQLDGLLELLSMLTAALAPGAANSPTNPVILI
VDKVLQDGLNGDGLMTFAELINFPQVALRHVEPEPLAPSPQEQAVGRQSLLAKSPLRQE
TQEAPGPREAKGQVEARRESLDPVQEPGQAEADGDPVGPGEAGQAEAKGDAPGPRG
EAEGQAEAKGDAPGPRGEAGQAEARENGEEAKELPGETLESKNTQNDFEVHIVQVENE
I

SEQ. ID. NO. 186 HG1018450P1 NP_006560:NM_006569_1-19
MLPLTMTVLIILLPTGQA

SEQ. ID. NO. 187 HG1018451P1 NP_006560:NM_006569_1-18
MLPLTMTVLIILLPTGQ

SEQ. ID. NO. 188 HG1018452P1 NP_006560:NM_006569_1-21
MLPLTMTVLIILLPTGQAAP

SEQ. ID. NO. 189 HG1018453P1 NP_006856:NM_006865
MTSILTVLILCLGLSDPRTHVQAGPLPKPTLWAEPPGSVITQGSPTLRCQGSLETQYHL
YREKKTALWITRIPQELVKKQFPILSITWEHAGRYCCYGSHTAGLSSESDPLELVVTG
AYSKPTLSALPSPVVTSGGNVTIQCDSQVAFDGFILCKEGEDEHPQCINSHSHARGSSRA
IFSVPVSPSRRSVRCYDSDRAPYVWSLPSDLGLLVPGVSKPSLSVQPGPVVAPGE
KLTFQCGSDAGYDRFVLYKEWRDFLQRPGRQPQAGLSQANFTLGPVRSYGGQYTCGGA
YNLSSEWSAPSDPLDILITGQIRARPFLSVREGPTVAGENVTLLCQSQGGMHTFLLTKE
GAADSPRLKSKRQSHKYQAEFPMSPTSAHACTYRCYGLSSNPYLLTHPSDPLELVVS
GAAETLSPPNKSDSKAGE

SEQ. ID. NO. 190 HG1018454P1 NP_006856:NM_006865_1-15
MTSILTVLILCLGLSL

SEQ. ID. NO. 191 HG1018455P1 NP_036577:NM_012445

MENPSPAALGKALCALLATLGAACQPLGGESIC SARAPAKYSITFTGKWSQTAFPPKQY
 PLFRPPAQWSSLLGAHSSDYSMMRNQYVNSGLRDFAEERGEAWALMKEIEAAGEALQSV
 HAVFSAPAVPSGTGQTSAELEVRHSLVSVFVRIVPSDPWFVGVDSLDLDCDGRWREQA
 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHFANSFYPRLKALPPIARVT
 LVRLRQSPRAFI PPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRIGTKS
 RTRYVRVQPANNGSPCELEEEAECPDNCV

SEQ.ID.NO. 192 HG1018456P1 NP_036577:NM_012445_1-26
 MENPSPAALGKALCALLATLGAAG

SEQ.ID.NO. 193 HG1018457P1 NP_036577:NM_012445_1-25
 MENPSPAALGKALCALLATLGAAG

SEQ.ID.NO. 194 HG1018458P1 NP_036577:NM_012445_1-24
 MENPSPAALGKALCALLATLGA

SEQ.ID.NO. 195 HG1018459P1 NP_036577:NM_012445_1-28
 MENPSPAALGKALCALLATLGAAGQP

SEQ.ID.NO. 196 HG1018460P1 NP_055070:NM_014255
 MKGWGLALLGALTAWARRSQDLHCGACRALVDELEWEIAQVDPKTIQMGSRINP
 DGSQSVVEVPYARSEAHITELLEICDRMKEYGEQIDPSTHRKNVVRVVRNGESSELDL
 QGIRIDSDISGTLKFACESIVEEYEDELIEFFSREADNVKDKLCSKRFDLCDHALHISHD
 EL

SEQ.ID.NO. 197 HG1018461P1 NP_055070:NM_014255_1-20
 MKGWGLALLGALLGTAWA

SEQ.ID.NO. 198 HG1018462P1 NP_055070:NM_014255_1-18
 MKGWGLALLGALLGTA

SEQ.ID.NO. 199 HG1018463P1 NP_055070:NM_014255_1-16
 MKGWGLALLGALLG

SEQ.ID.NO. 200 HG1018464P1 NP_055582:NM_014767
 MRAPGCGRLVPLLLLLAAALAEADAKGLKEGETPGNFMEDQWLSSISQYSGKIKHWN
 FRDEVEDDYIKSWEDNQDEALDITKDPCCQVKCSRHKVCIAQGYQRAMCISRKKLEHR
 IKQPTVKLHGNKDSICKPCHMAQLASVCGSDGHTYSSVCKLEQQACLSSKQLAVRCEGPC
 PCPTQAATSTADGKPTCTGQDLADLGDRLRDWFOILLHENSQNGSSASSVAGPASGLDK
 SLGASCKDSIGWMFSKLDTSADLFDQTELAAILNDKYEVCI RPPFFNSCDTYKDG RVSTA
 EWCFCFWREKPPCLAELERIQIQEAAKKKPGIFIPSCDEDDGYRKMQCQSSGDCWCVDQ
 LGLELTGTFTHGPSDCDDIVGFSGDFGSGVGEDEEEKETEAGEAEAEAGEADDG
 GYIW

SEQ.ID.NO. 201 HG1018465P1 NP_055582:NM_014767_1-24
MRAPGCGRLVLP1LL1AAAALAEQ

SEQ.ID.NO. 202 HG1018466P1 NP_055582:NM_014767_1-19
MRAPGCGRLVLP1LL1AAA

SEQ.ID.NO. 203 HG1018467P1 NP_055582:NM_014767_1-22
MRAPGCGRLVLP1LL1AAAALA

SEQ.ID.NO. 204 HG1018468P1 NP_055582:NM_014767_1-20
MRAPGCGRLVLP1LL1AAAA

SEQ.ID.NO. 205 HG1018469P1 NP_055582:NM_014767_1-26
MRAPGCGRLVLP1LL1AAAALAEQDA

SEQ.ID.NO. 206 HG1018470P1 NP_055582:NM_014767_1-21
MRAPGCGRLVLP1LL1AAAAL

SEQ.ID.NO. 207 HG1018471P1 NP_055697:NM_014882
MSLGQSACFLSIARSRSVMTGEQMAAFPSSTPNPLERPIKMWLKKQRSIVKNWQORY
FVLRAQQLYYKDEEDTKPGCMYLPQCTIKEIATNPEEAGKFVEIIPASWDQNRMGQD
SYVLMASSQAEEMEWKFLRRVAGTPCGVFGQRLDETVAEQKFGPHLVPILVEKCAEFI
LEHGRNEEGIFRLPGQDNLVKQLRDAFDAGERPSFDRDTHVTVASLLKLYLRDLPEPVV
PWSQYEGFLLCGLTNADAQELKQSLPRDNYSLLSYICRFLHEIQLNCAVNM
SVDNLATVIGVNLIRSKVEDPAVIMRGTPQIQRVMTMMIRDHEVLPFKSKDIPLSPPAQK
NDPKKAPVARSSVGDATEDLIRISRTDSFSSMTSDSTTSPTGQPSDAFEDSSKVPRE
KPGDWKMQSRKRTQTLPNRKCFLTSAFQGANSSKWEIFKNEFWSSEAKAGEGHRRTMS
QDLRQLSDSQRTSTVDNVPSLPGSPGEEASALSSQACDSKGDTLASPNSETGPGKNSGE
BEIDSLQRMVQELRKEIETQKMYEEQIKNLEKENYDVWAKVVRNLEELEKEKKKSAALE
ISLRNMERSREDVEKRNKALEEEVKEFVKSMKEPKTEA

SEQ.ID.NO. 208 HG1018472P1 NP_055697:NM_014882_1-18
MSLGQSACFLSIARSRS

SEQ.ID.NO. 209 HG1018473P1 NP_056965:NM_015881
MQRLGATLLCLLLAAAVTAPAPATATSAVPKPGPALSYPOEEATINEMFREVEELMED
TQHKLRSAVEEMEAEEAAAKASSEVNLANLPPSYHNETHNTDQKGNNTIHHVREIHKITN
NQTCQMVFSETVITSVGDEEGRRSHECIIDECCGFSMYCQFASFYTCQPCRGMCTR
DSECCGDLQCVWGHCTKMATRSGNGTICDNQDCQPGCLCCAFQGLFPVCTPLPVEGEL
CHDPASRLDLITWELEPDGALDRCPCASGLLCQPHSHSLVYVCKPTFVGSRDQDGEILL
PREVPDEYEVGSFMEEVRQLEDLERSLITEEMALGEPAAAAALLGGEHI

SEQ.ID.NO. 210 HG1018474P1 NP_056965:NM_015881_1-18
MQRLGATLLCLLLAAAVP

SEQ.ID.NO. 211 HG1018475P1 NP_056965:NM_015881_1-19
MORLGATLLCLLLAAAVPT

SEQ.ID.NO. 212 HG1018476P1 NP_056965:NM_015881_1-22
MORLGATLLCLLLAAAVPTAPA

SEQ.ID.NO. 213 HG1018477P1 NP_056965:NM_015881_1-16
MORLGATLLCLLLAAA

SEQ.ID.NO. 214 HG1018478P1 NP_056965:NM_015881_1-21
MORLGATLLCLLLAAAVPTAP

SEQ.ID.NO. 215 HG1018479P1 NP_057603:NM_016519
MSASKIPLFKMKDLILILCLLEMSFAVFFPQQSGTPGMASLSLETMRQLGSLQRLNTLS
QYSRYGFGKSFNSLWMHGLPPHSLPWMRPREHETQOYEYSLPVHPPPLPSQPSLKPPQ
PGLKPFLOSAAATNQATALKEALQPPIHGLHPLQEGELPLVQQQVAPSDKPKPELPG
VDFADPQPSLPGMDFPDQGPSLPLGLDFADPQGSTIFQIARLISHGPMQKQSFPLYPG
MLYVFFGANQINAPARLIGIMSSEEVAGGREDDPMAYGAMFPGFGMRPFGEGMPHNPAMGG
DFTLEFDSVAATKGPENEEGACQSPMPEANPDNLENPAFLTELEPAPHAGLLALPKDD
IPGLPRSPSGMKGLPSVTAAADPLMTPELADVITYADMTTSVDFQEEATMDTTMAP
NSLQTSMPGNKAQEPENMHDAWHFQEP

SEQ.ID.NO. 216 HG1018480P1 NP_057603:NM_016519_1-26
MSASKIPLFKMKDLILILCLLEMSFA

SEQ.ID.NO. 217 HG1018481P1 NP_057603:NM_016519_1-28
MSASKIPLFKMKDLILILCLLEMSFAVP

SEQ.ID.NO. 218 HG1018482P1 NP_149439:NM_033183
MEMFQGLLLLLLSMGGTWSKEPLRPRCRPINATLAVEKEGCPVCITVNTTICAGYCPT
MTRVLQGVLPALPQVVCNYRDRVFESIRLPGCPRGVNPVSYAVALSCQCALCRSTTDC
GGPKOHPLTCDPRFQDSSSKAPPSLPSPSRLPGPSDTFILPQ

SEQ.ID.NO. 219 HG1018483P1 NP_149439:NM_033183_1-18
MEMFQGLLLLLLSMGGT

SEQ.ID.NO. 220 HG1018484P1 NP_149439:NM_033183_1-20
MEMFQGLLLLLLSMGGTWA

SEQ.ID.NO. 221 HG1018485P1 NP_149439:NM_033183_1-16
MEMFQGLLLLLLSMG

SEQ.ID.NO. 222 HG1018486P1 NP_644808:NM_139279

MTMRSLLRTPFLCGLLWAFCAPGARAEEPAASFQPSGMLDKNTVHDQEHIMEHLEGVI
NKPEAEMSPOELQLHYFKMHDYDGNLLDGLSTALTTHVHKEEGSEQAPLMSEDELINI
IDGVLRDDDKNDGYIDYAEFAKSLQ

SEQ.ID.NO. 223 HG1018487P1 NP_644808:NM_139279_1-18
MTMRSLLRTPFLCGLLWA

SEQ.ID.NO. 224 HG1018488P1 NP_644808:NM_139279_1-20
MTMRSLLRTPFLCGLLWAF

SEQ.ID.NO. 225 HG1018489P1 NP_644808:NM_139279_1-26
MTMRSLLRTPFLCGLLWAFCAPGARA

SEQ.ID.NO. 226 HG1018490P1 NP_644808:NM_139279_1-23
MTMRSLLRTPFLCGLLWAFCAPG

SEQ.ID.NO. 227 HG1018491P1 NP_660295:NM_145252
MLLLLTLLALGGPTWAGMYGPGGKYFSTTEDYDHEITGLRVSVGLLLVKSQVVKLGDS
WDVKLGALCNTQEVTLQPGEYITKVFAFAQLRGMVMTSKDRYFYFGKLDGQISSAY
PSQEGQVLGYGYQYQLLGIKSIGFEWNYPLEEPTTEPPVNLTYSANSPVGR

SEQ.ID.NO. 228 HG1018492P1 NP_660295:NM_145252_1-13
MLLLLTLLALGGP

SEQ.ID.NO. 229 HG1018493P1 NP_660295:NM_145252_1-16
MLLLLTLLALGGPTWA

SEQ.ID.NO. 230 HG1018494P1 NP_660295:NM_145252_1-14
MLLLLTLLALGGPT

SEQ.ID.NO. 231 HG1018495P1 NP_660295:NM_145252_1-17
MLLLLTLLALGGPTWAG

SEQ.ID.NO. 232 HG1018496P1 NP_689534:NM_152321
MEAAPSRFMFLFLTCELAEEVAEVEKSSDGPAAQEPWLTDPVPAAMEFIAATEVAV
IGFFQDLEIPAVFILHSMVQKFPVSGISTDSEVLTHYNTGNTICLRLVDNEQLNLE
DEDIESIDATKLSRFIEINSLHMVTEYNPVTIVIGLFSVIQIHLILIMNKASPEYENMH
RYQKAQKLFQGKILFILVDGSMKENGKVISFFKLKESQLPALAIYQTLDDDEWDTLPTAEV
SVEHVQNFCDGFLSGKLLKENRESEGKTPKVEL

SEQ.ID.NO. 233 HG1018497P1 NP_689534:NM_152321_1-25
MEAAPSRFMFLFLTCELAEEVAA

SEQ.ID.NO. 234 HG1018498P1 NP_689534:NM_152321_1-21

MEAAPSRFMFLFLLTCELAA

SEQ. ID. NO. 235 HG1018499P1 NP_689848:NM_152635
MPPFLLLTCLFITGTSVPVALDPCSAIISLNEPWRNTDQHLDSEQGPPLCDNHVNGEWY
HFTGMAGDAMPTFCIPENHGGTHAPVWLNGSHFLIEGDGIVORQACASNGNCLWNTTIVE
VKACPGGYVYRLTKPSCFHYVCGHFYDICEDECHGSCSDTSECTCAPGTVLGPDROTIC
FDENECEQNNNGCSEICVNLKNSYRCECGVGRVLRSDGKTCEDEVEGCHNNNGGCSHSCLG
SEKGYQCECPRLGLVSEDNHTCQVPVLCKSNALVNIIPRELVGGLLEFLTNTSCRGVSNNG
THVNILFSLKTCGTVDVNDKIVASNLVTGLPKQTPGSSGDFIIRTSKLLIPVTCEFPFR
LYTISEGYVPNLNSPLFMSRNLHGIFPFTLEIFKDNFEPEEPYREALPTLKLKRDLSLYFGI
BPVVHVSGLSESVESCFATPTSKIDEVLKYLLRPDGCVDSDSVKQYTSRDHLAKHFQVPV
PKFVGDHKEVFLHCRVLVCGVLDESRCAQCGCHRRMRRGAGGEDSAGLQGGTTLTGPIR
IDWED

SEQ. ID. NO. 236 HG1018500P1 NP_689848:NM_152635_1-18
MPPFLLLTCLFITGTSVS

SEQ. ID. NO. 237 HG1018501P1 NP_689848:NM_152635_1-16
MPPFLLLTCLFITGTS

SEQ. ID. NO. 238 HG1018502P1 NP_689848:NM_152635_1-15
MPPFLLLTCLFITGT

SEQ. ID. NO. 239 HG1018503P1 NP_689968:NM_152755
MGPVRLGILLFLFLAVHEAWGLKEEDDDTERLPKCEVCKLLSTELQAELSRTGRSRE
VLELQGVLDTGKRKRHPYSVSETRLEALENLCEILIDYSVHAERKGSILRYAKGQSQTM
ATLKLGVQKGVKVDLGLPLELWDEPSVEVTYLLKQCEMTLEEFEDIVGDWYFHHQEQPLQ
NFLCEGHVLPAAETACLQETWTGKEITDGEKTEGEEEEEEEEEEEEEGGDKMTKTGSH
PKLDREDL

SEQ. ID. NO. 240 HG1018504P1 NP_689968:NM_152755_1-21
MGPVRLGILLFLFLAVHEAWA

SEQ. ID. NO. 241 HG1018505P1 NP_766630:NM_173042
MRHNWTPDLSPLWVLLCAHVVTLLVRATPVSQTTAATASVRSTKDCPCSPQPPVFPAAK
QCPALEVTWPEVEVPLNGTILSLSCVACSRFPNFSILYWLNGSFIHLPRLWEGSTSR
RGSTGTQLCKALVLEQLTPALHSTNFSCVLVDPEQVVQRHVVLVLAQLWAGLRATLPTQEA
LPSSHSPQQQG

SEQ. ID. NO. 242 HG1018506P1 NP_766630:NM_173042_1-29
MRHNWTPDLSPLWVLLCAHVVTLLVRAT

SEQ. ID. NO. 243 HG1018507P1 NP_766630:NM_173042_1-24
MRHNWTPDLSPLWVLLCAHVVTLL

SEQ. ID. NO. 244 HG1018508P1 NP_766630:NM_173042_1-28
MRHNWTPDLSPLWVLLCAHVVTLLVRA

SEQ. ID. NO. 245 HG1018509P1 NP_776214:NM_173842
MEICRLRSHLITLTLFLFHSETICRPSGRKSKMQAFRIWDVNQKTFYLRNNQLVAGYL
QGNVNLEEKIDVVPFIEHALFLGIHGGKMLSCVKSQDETRLQLEAVNITDLSENKQD
KRFAPIRSDSGPTTSFESACPGWFLCTAMEADQPVSLTNMPDEGVMVTKFYFQEDE

SEQ. ID. NO. 246 HG1018510P1 NP_776214:NM_173842_1-23
MEICRLRSHLITLTLFLFHSET

SEQ. ID. NO. 247 HG1018511P1 NP_776214:NM_173842_1-25
MEICRLRSHLITLTLFLFHSETIC

SEQ. ID. NO. 248 HG1018512P1 NP_783165:NM_175575
MWAPRCRRFWSRWEQVAALLLLLLLVGPPRSIALPPIRYSHAGICPNDMNPNLWVDAQS
TCRRECTQECETYEKCCPNVCGTKSCVAARYMDVKGKGPVGMPEKATCDHFMCLOQG
SECDIWDGQPVCKCKDRCEKPSFTCASDGLTYYNRCYMDAEACSKGITLAVVTCRYHFT
WPNTSPPPETTMHPTTASPETPELDMAAPALLNNPVHQSVTMGETVSFLCDVVGRPRPE
ITWEKQLEDRENVVMRPNHVRGNVVVNTIAQLVIYNAQLQDAGIYCTCTARNVAGVLRADF
PLSVVRGHQAATSESSPNGTAFPAECLKPPDSEDCGEEQTERWHFDAQANNCLTFTFGH
CHRNLFHETFEACMLACMSGPLAACLALQGPCKAYAPRAYNSQTGCQCFYVGCE
GNGNFESEACEESCFFPRGNQRCRACKPRQKLVTSFCRSDFVILGRVSELTEEPDSGR
ALVTVDEVLDKEMGLKFLGQEPLEVTLLHVDWACPCPNVTVSEMPLIIMGEVDGGWAML
RPDSFVGASSARRVRKIREVMHKKTCDVLKEFLGLH

SEQ. ID. NO. 249 HG1018513P1 NP_783165:NM_175575_1-32
MWAPRCRRFWSRWEQVAALLLLLLLVGPPRS

SEQ. ID. NO. 250 HG1018514P1 NP_783165:NM_175575_1-34
MWAPRCRRFWSRWEQVAALLLLLLLVGPPRSIA

SEQ. ID. NO. 251 HG1018515P1 NP_783165:NM_175575_1-29
MWAPRCRRFWSRWEQVAALLLLLLLVGVP

SEQ. ID. NO. 252 HG1018516P1 NP_783165:NM_175575_1-30
MWAPRCRRFWSRWEQVAALLLLLLLVGPP

SEQ. ID. NO. 253 HG1018517P1 NP_783165:NM_175575_1-27
MWAPRCRRFWSRWEQVAALLLLLLLG

SEQ. ID. NO. 254 HG1018856P1 27482680:27482679
MWCASPVAVAVAFCAGLLVSHPVLTQGEAGGRPGADCEVCKEFLNRFYKSLIDRGVNFSL
DTIEKELISFCLDTKGKENRLCYLGGATKDAATKIISEVTRPMSVHMPAMKICEKLLKKLD
SQICELKYEKTLDLASVDLFRMRVAELKQIILHSWGEECRACAECTDYVNLIQELAPKYAA
THPKTEL

SEQ. ID. NO. 255 HG1018857P1 27482680:27482679_1-26
MWCASPVAVAVAFCAGLLVSHPVLTQG

SEQ. ID. NO. 256 HG1018858P1 27482680:27482679_1-24
MWCASPVAVAVAFCAGLLVSHPVLT