

**(12) STANDARD PATENT**  
**(19) AUSTRALIAN PATENT OFFICE**

(11) Application No. **AU 2011248489 B2**

(54) Title  
**Innovative discovery of therapeutic, diagnostic, and antibody compositions related to protein fragments of alanyl tRNA synthetases**

(51) International Patent Classification(s)  
**C12N 9/00** (2006.01) **C12N 15/113** (2010.01)  
**A61K 38/43** (2006.01) **C12N 15/52** (2006.01)  
**C07K 16/40** (2006.01) **G01N 33/68** (2006.01)  
**C07K 19/00** (2006.01)

(21) Application No: **2011248489** (22) Date of Filing: **2011.04.28**

(87) WIPO No: **WO11/139853**

(30) Priority Data

(31) Number	(32) Date	(33) Country
<b>61/329,048</b>	<b>2010.04.28</b>	<b>US</b>
<b>61/328,867</b>	<b>2010.04.28</b>	<b>US</b>
<b>61/328,871</b>	<b>2010.04.28</b>	<b>US</b>

(43) Publication Date: **2011.11.10**

(44) Accepted Journal Date: **2016.10.06**

(71) Applicant(s)  
**aTYR Pharma, Inc.;Pangu Biopharma Limited**

(72) Inventor(s)  
**Greene, Leslie Ann;Chiang, Kyle P.;Hong, Fei;Vasserot, Alain P.;Lo, Wing-Sze;Watkins, Jeffry D.;Mendlein, John D.;Quinn, Cheryl L.**

(74) Agent / Attorney  
**Davies Collison Cave Pty Ltd, Level 15 1 Nicholson Street, MELBOURNE, VIC, 3000**

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

(19) World Intellectual Property Organization  
International Bureau

(43) International Publication Date  
10 November 2011 (10.11.2011)



(10) International Publication Number  
**WO 2011/139853 A3**

(51) International Patent Classification:

*C12N 9/00* (2006.01) *C12N 15/113* (2010.01)  
*C12N 15/52* (2006.01) *C07K 16/40* (2006.01)  
*C07K 19/00* (2006.01) *A61K 38/43* (2006.01)  
*G01N 33/68* (2006.01)

(21) International Application Number:

PCT/US2011/034387

(22) International Filing Date:

28 April 2011 (28.04.2011)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/329,048	28 April 2010 (28.04.2010)	US
61/328,867	28 April 2010 (28.04.2010)	US
61/328,871	28 April 2010 (28.04.2010)	US

(71) Applicants (for all designated States except US): **ATYR PHARMA, INC.** [US/US]; 3565 General Atomics Court, Suite 103, San Diego, California 92121 (US). **PANGU BIOPHARMA LIMITED** [CN/CN]; 18/F Edinburgh Tower, The Landmark, 15 Queen's Road Central, Hong Kong (CN).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **GREENE, Leslie Ann** [US/US]; 13941 Davenport Avenue, San Diego, California 92129 (US). **CHIANG, Kyle, P.** [US/US]; 943 Torrance Street, Apt. 6, San Diego, California 92103 (US). **HONG, Fei** [CN/US]; 7979 Caminito Dia, #2, San Diego, California 92122 (US). **VASSEROT, Alain, P.** [CH/US]; 6421 La Paloma Street, Carlsbad, California 92009 (US). **LO, Wing-Sze** [CN/CN]; RM 815, Sui Ming House, Siu Sai Wan Estate, Chai Wan, Hong Kong (CN). **WATKINS, Jeffry, D.** [US/US]; 3442 Fortuna Ranch Road, Encinitas, California 92024 (US). **MENDLEIN,**

**John, D.** [US/US]; 1550 Neptune Avenue, Encinitas, California 92024 (US). **QUINN, Cheryl, L.** [US/US]; 5392 Renaissance Avenue, San Diego, California 92122 (US).

(74) Agents: **ROGEL, Mark, E.** et al.; Colley LLP, 777 6th Street NW, Suite 1100, Washington DC 20001 (US).

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

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

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

- with international search report (Art. 21(3))

[Continued on next page]

(54) Title: INNOVATIVE DISCOVERY OF THERAPEUTIC, DIAGNOSTIC, AND ANTIBODY COMPOSITIONS RELATED TO PROTEIN FRAGMENTS OF ALANYL TRNA SYNTHETASES

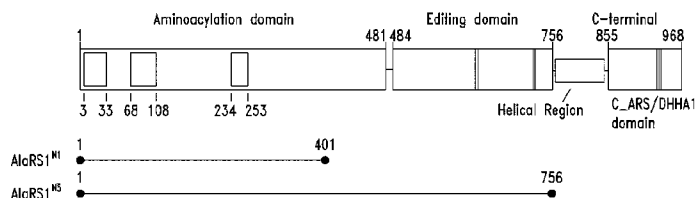


FIG. 1A

(57) Abstract: Provided are compositions comprising newly identified protein fragments of aminoacyl-tRNA synthetases, polynucleotides that encode them and complements thereof, related agents, and methods of use thereof in diagnostic, drug discovery, research, and therapeutic applications.



- 
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
- *with sequence listing part of description (Rule 5.2(a))*
- (88) Date of publication of the international search report:**  
10 May 2012

**INNOVATIVE DISCOVERY OF THERAPEUTIC, DIAGNOSTIC, AND ANTIBODY  
COMPOSITIONS RELATED TO PROTEIN FRAGMENTS OF ALANYL tRNA SYNTHETASES**

**CROSS-REFERENCE TO RELATED APPLICATIONS**

[0001] This application claims the benefit under 35 U.S.C. § 119(e) of US provisional patent application No. 61/329,048 filed on 4/28/2010, US provisional patent application No. 61/328,867 filed 4/28/2010, and US provisional patent application No. 61/328,871 filed 4/28/2010, the entire contents of each of which, are incorporated herein by reference.

**STATEMENT REGARDING SEQUENCE LISTING**

[0002] The Sequence Listing associated with this application is provided in text format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The name of the text file containing the Sequence Listing is 120161\_426PC\_SEQUENCE\_LISTING\_txt. The text file is about 281 KB, was created on 4/22/2011, and is being submitted electronically via EFS-Web.

**TECHNICAL FIELD**

[0003] The present invention relates generally to compositions comprising newly identified protein fragments of aminoacyl-tRNA synthetases and other proteins, polynucleotides that encode them and complements thereof, related agents, and methods of use thereof in diagnostic, drug discovery, research, and therapeutic applications.

**BACKGROUND**

[0004] For over four decades, aminoacyl-tRNA synthetases (AARSs) were thought of as essential housekeeping proteins that catalyze the aminoacylation of tRNA molecules as part of the decoding of genetic information during the process of protein translation. AARSs have been extensively studied in this respect, and many of their full-length sequences were cloned for sequence analysis and to provide a rich source of biochemical experimentation. Some fragments of AARSs, and other proteins, however, possess unexpected activities not associated with aminoacylation, including extracellular signaling activities that modulate pathways beyond protein translation. Generally, these unexpected activities are not observed in the context of the full-length or parental protein sequences; instead, they are observed following removal or resection of AARS protein fragments from their parental sequences, or by expressing and



sufficiently purifying fragment AARS sequences and then testing for novel, non-synthetase related activities.

**[0005]** While the full-length sequences of AARS have been known for some time, no systematic experimental analysis has been conducted to elucidate such AARS protein fragments, or protein fragments from related or associated proteins, or to evaluate the potential role of the full length AARS proteins for novel biological activities outside of the context of amino acid synthesis. .” In portions of this specification, such AARS protein fragments, AARS domains, or AARS alternative splice variants are referred to herein as “resectins”. In its broadest context, the term “resectin” refers to a portion of a protein which has been excised or restricted (either by means of proteolysis, alternative splicing, mutagenesis, or recombinant genetic engineering) from the context of its native full-length or parental protein sequence, which often otherwise masks its novel biological activities. Likewise, no systematic experimental analysis has been conducted to explore the use of such resectins as biotherapeutic agents, diagnostic agents, or drug targets in the treatment of various medical conditions, or their potential association with human diseases. As essential housekeeping genes with a known function in mammals that is critical to life, AARSs were neither considered as drug targets in mammals, nor were they parsed out by standard genomic sequencing, bioinformatic, or similar efforts to identify resectins having non-synthetase activities. Standard biochemical research efforts have similarly been directed away from characterizing the biological properties of AARS resectins and their potential therapeutic and diagnostic relevance, mainly due to the previously understood role of their corresponding full-length parental AARSs.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

**[0006]** **Figure 1** shows the domain structure of the Alanyl aminoacyl tRNA synthetase overlaid with the relative positions and sizes of the N-terminal AARS polypeptides shown schematically. Figure 1A representing fragments identified from mass spectrometry analysis, Figure 1B representing the fragments identified from deep sequencing of transcriptomes, and Figure 1C representing fragments identified from bioinformatics analysis.

**[0007]** **Figure 2** shows the domain structure of the Alanyl aminoacyl tRNA synthetase overlaid with the relative positions and sizes of the C-terminal AARS polypeptides shown schematically. Figure 2A representing fragments identified from mass spectrometry analysis, Figure 2B representing the fragments identified from deep sequencing of transcriptomes, and Figure 2C representing fragments identified from bioinformatics analysis.

[0008] **Figure 3** shows the domain structure of the Alanyl aminoacyl tRNA synthetase overlaid schematically with the relative positions and sizes of the Internal AARS polypeptides identified from bioinformatics analysis.

#### **BRIEF SUMMARY OF THE INVENTION**

[0009] Embodiments of the present invention relate generally to the discovery of protein fragments of aminoacyl-tRNA synthetases (AARSs), which possess non-canonical biological activities, such as extracellular signaling activities, and/or other characteristics of therapeutic and diagnostic relevance. The AARSs are universal and essential elements of the protein synthesis machinery found in all organisms, but human AARSs and their associated proteins have naturally-occurring resected variants, with potent cell signaling activities that contribute to normal functioning of humans. The activities of these protein fragments are distinct from the protein synthesis activities commonly known for AARSs, and the present invention includes the discovery and development of these resected proteins as new biotherapeutic agents, new discovery research reagents, and as new antigens/targets for directed biologics and diagnostic agents that can be used to potentially treat or diagnose a wide variety of human diseases, such as inflammatory, hematological, neurodegenerative, autoimmune, hematopoietic, cardiovascular, and metabolic diseases or disorders.

[0010] The AARS protein fragment(s) of the present invention may therefore be referred to as “resectins,” or alternatively as “appendacrines.” As noted above, the term “resectin” derives from the process of excising or resecting a given AARS protein fragment from the context of its full-length parent AARS sequence, which typically masks its non-canonical activities. In certain instances, the AARS protein fragments and polynucleotides of the present invention were identified through the occurrence of this resection process, whether naturally-occurring (*e.g.*, proteolytic, splice variant), artificially-induced, or predicted. The term “appendacrine” derives from a combination of “append” (from Latin – appender) and to “separate” or “discern” (from Greek – crines),” and also reflects the separation of one or more appended domains of the AARS protein fragments from their corresponding full-length or parent AARS sequences.

[0011] Although a few AARS fragments have been previously shown to have non-synthetase activities, the expression, isolation, purification, and characterization of such fragments for biotherapeutic, discovery, or diagnostic utility is limited, and persons skilled in the art would not have readily appreciated such activities to associate with each member of the entire family of AARSs, or with alternative fragments. Here, a methodical approach was utilized to discover and verify AARS protein fragments for

the 20 mitochondrial and 20 cytosolic AARSs (and associated proteins) for biotherapeutic discovery and diagnostic utility. For instance, certain of the present AARS protein fragment(s) and polynucleotides that encode them are identified from biological samples using mass spectrometry (MS), mainly to identify proteolytic fragments, and others were identified by deep sequencing techniques, mainly to identify splice variants. Other AARS protein fragment(s) are identified using *in silico* predictions of amino acid sequences, such as by computationally comparing synthetases from humans and lower organisms along with key demarcations (*e.g.*, protease sites); this approach utilized sequence analysis of the full-length AARS based on specific criteria to discern proteolytic fragments and functional domains possessing non-canonical biological activities.

[0012] Novel resectins of the AARSs are unexpected, and their differential expression is also unexpected. Specific resections are typically seen under different treatments (*e.g.*, from cells grown in media with or without serum), at different stages of growth (*e.g.*, adult brain vs. fetal brain) and for different tissue types (*e.g.*, pancreas vs. liver). The pattern of expression is not the same for all aminoacyl tRNA synthetases despite the fact that the canonical functions for all aminoacyl tRNA synthetases are needed in the same cell locations and in relatively proportional amounts. One would not expect the levels of an aminoacyl tRNA synthetase activity to increase without an increase in the amounts of other aminoacyl tRNA synthetase activities at the same time. The mass spectrometry and deep sequencing data indicates that aminoacyl tRNA synthetase resectins do have varying levels and do occur in different sites and at different stages

[0013] In addition, AARS protein fragments can be expressed and purified to sufficiently high purity to discern their biological properties. Previously, fragments were often not of sufficient purity, folding, and stability to enable proper biological characterization of non-synthetase activities. Cell based assays, for instance, are used in conjunction with sufficiently pure, stable, soluble and folded resectins to reveal their important biotherapeutic, discovery or diagnostic activities.

[0014] In particular, embodiments of the present invention relate to protein fragments of Alanine tRNA synthetases, related agents and compositions of biotherapeutic, discovery, or diagnostic utility, and methods of use thereof. The compositions of the present invention are useful in a variety of diagnostic, drug discovery, and therapeutic applications, as described herein. Preferably, the AARS proteins and fragments are purified and stored in suitable condition to the extent required for such biotherapeutic, discovery, or diagnostic uses.

**[0015]** Certain embodiments include compositions, comprising an isolated aminoacyl-tRNA synthetase (AARS) protein fragment of at least about 100, 90, 80, 70, 60, 50 or 40 amino acids that comprises an amino acid sequence as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, and has a solubility of at least about 5 mg/ml, and wherein the composition has a purity of at least about 95% on a protein basis, and less than about 10 EU / mg protein endotoxin. In one aspect, the composition is a therapeutic composition. In specific embodiments, the composition is substantially serum free. In some embodiments the AARS protein fragment comprises a non-canonical activity. In some embodiments, the non-canonical biological activity is selected from modulation of extracellular signaling, modulation of cell proliferation, modulation of cell differentiation, modulation of gene transcription, modulation of cytokine production or activity, modulation of cytokine receptor activity, and modulation of inflammation. In some embodiments, the AARS protein fragment has an EC<sub>50</sub> of less than about 1 nM, about 5 nM, about 10 nM, about 50 nM, about 100 nM or about 200 nM for a cell-based non-canonical biological activity.

**[0016]** In certain embodiments the AARS protein fragment is fused to a heterologous polypeptide. In some embodiments, the AARS fusion protein substantially retains a non-canonical activity of the AARS protein fragment. In some embodiments, the AARS fusion protein suppresses a non-canonical activity of the AARS protein fragment. In some embodiments, the heterologous polypeptide is attached to the N-terminus of the AARS protein fragment. In some embodiments, the heterologous polypeptide is attached to the C-terminus of the AARS protein fragment. In one aspect of any of these embodiments the heterologous polypeptide is selected from the group consisting of purification tags, epitope tags, targeting sequences, signal peptides, membrane translocating sequences, and PK modifiers.

**[0017]** In certain embodiments, the composition comprises an AARS protein fragment at a concentration of least about 10 mg/mL. In certain embodiments the composition comprises an AARS protein fragment which is at least 90% monodisperse. In certain embodiments the composition comprises less than about 3 % high molecular weight aggregated proteins. In certain embodiments the composition exhibits less than 3% aggregation when stored at a concentration of at least 10 mg/ mL in PBS for one week at 4 °C. In certain embodiments the composition exhibits less than 3% aggregation when stored at a concentration of at least 10 mg/ mL in PBS for one week at room temperature.

**[0018]** Various assays for measuring such features of resectins are described herein and may be used to define aspects of the invention. In certain aspects, these features

will be preferable for biotherapeutic utility of the AARS protein fragments described herein.

**[0019]** Certain embodiments include compositions, comprising an isolated aminoacyl-tRNA synthetase (AARS) protein fragment of at least 100 amino acids that differs from an amino acid sequence set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9 by substitution, deletion, and/or addition of about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids, wherein the altered protein fragment substantially retains a non-canonical activity of the unaltered protein, or has a dominant negative phenotype in relation to the non-canonical activity, wherein the protein fragment has a solubility of at least about 5 mg/ml, and wherein the composition has a purity of at least about 95% on a protein basis and less than about 10 EU / mg protein endotoxin. In specific embodiments, the composition is substantially serum free.

**[0020]** Other embodiments include compositions, comprising an isolated antibody that specifically binds to an isolated aminoacyl-tRNA synthetase (AARS) protein fragment as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, wherein affinity of the antibody for the AARS protein fragment is about 10X stronger than its affinity for a corresponding full-length AARS polypeptide. One of the surprising aspects of the present invention includes certain resectins possessing “new” surfaces accessible to antibody or other directed biologics, whereas the full length AARS “hides” or covers these surfaces with other sequences or adjacent domains. The process of resecting can also create greater aqueous accessibility for revealing previously unidentified biological activities. Some embodiments include compositions, comprising an isolated antibody that specifically binds to an isolated aminoacyl-tRNA synthetase (AARS) protein fragment as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, wherein the antibody has an affinity of at least about 10 nM for the AARS protein fragment, and an affinity of at least about 100 nM for a corresponding full-length AARS polypeptide. In some embodiments, the antibody binds to an epitope located within an AARS polypeptide unique splice junction as set forth in any of Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, or to an amino acid sequence C-terminal of this splice site. In certain embodiments, the antibody antagonizes the non-canonical activity of the AARS protein fragment. Such antagonists may optionally bind the corresponding parental or full-length AARS.

**[0021]** Other aspects relate to bioassay systems, comprising a substantially pure aminoacyl-tRNA synthetase (AARS) protein fragment of at least 100 amino acids that comprises an amino acid sequence as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, and a binding partner that binds to the AARS protein fragment. In one

aspect, the binding partner is selected from the group consisting of a cellular surface receptor protein, nucleic acid, lipid membrane, cell regulatory protein, enzyme, and transcription factor. Optionally, such a receptor may be part of a cell, preferably a cell relevant to the revealed biology of the resectin.

[0022] Certain embodiments include cellular compositions, comprising an isolated aminoacyl-tRNA synthetase (AARS) protein fragment of at least 100 amino acids that comprises an amino acid sequence as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, and an engineered population of cells in which at least one cell comprises a polynucleotide encoding said AARS protein fragment. In one aspect, the cells are capable of growing in a serum free medium.

[0023] Also included are detection systems, comprising a substantially pure aminoacyl-tRNA synthetase (AARS) protein fragment of at least 50 or 100 amino acids that comprises an amino acid sequence as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, a cell that comprises a cell-surface receptor or an extracellular portion thereof that binds to the protein fragment, and a molecule of less than about 2000 daltons, or a second polypeptide, which modulates binding or interaction between the AARS protein fragment and the extracellular receptor.

[0024] Particular embodiments include diagnostic systems, comprising a substantially pure aminoacyl-tRNA synthetase (AARS) protein fragment of at least 100 amino acids that comprises an amino acid sequence as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, and a cell that comprises a cell-surface receptor or an extracellular portion thereof that binds to the AARS protein fragment, wherein the system or cell comprises an indicator molecule that allows detection of a change in the levels or activity of the cell-surface receptor or extracellular portion thereof.

[0025] Certain embodiments include cellular growth devices, comprising an isolated aminoacyl-tRNA synthetase (AARS) protein fragment of at least 100 amino acids that comprises an amino acid sequence as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, an engineered population of cells in which at least one cell comprises a polynucleotide encoding said AARS protein fragment, at least about 10 liters of serum-free cell media, and a sterile container. In specific embodiments, the cells utilized for any of the methods or compositions described herein are capable of growing in serum-free media, optionally with an antibiotic and an inducer.

[0026] Some embodiments relate to antisense or RNA interference (RNAi) agents, comprising a sequence that is targeted against a unique splice junction of an AARS splice variant as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9.

[0027] Also included are therapeutic compositions, comprising an isolated aminoacyl-tRNA synthetase (AARS) protein fragment of at least 100 amino acids that

comprises an amino acid sequence as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, wherein the protein fragment specifically binds to a binding partner and has a solubility of at least about 5 mg/ml, and wherein the composition has a purity of at least about 95% on a protein basis. In some aspects, the composition may have less than 10 EU endotoxin / mg protein.

**[0028]** Also included are compositions, comprising an isolated aminoacyl-tRNA synthetase (AARS) protein fragment of at least 100 amino acids that is at least 80%, 85%, 90%, 95%, 98%, or 100% identical to an amino acid sequence set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, wherein the protein fragment has a solubility of at least about 5 mg/ml, and wherein the composition has a purity of at least about 95% on a protein basis and less than 10 EU endotoxin / mg protein. In any of these embodiments, the compositions may comprise an AARS protein fragment that is at least about 50%, about 60%, about 70%, about 80%, about 90% or about 95% monodisperse with respect to its apparent molecular mass. In another aspect of any of these embodiments, the compositions comprise less than about 10 % (on a protein basis) high molecular weight aggregated proteins, or less than about 5 % high molecular weight aggregated proteins, or less than about 4% high molecular weight aggregated proteins, or less than about 3% high molecular weight aggregated proteins, or less than about 2 % high molecular weight aggregated proteins, or less than about 1% high molecular weight aggregated proteins.

**[0029]** In another aspect of any of these embodiments, the compositions exhibits less than about 10% aggregation when stored at a concentration of at least 10 mg/ mL in PBS for one week at 4 °C, or less than about 5% aggregation when stored at a concentration of at least 10 mg/ mL in PBS for one week at 4 °C, or less than about 3% aggregation when stored at a concentration of at least 10 mg/ mL in PBS for one week at 4 °C, or less than about 2% aggregation when stored at a concentration of at least 10 mg/ mL in PBS for one week at 4 °C, or less than about 1% aggregation when stored at a concentration of at least 10 mg/ mL in PBS for one week at 4 °C.

**[0030]** Certain embodiments include compositions, comprising a substantially pure aminoacyl-tRNA synthetase (AARS) protein fragment of at least 100 amino acids that comprises an amino acid sequence as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, and at least one covalently or non-covalently moiety attached thereto. In some embodiments, the moiety is a detectable label. In some embodiments, the moiety is a water soluble polymer. In some embodiments, the moiety is PEG. In one aspect of any of these embodiments, the moiety is attached to the N-terminus of the protein fragment. In one aspect of any of these embodiments, the moiety is attached to the C-terminus of the protein fragment.

[0031] Particular embodiments include compositions, comprising a solid substrate attached to an isolated aminoacyl-tRNA synthetase (AARS) protein fragment of at least 100 amino acids that comprises an amino acid sequence as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, or a biologically active fragment or variant thereof, wherein the protein fragment has a solubility of at least about 5 mg/ml, and the composition has a purity of at least about 95% on a protein basis.

[0032] Also included are compositions, comprising a binding agent that specifically binds to an isolated aminoacyl-tRNA synthetase (AARS) protein fragment as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, wherein the binding agent has an affinity of at least about 1 nM for the protein fragment. In one aspect, the binding agent binds to an epitope located within an AARS polypeptide unique splice junction as set forth in any of Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, or to an amino acid sequence C-terminal of this splice site. In some embodiments, the binding agent antagonizes a non-canonical activity of the AARS polypeptide.

[0033] Certain embodiments include isolated aminoacyl-tRNA synthetase (AARS) polypeptides, comprising an amino acid sequence of an AARS protein fragment as described herein, an amino acid sequence encoded by an AARS polynucleotide as described herein, or a variant or fragment thereof. Certain AARS polypeptides comprise an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, or 100% identical to an AARS reference sequence as disclosed in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, or Table E2. Certain AARS polypeptides consist essentially of an amino acid sequence that is at least 80%, 85%, 90%, 95%, 98%, or 100% identical to an AARS reference sequence as disclosed in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, or Table E2. In certain embodiments, the polypeptide comprises a non-canonical biological activity. In specific embodiments, the non-canonical biological activity is selected from modulation of cell signaling (*e.g.*, extracellular signaling), modulation of cell proliferation, modulation of cell migration, modulation of cell differentiation, modulation of apoptosis or cell death, modulation of angiogenesis, modulation of cell binding, modulation of cellular metabolism, modulation of cellular uptake, modulation of gene transcription, or secretion, modulation of cytokine production or activity, modulation of cytokine receptor activity, and modulation of inflammation.

[0034] Other aspects include antibodies and other binding agents that exhibit binding specificity for an isolated AARS polypeptide as described herein, a binding partner of the AARS polypeptide, or the complex of both. In some embodiments, the affinity of the antibody or binding agent for the AARS polypeptide is about 10X stronger than its affinity for a corresponding full-length AARS polypeptide. In specific embodiments, the binding agent is selected from a peptide, peptide mimetic, an



adnectin, an aptamer, and a small molecule. In certain embodiments, the antibody or binding agent antagonizes a non-canonical activity of the AARS polypeptide. In other embodiments, the antibody or binding agent agonizes a non-canonical activity of the AARS polypeptide.

**[0035]** Certain embodiments include isolated aminoacyl-tRNA synthetase (AARS) polynucleotides, comprising a nucleotide sequence of an AARS polynucleotide as described herein, a nucleotide sequence that encodes an AARS protein fragment as described herein, or a variant, a fragment, or a complement thereof. Certain AARS polynucleotides comprise a nucleotide sequence that is at least 80%, 85%, 90%, 95%, 98%, or 100% identical to an AARS reference polynucleotide, or a complement thereof, as disclosed in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, or Table E2. In some embodiments, the nucleotide sequence is codon optimized for bacterial expression. In one aspect, the nucleotide sequence is at least 80% identical a polynucleotide sequence disclosed in Table E2.

**[0036]** Specific AARS polynucleotides consist essentially of a nucleotide sequence that is at least 80%, 85%, 90%, 95%, 98%, or 100% identical to an AARS reference polynucleotide, or a complement thereof, as disclosed in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, or Table E2. Other AARS polynucleotides comprise or consist essentially of a nucleotide sequence that specifically hybridizes to an AARS reference polynucleotide, as disclosed in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, or Table E2. In certain embodiments, the polynucleotide is selected from a primer, a probe, and an antisense oligonucleotide. In specific embodiments, the primer, probe, or antisense oligonucleotide is targeted to a specific or unique splice junction, and / or sequence 3' of this splice site within an AARS polynucleotide.

**[0037]** Certain embodiments include methods of determining presence or levels of an AARS protein fragment in a sample, comprising contacting the sample with one or more binding agents that specifically bind to an AARS protein fragment as described herein, detecting the presence or absence of the binding agent, and thereby determining the presence or levels of the AARS protein fragment. Other embodiments include methods of determining presence or levels of an AARS protein fragment in a sample, comprising analyzing the sample with a detector that is capable of specifically identifying a protein fragment as described herein, and thereby determining the presence or levels of the AARS protein fragment. In specific embodiments, the detector is a mass spectrometer (MS), a flow cytometer, a protein imaging device, an enzyme-linked immunosorbent assays (ELISA), or a protein microarray. Certain embodiments comprise comparing the presence or levels of the AARS protein fragment to a control sample or a predetermined value. Certain embodiments comprise

characterizing the state of the sample to distinguish it from the control. In specific embodiments, the sample and control comprise a cell or tissue, and the method comprises distinguishing between cells or tissues of different species, cells of different tissues or organs, cells at different cellular developmental states, cells at different cellular differentiation states, cells at different physiological states, or healthy and diseased cells. For instance, selected resectins may be more abundant under conditions such as stress or insult.

**[0038]** Certain embodiments include discovery methods of, and related compositions for, identifying a compound that specifically binds to an aminoacyl-tRNA synthetase (AARS) polypeptide as described herein, or one or more of its cellular binding partners, comprising a) combining the AARS polypeptide or its cellular binding partner or both with at least one test compound under suitable conditions, and b) detecting binding of the AARS polypeptide or its cellular binding partner or both to the test compound, thereby identifying a compound that specifically binds to the AARS polypeptide or its cellular binding partner or both. In certain embodiments, the test compound is a polypeptide or peptide, an antibody or antigen-binding fragment thereof, a peptide mimetic, or a small molecule. In certain embodiments, the test compound agonizes a non-canonical biological activity of the AARS polypeptide or its cellular binding partner. In other embodiments, the test compound antagonizes a non-canonical biological activity of the AARS polypeptide or its cellular binding partner. Certain embodiments include a compound identified by the above-method, such as an agonist (*e.g.*, small molecule, peptide).

**[0039]** Certain embodiments include methods of determining presence or levels of a polynucleotide sequence of an AARS splice variant in a sample, comprising contacting the sample with one or more oligonucleotides that specifically hybridize to an AARS polynucleotide as described herein, detecting the presence or absence of the oligonucleotides in the sample, and thereby determining the presence or levels of the polynucleotide sequence of the AARS splice variant. Other embodiments include methods of determining presence or levels of a polynucleotide sequence of an AARS splice variant in a sample, comprising contacting the sample with at least two oligonucleotides that specifically amplify an AARS polynucleotide as described herein, performing an amplification reaction, detecting the presence or absence of an amplified product, and thereby determining presence or levels of the polynucleotide sequence of the AARS splice variant. In specific embodiments, the oligonucleotide(s) specifically hybridize to or specifically amplify a splice junction that is unique to the AARS splice variant. Certain embodiments include comparing the presence or levels of the AARS protein fragment or splice variant to a control sample or a predetermined value. Certain

embodiments include characterizing the state of the sample to distinguish it from the control. In specific embodiments, the sample and control comprise a cell or tissue, and the method comprises distinguishing between cells or tissues of different species, cells of different tissues or organs, cells at different cellular developmental states, cells at different cellular differentiation states, or healthy and diseased cells.

**[0040]** Some embodiments include pharmaceutical compositions, comprising an AARS polynucleotide described herein, an AARS polypeptide described herein, a binding agent as described herein, or a compound identified by the above-method or described herein, and a pharmaceutically acceptable excipient or carrier.

**[0041]** Certain embodiments include methods of modulating a cellular activity of a cell, comprising contacting the cell with an AARS polynucleotide described herein, an AARS polypeptide described herein, a binding agent described herein, a compound of the above-method or described herein, or a pharmaceutical composition described herein. In specific embodiments, the cellular activity is selected from cell proliferation, cell migration, cell differentiation, apoptosis or cell death, cell signaling, angiogenesis, cell binding, cellular uptake, cell secretion, metabolism, cytokine production or activity, cytokine receptor activity, gene transcription, and inflammation. In one aspect, the cell is selected from the group consisting of pre-adipocytes, bone marrow, neutrophils, blood cells, hepatocytes, astrocytes, mesenchymal stem cells, and skeletal muscle cells.

**[0042]** In certain embodiments, the cell is in a subject. Certain embodiments comprise treating the subject, wherein the subject has a condition associated with a neoplastic disease, an immune system disease or condition, an infectious disease, a metabolic disease, an inflammatory disorder, neuronal/neurological disease, a muscular/cardiovascular disease, a disease associated with aberrant hematopoiesis, a disease associated with aberrant angiogenesis, or a disease associated with aberrant cell survival.

**[0043]** Also included are processes for manufacturing a pharmaceutical compound, comprising: a) performing an *in vitro* screen of one or more candidate compounds in the presence an AARS protein fragment of at least 100 amino acids that comprises an amino acid sequence as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, to identify a compound that specifically binds to the AARS protein fragment; b) performing a cell-based or biochemical or receptor assay with the compound identified in step a), to identify a compound that modulates one or more non-canonical activities of the AARS protein fragment; c) optionally assessing the structure-activity relationship (SAR) of the compound identified in step b), to correlate its structure with modulation of the non-canonical activity, and optionally derivatizing the compound to alter its ability to modulate the non-canonical activity; and d) producing sufficient amounts of

the compound identified in step b), or the derivatized compound in step c), for use in humans, thereby manufacturing the pharmaceutical compound.

**[0044]** Other embodiments include processes for manufacturing a pharmaceutical compound, comprising: a) performing an *in vitro* screen of one or more candidate compounds in the presence a cell-surface receptor or an extracellular portion thereof that specifically binds to an AARS protein fragment of Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, to identify a compound that specifically binds to the cell-surface receptor or extracellular portion thereof; b) performing a cell-based or biochemical or receptor assay with the compound identified in step a), to identify a compound that modulates one or more non-canonical activities of the AARS protein fragment; c) optionally assessing the structure-activity relationship (SAR) of the compound identified in step b), to correlate its structure with modulation of the non-canonical activity, and optionally derivatizing the compound to alter its ability to modulate the non-canonical activity; and d) producing sufficient amounts of the compound identified in step b), or the derivatized compound in step c), for use in humans, thereby manufacturing the pharmaceutical compound.

**[0045]** Some embodiments include a cellular composition, comprising an engineered population of cells in which at least one cell comprises a polynucleotide encoding a heterologous full length aminoacyl-tRNA synthetase (AARS) protein, wherein the cells are capable of growing in a serum-free medium. In one aspect, the full length aminoacyl-tRNA synthetase (AARS) protein comprises a heterologous purification or epitope tag to facilitate purification of an AARS protein fragment. In another aspect, the full length aminoacyl-tRNA synthetase (AARS) protein comprises a heterologous proteolysis site to enable production of the AARS protein fragment upon cleavage.

**[0046]** Some embodiments include a method for producing an AARS polypeptide as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, or Table E2 in situ within a cell, comprising; i) expressing a heterologous full length aminoacyl-tRNA synthetase (AARS) protein within the cell, wherein the cell comprises a protease capable of cleaving the heterologous full length aminoacyl-tRNA synthetase (AARS) protein to produce the AARS polypeptide.

**[0047]** Some embodiments include a method for producing an AARS polypeptide as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, or Table E2 comprising contacting an isolated full length aminoacyl-tRNA synthetase (AARS) protein with a protease that is capable of cleaving the full length aminoacyl-tRNA synthetase (AARS) protein and producing an AARS polypeptide.

2011248489 14 Sep 2016

**[0048]** Some embodiments include an engineered full length aminoacyl-tRNA synthetase (AARS) protein comprising a heterologous proteolysis site to enable the proteolytic generation of an AARS protein fragment as set forth in any of Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9 or Table E2.

**[0049]** Some embodiments include a composition, comprising an isolated full length aminoacyl-tRNA synthetase protein, wherein the composition has a purity of at least about 95% on a protein basis, less than about 10 EU endotoxin / mg protein, and is substantially serum free. In one aspect, the full length aminoacyl-tRNA synthetase protein is present at a concentration of at least 10 mg / mL, and is at least 90% monodisperse.

**[0050]** A further embodiment includes a method of treating a disease or disorder mediated by the dysregulation of the expression, activity or spatiotemporal location of a tRNA synthetase via the administration of an AARS protein fragment, or nucleic acid encoding the AARS protein fragment, as set forth in any of Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, or Table E2. In one aspect of this embodiment, the disease is selected cancer, neuropathy, diabetes, and inflammatory disorders.

**[0050a]** In an aspect, the present invention relates to a therapeutic composition, comprising an isolated aminoacyl-tRNA synthetase (AARS) polypeptide that is at least 95%, 98%, or 100% identical to SEQ ID NO: 81 or 142, or a fragment thereof which is 200 or more contiguous amino acids of SEQ ID NO: 81 or 142, wherein the polypeptide has an extracellular signaling activity and has a solubility of at least about 5 mg/mL, and wherein the composition has a purity of at least about 95% on a protein basis and less than about 10 EU endotoxin / mg protein.

**[0050b]** In another aspect, the present invention relates to a cellular composition or growth device, comprising an aminoacyl-tRNA synthetase (AARS) polypeptide that is at least 95%, 98%, or 100% identical to SEQ ID NO: 81 or 142, or a fragment thereof which is 200 or more contiguous amino acids of SEQ ID NO: 81 or 142, and an element selected from the group consisting of

- (i) an engineered population of cells in which at least one cell comprises a polynucleotide encoding said AARS polypeptide, wherein the cells are capable of growing in a serum-free medium, and
- (ii) an engineered population of cells in which at least one cell comprises a polynucleotide encoding said AARS polypeptide, at least about 10 liters of a

2011248489 14 Sep 2016

serum- free growth medium, and a sterile container.

**[0050c]** In another aspect, the present invention relates to a method of identifying a compound that specifically binds to an aminoacyl-tRNA synthetase (AARS) polypeptide that is at least 95%, 98%, or 100% identical to SEQ ID NO: 81 or 142, or a fragment thereof which is 200 or more contiguous amino acids of SEQ ID NO: 81 or 142, comprising

- a) combining the AARS polypeptide with at least one test compound under suitable conditions, and
- b) detecting binding of the AARS polypeptide to the test compound, thereby identifying a compound that specifically binds to the AARS polypeptide.

**[0050d]** In another aspect, the present invention relates to an engineered full length aminoacyl-tRNA synthetase (AARS) protein comprising a heterologous proteolysis site to enable the proteolytic generation of an AARS polypeptide that is at least 95%, 98%, or 100% identical to SEQ ID NO: 81 or 142, or a fragment thereof which is 200 or more contiguous amino acids of SEQ ID NO: 81 or 142.

**[0050e]** In another aspect, the present invention relates to a method for treating a neurological disease in a subject in need thereof comprising administering to the subject a therapeutic composition in accordance with the present invention.

**[0050f]** In another aspect, the present invention relates to a method for treating chronic granulomatous disease (CGD) or a condition where the oxidative burst of granulocytes is impaired in a subject in need thereof comprising administering to the subject an effective amount of a therapeutic composition in accordance with the present invention.

**[0050g]** In another aspect, the present invention relates to a method for treating an inflammatory disease in a subject in need thereof comprising administering to the subject an effective amount of a therapeutic composition in accordance with the present invention.

**[0050h]** The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that that prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavour to which this specification relates.

2011248489 14 Sep 2016

**DETAILED DESCRIPTION OF THE INVENTION****[0051] TABLE OF CONTENTS**

<b>[0052]</b>	I.	OVERVIEW .....	15
<b>[0053]</b>	II.	DEFINITIONS .....	15
<b>[0054]</b>	III.	PURIFIED AARS PROTEIN FRAGMENTS AND VARIANTS .....	28
<b>[0055]</b>	IV.	AARS POLYNUCLEOTIDES .....	87
<b>[0056]</b>	V.	ANTIBODIES .....	99
<b>[0057]</b>	VI.	ANTIBODY ALTERNATIVES AND OTHER BINDING AGENTS .....	104
<b>[0058]</b>	VII.	BIOASSAYS AND ANALYTICAL ASSAYS .....	109
<b>[0059]</b>	VIII.	EXPRESSION AND PURIFICATION SYSTEMS .....	111
<b>[0060]</b>	IX.	DIAGNOSTIC METHODS AND COMPOSITIONS .....	124
<b>[0061]</b>	X.	ANTISENSE AND RNAI AGENTS .....	139
<b>[0062]</b>	A.	ANTISENSE AGENTS .....	140
<b>[0063]</b>	B.	RNA INTERFERENCE AGENTS .....	148
<b>[0064]</b>	XI.	DRUG DISCOVERY .....	156
<b>[0065]</b>	XII.	METHODS OF USE .....	164
<b>[0066]</b>	XIII.	PHARMACEUTICAL FORMULATIONS, ADMINISTRATION AND KITS .....	168
<b>[0067]</b>	XIV.	EXAMPLES .....	177

## I. OVERVIEW

[0068] The current invention is directed, at least in part, to the discovery of novel AARS polypeptides, and methods for their preparation and use, that represent the transformation of native wild type proteins into new forms that exhibit markedly different characteristics compared to the naturally occurring full length Alanyl tRNA synthetase genes. Such AARS polypeptides were identified based on extensive sequence, and mass spectrum analysis of expressed Alanyl tRNA synthetase in different tissues, followed by the systematic production and testing of each potential AARS polypeptide to identify protein sequences that represent stable and soluble protein domains which exhibit novel biological activities.

[0069] Based on this analysis at least two novel families of AARS polypeptides derived from Alanyl tRNA synthetase have been identified.

[0070] In one aspect, such Alanyl tRNA synthetase derived AARS polypeptides comprise polypeptide sequences comprising approximately the first 400 to 530 amino acids of Alanyl tRNA synthetase.

[0071] In a second aspect, such Alanyl tRNA synthetase derived AARS polypeptides comprise polypeptide sequences comprising approximately the last 221 to 210 amino acids of Alanyl tRNA synthetase.

[0072] These new AARS polypeptide families represent novel, previously unknown protein products which exhibit *inter alia* i) novel biological activity, ii) favorable protein stability and aggregation characteristics, and iii) the ability to expressed and produced at high level in prokaryotic expression systems, which are materially different from the intact wild type protein.

## II. DEFINITIONS

[0073] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, preferred methods and materials are described. For the purposes of the present invention, the following terms are defined below.

[0074] The articles “a” and “an” are used herein to refer to one or to more than one (*i.e.*, to at least one) of the grammatical object of the article. By way of example, “an element” means one element or more than one element.

[0075] By “about” is meant a quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length that varies by as much as 30, 25, 20, 25, 10,



9, 8, 7, 6, 5, 4, 3, 2 or 1% to a reference quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length.

**[0076]** An “agonist” refers to a molecule that intensifies or mimics an activity. For example, a non-canonical biological activity of an AARS, or another protein. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition that modulates the activity of an AARS either by directly interacting with the AARS or its binding partner, or by acting on components of the biological pathway in which the AARS participates. Included are partial and full agonists.

**[0077]** As used herein, the term “amino acid” is intended to mean both naturally occurring and non-naturally occurring amino acids as well as amino acid analogs and mimetics. Naturally occurring amino acids include the 20 (L)-amino acids utilized during protein biosynthesis as well as others such as 4-hydroxyproline, hydroxylysine, desmosine, isodesmosine, homocysteine, citrulline and ornithine, for example. Non-naturally occurring amino acids include, for example, (D)-amino acids, norleucine, norvaline, p-fluorophenylalanine, ethionine and the like, which are known to a person skilled in the art. Amino acid analogs include modified forms of naturally and non-naturally occurring amino acids. Such modifications can include, for example, substitution or replacement of chemical groups and moieties on the amino acid or by derivitization of the amino acid. Amino acid mimetics include, for example, organic structures which exhibit functionally similar properties such as charge and charge spacing characteristic of the reference amino acid. For example, an organic structure which mimics Arginine (Arg or R) would have a positive charge moiety located in similar molecular space and having the same degree of mobility as the  $\epsilon$ -amino group of the side chain of the naturally occurring Arg amino acid. Mimetics also include constrained structures so as to maintain optimal spacing and charge interactions of the amino acid or of the amino acid functional groups. Those skilled in the art know or can determine what structures constitute functionally equivalent amino acid analogs and amino acid mimetics.

**[0078]** In certain aspects, the use of non-natural amino acids can be utilized to modify (*e.g.*, increase) a selected non-canonical activity of an AARS protein fragment, or to alter the *in vivo* or *in vitro* half-life of the protein. Non-natural amino acids can also be used to facilitate (selective) chemical modifications (*e.g.*, pegylation) of an AARS protein. For instance, certain non-natural amino acids allow selective attachment of polymers such as PEG to a given protein, and thereby improve their pharmacokinetic properties.

[0079] Specific examples of amino acid analogs and mimetics can be found described in, for example, Roberts and Vellaccio, *The Peptides: Analysis, Synthesis, Biology*, Eds. Gross and Meinhofer, Vol. 5, p. 341, Academic Press, Inc., New York, N.Y. (1983), the entire volume of which is incorporated herein by reference. Other examples include peralkylated amino acids, particularly permethylated amino acids. See, for example, *Combinatorial Chemistry*, Eds. Wilson and Czarnik, Ch. 11, p. 235, John Wiley & Sons Inc., New York, N.Y. (1997), the entire book of which is incorporated herein by reference. Yet other examples include amino acids whose amide portion (and, therefore, the amide backbone of the resulting peptide) has been replaced, for example, by a sugar ring, steroid, benzodiazepine or carbo cycle. See, for instance, *Burger's Medicinal Chemistry and Drug Discovery*, Ed. Manfred E. Wolff, Ch. 15, pp. 619-620, John Wiley & Sons Inc., New York, N.Y. (1995), the entire book of which is incorporated herein by reference. Methods for synthesizing peptides, polypeptides, peptidomimetics and proteins are well known in the art (see, for example, U.S. Pat. No. 5,420,109; M. Bodanzsky, *Principles of Peptide Synthesis* (1st ed. & 2d rev. ed.), Springer-Verlag, New York, N.Y. (1984 & 1993), see Chapter 7; Stewart and Young, *Solid Phase Peptide Synthesis*, (2d ed.), Pierce Chemical Co., Rockford, Ill. (1984), each of which is incorporated herein by reference). Accordingly, the AARS polypeptides of the present invention may be composed of naturally occurring and non-naturally occurring amino acids as well as amino acid analogs and mimetics.

[0080] The term "antagonist" refers to a molecule that reduces or attenuates an activity. For example, a non-canonical biological activity of an AARS, or another protein. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition that modulates the activity of an AARS or its binding partner, either by directly interacting with the AARS or its binding partner or by acting on components of the biological pathway in which the AARS participates. Included are partial and full antagonists.

[0081] The term "aminoacyl-tRNA synthetase" (AARS) refers generally to enzymes that in their natural or wild-type form are capable of catalyzing the esterification of a specific amino acid or its precursor to one of all its compatible cognate tRNAs to form an aminoacyl-tRNA. In this "canonical" activity, aminoacyl-tRNA synthetases catalyze a two-step reaction: first, they activate their respective amino acid by forming an aminoacyl-adenylate, in which the carboxyl of the amino acid is linked in to the alpha-phosphate of ATP by displacing pyrophosphate, and then, when the correct tRNA is bound, the aminoacyl group of the aminoacyl-adenylate is transferred to the 2' or 3' terminal OH of the tRNA.

[0082] Class I aminoacyl-tRNA synthetases typically have two highly conserved sequence motifs. These enzymes aminoacylate at the 2'-OH of an adenosine nucleotide, and are usually monomeric or dimeric. Class II aminoacyl-tRNA synthetases typically have three highly conserved sequence motifs. These enzymes aminoacylate at the 3'-OH of the same adenosine, and are usually dimeric or tetrameric. The active sites of class II enzymes are mainly made up of a seven-stranded anti-parallel  $\beta$ -sheet flanked by  $\alpha$ -helices. Although phenylalanine-tRNA synthetase is class II, it aminoacylates at the 2'-OH.

[0083] AARS polypeptides include sources of mitochondrial and cytoplasmic forms of tyrosyl-tRNA synthetase (TyrRS), a tryptophanyl-tRNA synthetase (TrpRS), a glutaminyl-tRNA synthetase (GlnRS), a glycyl-tRNA synthetase (GlyRS), a histidyl-tRNA synthetase (HisRS), a seryl-tRNA synthetase (SerRS), a phenylalanyl-tRNA synthetase (PheRS), an alanyl-tRNA synthetase (AlaRS), an asparaginyl-tRNA synthetase (AsnRS), an aspartyl-tRNA synthetase (AspRS), a cysteinyl-tRNA synthetase (CysRS), a glutamyl-tRNA synthetase (GluRS), a prolyl-tRNA synthetase (ProRS), an arginyl-tRNA synthetase (ArgRS), an isoleucyl-tRNA synthetase (IleRS), a leucyl-tRNA synthetase (LeuRS), a lysyl-tRNA synthetase (LysRS), a threonyl-tRNA synthetase (ThrRS), a methionyl-tRNA synthetases (MetRS), or a valyl-tRNA synthetase (ValRS). The wild-type or parental sequences of these AARS polypeptides are known in the art.

[0084] By "coding sequence" is meant any nucleic acid sequence that contributes to the code for the polypeptide product of a gene. By contrast, the term "non-coding sequence" refers to any nucleic acid sequence that does not contribute to the code for the polypeptide product of a gene.

[0085] Throughout this specification, unless the context requires otherwise, the words "comprise," "comprises," and "comprising" will be understood to imply the inclusion of a stated step or element or group of steps or elements but not the exclusion of any other step or element or group of steps or elements.

[0086] By "consisting of" is meant including, and limited to, whatever follows the phrase "consisting of." Thus, the phrase "consisting of" indicates that the listed elements are required or mandatory, and that no other elements may be present. By "consisting essentially of" is meant including any elements listed after the phrase, and limited to other elements that do not interfere with or contribute to the activity or action specified in the disclosure for the listed elements. Thus, the phrase "consisting essentially of" indicates that the listed elements are required or mandatory, but that other elements are optional and may or may not be present depending upon whether or not they materially affect the activity or action of the listed elements.

[0087] The recitation “endotoxin free” or “substantially endotoxin free” relates generally to compositions, solvents, and/or vessels that contain at most trace amounts (e.g., amounts having no clinically adverse physiological effects to a subject) of endotoxin, and preferably undetectable amounts of endotoxin. Endotoxins are toxins associated with certain bacteria, typically gram-negative bacteria, although endotoxins may be found in gram-positive bacteria, such as *Listeria monocytogenes*. The most prevalent endotoxins are lipopolysaccharides (LPS) or lipo-oligo-saccharides (LOS) found in the outer membrane of various Gram-negative bacteria, and which represent a central pathogenic feature in the ability of these bacteria to cause disease. Small amounts of endotoxin in humans may produce fever, a lowering of the blood pressure, and activation of inflammation and coagulation, among other adverse physiological effects.

[0088] Therefore, in pharmaceutical production of AARS polypeptides, it is often desirable to remove most or all traces of endotoxin from drug products and/or drug containers, because even small amounts may cause adverse effects in humans. A depyrogenation oven may be used for this purpose, as temperatures in excess of 300°C are typically required to break down most endotoxins. For instance, based on primary packaging material such as syringes or vials, the combination of a glass temperature of 250°C and a holding time of 30 minutes is often sufficient to achieve a 3 log reduction in endotoxin levels. Other methods of removing endotoxins are contemplated, including, for example, chromatography and filtration methods, as described herein and known in the art. Also included are methods of producing AARS polypeptides in, and isolating them from, eukaryotic cells such as mammalian cells, to reduce if not eliminate the risk of endotoxins being present in a composition of the invention. Preferred are methods of producing AARS polypeptides in and isolating them from serum free cells. Such compositions comprising AARS polypeptides, represent new formulations which exhibit novel and new biological and therapeutic characteristics not found in AARS polypeptide compositions contaminated with serum or endotoxin which have the potential to bind to and alter the novel biological properties of the AARS polypeptides.

[0089] Endotoxins can be detected using routine techniques known in the art. For example, the Limulus Amebocyte Lysate assay, which utilizes blood from the horseshoe crab, is a very sensitive assay for detecting presence of endotoxin, and reagents, kits and instrumentation for the detection of endotoxin based on this assay are commercially available, for example from the Lonza Group. In this test, very low levels of LPS can cause detectable coagulation of the limulus lysate due a powerful enzymatic cascade that amplifies this reaction. Endotoxins can also be quantitated by

enzyme-linked immunosorbent assay (ELISA). To be substantially endotoxin free, endotoxin levels may be less than about 0.001, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.08, 0.09, 0.1, 0.5, 1.0, 1.5, 2, 2.5, 3, 4, 5, 6, 7, 8, 9, or 10 EU /mg of protein. Typically, 1 ng lipopolysaccharide (LPS) corresponds to about 1-10 EU.

[0090] In certain embodiments, the “purity” of any given agent (*e.g.*, AARS protein fragment) in a composition may be specifically defined. For instance, certain compositions may comprise an agent that is at least 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% pure, including all decimals in between, as measured, for example and by no means limiting, by high pressure liquid chromatography (HPLC), a well-known form of column chromatography used frequently in biochemistry and analytical chemistry to separate, identify, and quantify compounds.

[0091] As used herein, the terms “function” and “functional” and the like refer to a biological, enzymatic, or therapeutic function.

[0092] By “gene” is meant a unit of inheritance that may occupy a specific locus on a chromosome and consists of transcriptional and/or translational regulatory sequences and/or a coding region and/or non-translated sequences (*i.e.*, introns, 5’ and 3’ untranslated sequences).

[0093] “Homology” refers to the percentage number of amino acids that are identical or constitute conservative substitutions. Homology may be determined using sequence comparison programs such as GAP (Deveraux *et al.*, 1984, *Nucleic Acids Research* 12, 387-395), which is incorporated herein by reference. In this way sequences of a similar or substantially different length to those cited herein could be compared by insertion of gaps into the alignment, such gaps being determined, for example, by the comparison algorithm used by GAP.

[0094] The term “host cell” includes an individual cell or cell culture that can be or has been a recipient of any recombinant vector(s), isolated polynucleotide, or polypeptide of the invention. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation and/or change. A host cell includes cells transfected or infected *in vivo* or *in vitro* with a recombinant vector or a polynucleotide of the invention. A host cell which comprises a recombinant vector of the invention is a recombinant host cell.

[0095] By “isolated” is meant material that is substantially or essentially free from components that normally accompany it in its native state. For example, an “isolated polynucleotide,” as used herein, includes a polynucleotide that has been purified from the sequences that flank it in its naturally-occurring state, *e.g.*, a DNA fragment which has been removed from the sequences that are normally adjacent to the fragment.

Alternatively, an “isolated peptide” or an “isolated polypeptide” and the like, as used herein, includes the *in vitro* isolation and/or purification of a peptide or polypeptide molecule from its natural cellular environment, and from association with other components of the cell; *i.e.*, it is not significantly associated with *in vivo* substances.

[0096] The term “mRNA” or sometimes refer by “mRNA transcripts” as used herein, include, but not limited to pre-mRNA transcript(s), transcript processing intermediates, mature mRNA(s) ready for translation and transcripts of the gene or genes, or nucleic acids derived from the mRNA transcript(s). Transcript processing may include splicing, editing and degradation. As used herein, a nucleic acid derived from an mRNA transcript refers to a nucleic acid for whose synthesis the mRNA transcript or a subsequence thereof has ultimately served as a template. A cDNA reverse transcribed from an mRNA, an RNA transcribed from that cDNA, a DNA amplified from the cDNA, an RNA transcribed from the amplified DNA, etc., are all derived from the mRNA transcript and detection of such derived products is indicative of the presence and/or abundance of the original transcript in a sample. Thus, mRNA derived samples include, but are not limited to, mRNA transcripts of the gene or genes, cDNA reverse transcribed from the mRNA, cRNA transcribed from the cDNA, DNA amplified from the genes, RNA transcribed from amplified DNA, and the like.

[0097] “Non-canonical” activity as used herein, refers generally to either i) a new activity possessed by an AARS polypeptide of the invention that is not possessed to any significant degree by the intact native full length parental protein, or ii) an activity that was possessed by the by the intact native full length parental protein, where the AARS polypeptide either exhibits a significantly higher (*i.e.* at least 20% greater) specific activity compared to the intact native full length parental protein, or exhibits the activity in a new context; for example by isolating the activity from other activities possessed by the intact native full length parental protein. In the case of AARS polypeptides, non-limiting examples of non-canonical activities include extracellular signaling, RNA-binding, amino acid-binding, modulation of cell proliferation, modulation of cell migration, modulation of cell differentiation (*e.g.*, hematopoiesis, neurogenesis, myogenesis, osteogenesis, and adipogenesis), modulation of gene transcription, modulation of apoptosis or other forms of cell death, modulation of cell signaling, modulation of cellular uptake, or secretion, modulation of angiogenesis, modulation of cell binding, modulation of cellular metabolism, modulation of cytokine production or activity, modulation of cytokine receptor activity, modulation of inflammation, and the like.

[0098] The term “half maximal effective concentration” or “EC<sub>50</sub>” refers to the concentration of an AARS protein fragment, antibody or other agent described herein at

which it induces a response halfway between the baseline and maximum after some specified exposure time; the  $EC_{50}$  of a graded dose response curve therefore represents the concentration of a compound at which 50% of its maximal effect is observed. In certain embodiments, the  $EC_{50}$  of an agent provided herein is indicated in relation to a “non-canonical” activity, as noted above.  $EC_{50}$  also represents the plasma concentration required for obtaining 50% of a maximum effect *in vivo*. Similarly, the “ $EC_{90}$ ” refers to the concentration of an agent or composition at which 90% of its maximal effect is observed. The “ $EC_{90}$ ” can be calculated from the “ $EC_{50}$ ” and the Hill slope, or it can be determined from the data directly, using routine knowledge in the art. In some embodiments, the  $EC_{50}$  of an AARS protein fragment, antibody, or other agent is less than about 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 nM. Preferably, biotherapeutic composition will have an  $EC_{50}$  value of about 1nM or less.

[0099] The term “modulating” includes “increasing” or “stimulating,” as well as “decreasing” or “reducing,” typically in a statistically significant or a physiologically significant amount as compared to a control. Accordingly a “modulator” may be an agonist, an antagonist, or any mixture thereof depending upon the conditions used. An “increased” or “enhanced” amount is typically a “statistically significant” amount, and may include an increase that is 1.1, 1.2, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30 or more times (*e.g.*, 500, 1000 times) (including all integers and decimal points in between and above 1, *e.g.*, 1.5, 1.6, 1.7, 1.8, etc.) the amount produced by no composition (the absence of an agent or compound) or a control composition. A “decreased” or reduced amount is typically a “statistically significant” amount, and may include a 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% decrease in the amount produced by no composition (the absence of an agent or compound) or a control composition, including all integers in between. As one non-limiting example, a control in comparing canonical and non-canonical activities could include the AARS protein fragment of interest compared to its corresponding full-length AARS, or a fragment AARS having comparable canonical activity to its corresponding full-length AARS. Other examples of “statistically significant” amounts are described herein.

[00100] By “obtained from” is meant that a sample such as, for example, a polynucleotide extract or polypeptide extract is isolated from, or derived from, a particular source of the subject. For example, the extract can be obtained from a tissue or a biological fluid isolated directly from the subject. “Derived” or “obtained from” can also refer to the source of a polypeptide or polynucleotide sequence. For instance,

an AARS sequence of the present invention may be “derived” from the sequence information of an AARS proteolytic fragment or AARS splice variant, or a portion thereof, whether naturally-occurring or artificially generated, and may thus comprise, consist essentially of, or consist of that sequence.

**[00101]** The terms “polypeptide” and “protein” are used interchangeably herein to refer to a polymer of amino acid residues and to variants and synthetic and naturally occurring analogues of the same. Thus, these terms apply to amino acid polymers in which one or more amino acid residues are synthetic non-naturally occurring amino acids, such as a chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally-occurring amino acid polymers and naturally occurring chemical derivatives thereof. Such derivatives include, for example, post-translational modifications and degradation products including pyroglutamyl, iso-aspartyl, proteolytic, phosphorylated, glycosylated, oxidatized, isomerized, and deaminated variants of the AARS reference fragment.

**[00102]** The recitations “sequence identity” or, for example, comprising a “sequence 50% identical to,” as used herein, refer to the extent that sequences are identical on a nucleotide-by-nucleotide basis or an amino acid-by-amino acid basis over a window of comparison. Thus, a “percentage of sequence identity” may be calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, I) or the identical amino acid residue (*e.g.*, Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys and Met) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity.

**[00103]** Terms used to describe sequence relationships between two or more polynucleotides or polypeptides include “reference sequence,” “comparison window,” “sequence identity,” “percentage of sequence identity” and “substantial identity.” A “reference sequence” is at least 12 but frequently 15 to 18 and often at least 25 monomer units, inclusive of nucleotides and amino acid residues, in length. Because two polynucleotides may each comprise (1) a sequence (*i.e.*, only a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a “comparison window” to identify and compare local regions of sequence similarity. A “comparison window” refers to a



conceptual segment of at least 6 contiguous positions, usually about 50 to about 100, more usually about 100 to about 150 in which a sequence is compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. The comparison window may comprise additions or deletions (*i.e.*, gaps) of about 20% or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by computerized implementations of algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Drive Madison, WI, USA) or by inspection and the best alignment (*i.e.*, resulting in the highest percentage homology over the comparison window) generated by any of the various methods selected. Reference also may be made to the BLAST family of programs as for example disclosed by Altschul *et al.*, 1997, *Nucl. Acids Res.* 25:3389. A detailed discussion of sequence analysis can be found in Unit 19.3 of Ausubel *et al.*, "Current Protocols in Molecular Biology," John Wiley & Sons Inc, 1994-1998, Chapter 15.

**[00104]** Calculations of sequence similarity or sequence identity between sequences (the terms are used interchangeably herein) are performed as follows. To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In certain embodiments, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position.

**[00105]** The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

**[00106]** The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch, (1970, *J. Mol. Biol.* 48: 444-453) algorithm which

has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used unless otherwise specified) are a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frame shift gap penalty of 5.

**[00107]** The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of E. Meyers and W. Miller (1989, *Cabios*, **4**: 11-17) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

**[00108]** The nucleic acid and protein sequences described herein can be used as a “query sequence” to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.*, (1990, *J. Mol. Biol.*, **215**: 403-10). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997, *Nucleic Acids Res.*, **25**: 3389-3402). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used.

**[00109]** The term “solubility” refers to the property of an agent provided herein to dissolve in a liquid solvent and form a homogeneous solution. Solubility is typically expressed as a concentration, either by mass of solute per unit volume of solvent (g of solute per kg of solvent, g per dL (100 mL), mg/ml, etc.), molarity, molality, mole fraction or other similar descriptions of concentration. The maximum equilibrium amount of solute that can dissolve per amount of solvent is the solubility of that solute in that solvent under the specified conditions, including temperature, pressure, pH, and the nature of the solvent. In certain embodiments, solubility is measured at physiological pH. In certain embodiments, solubility is measured in water or a physiological buffer such as PBS. In certain embodiments, solubility is measured in a

biological fluid (solvent) such as blood or serum. In certain embodiments, the temperature can be about room temperature (*e.g.*, about 20, 21, 22, 23, 24, 25°C) or about body temperature (37°C). In certain embodiments, an agent such as an AARS protein fragment has a solubility of at least about 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, or 30 mg/ml at room temperature or at 37°C.

**[00110]** A “splice junction” as used herein includes the region in a mature mRNA transcript or the encoded polypeptide where the 3’ end of a first exon joins with the 5’ end of a second exon. The size of the region may vary, and may include 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or more (including all integers in between) nucleotide or amino acid residues on either side of the exact residues where the 3’ end of one exon joins with the 5’ end of another exon. An “exon” refers to a nucleic acid sequence that is represented in the mature form of an RNA molecule after either portions of a precursor RNA (introns) have been removed by cis-splicing or two or more precursor RNA molecules have been ligated by trans-splicing. The mature RNA molecule can be a messenger RNA or a functional form of a non-coding RNA such as rRNA or tRNA. Depending on the context, an exon can refer to the sequence in the DNA or its RNA transcript. An “intron” refers to a non-coding nucleic acid region within a gene, which is not translated into a protein. Non-coding intronic sections are transcribed to precursor mRNA (pre-mRNA) and some other RNAs (such as long noncoding RNAs), and subsequently removed by splicing during the processing to mature RNA.

**[00111]** A “splice variant” refers to a mature mRNA and its encoded protein that are produced by alternative splicing, a process by which the exons of the RNA (a primary gene transcript or pre-mRNA) are reconnected in multiple ways during RNA splicing. The resulting different mRNAs may be translated into different protein isoforms, allowing a single gene to code for multiple proteins.

**[00112]** A “subject,” as used herein, includes any animal that exhibits a symptom, or is at risk for exhibiting a symptom, which can be treated or diagnosed with an AARS polynucleotide or polypeptide of the invention. Also included are subjects for which it is desirable to profile levels of AARS polypeptides and/or polynucleotides of the invention, for diagnostic or other purposes. Suitable subjects (patients) include laboratory animals (such as mouse, rat, rabbit, or guinea pig), farm animals, and domestic animals or pets (such as a cat or dog). Non-human primates and, preferably, human patients, are included.

**[00113]** “Treatment” or “treating,” as used herein, includes any desirable effect on the symptoms or pathology of a disease or condition that can be effected by the non-

canonical activities of an AARS polynucleotide or polypeptide, as described herein, and may include even minimal changes or improvements in one or more measurable markers of the disease or condition being treated. Also included are treatments that relate to non-AARS therapies, in which an AARS sequence described herein provides a clinical marker of treatment. "Treatment" or "treating" does not necessarily indicate complete eradication or cure of the disease or condition, or associated symptoms thereof. The subject receiving this treatment is any subject in need thereof. Exemplary markers of clinical improvement will be apparent to persons skilled in the art.

[00114] The practice of the present invention will employ, unless indicated specifically to the contrary, conventional methods of molecular biology and recombinant DNA techniques within the skill of the art, many of which are described below for the purpose of illustration. Such techniques are explained fully in the literature. See, e.g., Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual* (3<sup>rd</sup> Edition, 2001); *DNA Cloning: A Practical Approach*, vol. I & II (D. Glover, ed.); *Oligonucleotide Synthesis* (N. Gait, ed., 1984); *Oligonucleotide Synthesis: Methods and Applications* (P. Herdewijn, ed., 2004); *Nucleic Acid Hybridization* (B. Hames & S. Higgins, eds., 1985); *Nucleic Acid Hybridization: Modern Applications* (Buzdin and Lukyanov, eds., 2009); *Transcription and Translation* (B. Hames & S. Higgins, eds., 1984); *Animal Cell Culture* (R. Freshney, ed., 1986); Freshney, R.I. (2005) *Culture of Animal Cells, a Manual of Basic Technique*, 5<sup>th</sup> Ed. Hoboken NJ, John Wiley & Sons; B. Perbal, *A Practical Guide to Molecular Cloning* (3<sup>rd</sup> Edition 2010); Farrell, R., *RNA Methodologies: A Laboratory Guide for Isolation and Characterization* (3<sup>rd</sup> Edition 2005), *Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA* Methods in Enzymology, Academic Press; *Using Antibodies: A Laboratory Manual: Portable Protocol NO. 1* by Edward Harlow, David Lane, Ed Harlow (1999, Cold Spring Harbor Laboratory Press, ISBN 0-87969-544-7); *Antibodies: A Laboratory Manual* by Ed Harlow (Editor), David Lane (Editor) (1988, Cold Spring Harbor Laboratory Press, ISBN 0-87969-3,4-2), 1855. *Handbook of Drug Screening*, edited by Ramakrishna Seethala, Prabhavathi B. Fernandes (2001, New York, N.Y., Marcel Dekker, ISBN 0-8247-0562-9); and *Lab Ref: A Handbook of Recipes, Reagents, and Other Reference Tools for Use at the Bench*, Edited Jane Roskams and Linda Rodgers, (2002, Cold Spring Harbor Laboratory, ISBN 0-87969-630-3).

[00115] All publications, patents and patent applications cited herein are hereby incorporated by reference in their entirety.

### III. PURIFIED AARS PROTEIN FRAGMENTS AND VARIANTS FOR THERAPEUTICS AND OTHER APPLICATIONS

[00116] Surprisingly, and unlike their full-length parental sequences that are known only for their aminoacylation-activities, it has been found that AARS fragments possess biological activities *important for* biotherapeutic, discovery and diagnostic applications. Embodiments of the present invention therefore include full length proteins, mature protein isoforms and protein fragments of aminoacyl-tRNA synthetases (AARS), in addition to biologically active variants and fragments thereof. In certain embodiments, the proteins and fragments may arise through endogenous proteolysis, *in vitro* proteolysis, splice variation, or *in silico* prediction, among other mechanisms. The AARS protein fragments described herein, and variants thereof, may possess at least one “non-canonical” biological activity. The AARS protein fragment(s) of the present invention are also referred to herein as “AARS polypeptides” or “AARS reference polypeptides.” In certain embodiments, the AARS polypeptides provided herein comprise or consist essentially of all or a portion of the AARS polypeptide “reference sequence(s)” as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9 below, which represent the amino acid sequence(s) of various fragments of Alanyl tRNA synthetases. Mouse and human AARS protein sequences are highly related, typically differing by no more than a few amino acids within an entire sequence, a particular domain, or a particular protein fragment.

#### N-terminal AARS Polypeptides: (Tables 1, 2 & 3)

Table 1A			
N-terminal AARS polypeptides identified by MS			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID . NO.
AlaRS1 <sup>N</sup> <sub>1</sub>	Protein / Human / 1-401	MDSTLTASEIRQRFIDFFKRNEHTYVHSSATIP LDDPTLLFANAGMNQFKPIFLNTIDPSHPMAK LSRAANTQKCIRAGGKHNDLDDVGKDVYHH TFFEMLGWSWFGDYFKELACKMALELLTQEF GIPIERLYVTYFGGDEAAGLEADLECKQIWQN LGLDDTKILPGNMKDNFWEMGDTGPCGPCS EIHYDRIGGRDAAHLVNQDDPNVLEIWNLVFI QYNREADGILKPLPKKSIDTGMGLERLVSVLQ NKMSNYDTDLFVPYFEAIQKGTGARPYTGKV GAEDADGIDMAYRVLADHARTITVALADGG RPDNTGRGYVLRRLRAVRYAHEKLNASRG FFATLVDVVVQSLGDAFPELKKDPDMVKDII NEEEVQFLKTLRSRGRRLDRKIQSLG	SEQ.ID. NO.12
AlaRS1 <sup>N</sup>	DNA /	ATGGACTCTACTCTAACAGCAAGTGAAATC	SEQ.ID.

Table 1A N-terminal AARS polypeptides identified by MS			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID . NO.
1	Human /	CGGCAGCGATTTATAGATTTCTTCAAGAGG AACGAGCATACGTATGTTCACTCGTCTGCC ACCATCCCATTGGATGACCCCACTTTGCTCT TTGCCAATGCAGGCATGAACCAGTTTAAAC CCATTTTCCTGAACACAATTGACCCATCTCA CCCCATGGCAAAGCTGAGCAGAGCTGCCAA TACCCAGAAGTGCATCCGGGCTGGGGGCAA ACATAATGACCTGGACGATGTGGGCAAGGA TGTCTATCATCACACCTTCTTCGAGATGCTG GGCTCTTGGTCTTTTGGAGATTACTTTAAGG AATTGGCATGTAAGATGGCTCTGGAATCC TCACCCAAGAGTTTGGCATTCCCATTGAAA GACTTTATGTTACTTACTTTGGCGGGGATGA AGCAGCTGGCTTAGAAGCAGATCTGGAATG CAAACAGATCTGGCAAAATTTGGGGCTGGA TGACACCAAATCCTCCCAGGCAACATGAA GGATAACTTCTGGGAGATGGGTGACACGGG CCCCTGTGGTCCTTGCAGTGAGATCCACTA CGACCGGATTGGTGGTCGGGACGCCGCACA TCTTGTCAACCAGGACGACCCTAATGTGCT GGAGATCTGGAACCTTGTGTTTCATCCAGTA TAACAGGGAAGCTGATGGCATTCTGAAACC TCTTCCCAAGAAAAGCATTGACACAGGGAT GGGCCTGGAACGACTGGTATCTGTGCTGCA GAATAAGATGTCCAACTATGACACTGACCT TTTTGTCCCTTACTTTGAAGCCATTGAGAAG GGCACAGGTGCCCCGACCATACTGGGAA AGTTGGTGCTGAGGATGCCGATGGGATTGA CATGGCCTACCGGGTGCTGGCTGACCACGC TCGGACCATCACTGTGGCACTGGCTGATGG TGGCCGGCCTGACAACACAGGGCGTGGATA TGTGTTGAGACGGATTCTCCGCCGAGCTGT CCGATACGCCCATGAAAAGCTCAATGCCAG CAGGGGCTTCTTTGCTACGTTAGTGGATGTT GTCGTCCAGTCCCTGGGAGATGCATTTCT GAGCTGAAGAAGGACCCAGACATGGTGAA GGACATCATTAATGAAGAAGAGGTGCAGTT TCTCAAGACTCTCAGCAGAGGGCGTCGCAT CCTGGACAGGAAAATTCAGAGCCTGGGA	NO.13
AlaRS1 <sup>N</sup> 5	Protein / Human /  1-756	MDSTLTASEIRQRFIDFFKRNEHTYVHSSATIP LDDPTLLFANAGMNQFKPIFLNTIDPSHPMAK LSRAANTQKCIRAGGKHNDLDDVGKDVYHH TFFEMLGSWSFGDYFKELACKMALELLTQEF	SEQ.ID. NO.14

Table 1A N-terminal AARS polypeptides identified by MS			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID . NO.
		GIPIERLYVTYFGGDEAAGLEADLECKQIWQN LGLDDTKILPGNMKDNFWEMGDTGPCGPCS EIHVDRIGGRDAAHLVNQDDPNVLEIWNLVFI QYNREADGILKPLPKKSIDTGMGLERLVSVLQ NKMSNYDSDLFVPYFEAIQKGTGARPYTGKV GAEDADGIDMAYRVLADHARTITVALADGG RPDNTGRGYVLRRLRRAVRYAHEKLNASRG FFATLVDVVVQSLGDAFPELKKDPDMVKDII NEEVQFLKTLRGRRLDRKIQSLGDSKTIPTG DTAWLLYDTYGFPVDLTGLIAEEKGLVMD DGFEERKLAQLKSQKGAGGEDLIMLDIYAI EELRARGLEVTDSPKYNHLDSSGSYVFEN TVATVMALRREKMFVEEVSTGQECGVVLDK TCFYAEQGGQIYDEGYLVKVDDSSDKTEFT VKNAQVRGGYVLHIGTIYGDLKVGQVWLF DEPRRRPIMSNHTATHILNFALRSVLGEADQK GSLVAPDRLRFDFATAKGAMSTQQIKKAEI NEMIEAAKAVYTQDCPLAAAKAIQGLRAVFD ETYPDPVRVVSIGVPVSELLDDPSGPAGSLTS VEFCCGTHLRNSSHAGAFVIVTEEAIAKIRRI VAVT	
AlaRS1 <sup>N</sup> <sub>5</sub>	DNA / Human	ATGGACTCTACTCTAACAGCAAGTGAAATC CGGCAGCGATTTATAGATTTCTTCAAGAGG AACGAGCATACGTATGTTCACTCGTCTGCC ACCATCCCATTGGATGACCCCACTTTGCTCT TTGCCAATGCAGGCATGAACCAGTTTAAAC CCATTTTCCTGAACACAATTGACCCATCTCA CCCCATGGCAAAGCTGAGCAGAGCTGCCAA TACCCAGAAGTGCATCCGGGCTGGGGGCAA ACATAATGACCTGGACGATGTGGGCAAGGA TGTCTATCATCACACCTTCTTCGAGATGCTG GGCTCTTGGTCTTTTGGAGATTACTTTAAGG AATTGGCATGTAAGATGGCTCTGGAACCTC TCACCCAAGAGTTTGGCATTCCCATTGAAA GACTTTATGTTACTTACTTTGGCGGGGATGA AGCAGCTGGCTTAGAAGCAGATCTGGAATG CAAACAGATCTGGCAAAATTTGGGGCTGGA TGACACCAAATCCTCCCAGGCAACATGAA GGATAACTTCTGGGAGATGGGTGACACGGG CCCCTGTGGTCCTTGCAGTGAGATCCACTA CGACCGGATTGGTGGTCGGGACGCCGCACA TCTTGTCAACCAGGACGACCCTAATGTGCT GGAGATCTGGAACCTTGTGTTTCATCCAGTA TAACAGGGAAGCTGATGGCATTCTGAAACC	SEQ.ID. NO.15

Table 1A N-terminal AARS polypeptides identified by MS			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID . NO.
		TCTTCCCAAGAAAAGCATTGACACAGGGAT GGGCCTGGAACGACTGGTATCTGTGCTGCA GAATAAGATGTCCAACATGACACTGACCT TTTTGTCCCTTACTTTGAAGCCATTCAGAAG GGCACAGGTGCCCCGACCATACTGGGAA AGTTGGTGCTGAGGATGCCGATGGGATTGA CATGGCCTACCGGGTGCTGGCTGACCACGC TCGGACCATCACTGTGGCACTGGCTGATGG TGGCCGGCCTGACAACACAGGGCGTGGATA TGTGTTGAGACGGATTCTCCGCCGAGCTGT CCGATACGCCCATGAAAAGCTCAATGCCAG CAGGGGCTTCTTTGCTACGTTAGTGGATGTT GTCGTCCAGTCCCTGGGAGATGCATTTCTT GAGCTGAAGAAGGACCCAGACATGGTGAA GGACATCATTAATGAAGAAGAGGTGCAGTT TCTCAAGACTCTCAGCAGAGGGCGTCGCAT CCTGGACAGGAAAATTCAGAGCCTGGGAG ACAGCAAGACCATTCCCGGAGACACTGCTT GGCTCCTCTATGACACCTATGGGTTTCCAGT GGATCTGACTGGACTGATTGCTGAAGAGAA GGGCCTGGTGGTAGACATGGATGGCTTTGA AGAGGAGAGGAACTGGCCCAGCTGAAAT CACAGGGCAAGGGAGCTGGTGGGGAAGAC CTCATTATGCTGGACATTTACGCTATCGAA GAGCTCCGGGCACGGGGTCTGGAGGTCACA GATGATTCCCCAAAGTACAATTACCATTTG GACTCCAGTGGTAGCTATGTATTTGAGAAC ACAGTGGCTACGGTGATGGCTCTGCGCAGG GAGAAGATGTTTCGTGGAAGAGGTGTCCACA GGCCAGGAGTGTGGAGTGGTGCTGGACAA GACCTGTTTCTATGCTGAGCAAGGAGGCCA GATCTATGACGAAGGCTACCTGGTGAAGGT GGATGACAGCAGTGAAGATAAACAGAGT TTACAGTGAAGAATGCTCAGGTCCGAGGAG GGTATGTGCTACACATTGGAACCATCTACG GTGACCTGAAAGTGGGGGATCAGGTCTGGC TGTTTATTGATGAGCCCCGACGAAGACCCA TCATGAGCAACCACACAGCTACGCACATTC TGAAGTTCGCCCTGCGCTCAGTGCTTGGGG AAGCTGACCAGAAAGGCTCATTGGTTGCTC CTGACCGCCTCAGATTTGACTTTACTGCCA AGGGAGCCATGTCCACCCAACAGATCAAG AAGGCTGAAGAGATTGCTAATGAGATGATT GAGGCAGCCAAGGCCGTCTATACCCAGGAT	



Table 1A N-terminal AARS polypeptides identified by MS			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID . NO.
		TGCCCCCTGGCAGCAGCGAAAGCCATCCAG GGCCTACGGGCTGTGTTTGATGAGACCTAT CCTGACCCTGTGCGAGTCGTCTCCATTGGG GTCCCGGTGTCCGAGTTGCTGGATGACCCC TCTGGGCCTGCTGGCTCCCTGACTTCTGTTG AGTTCTGTGGGGGAACGCACCTGCGGAACT CGAGTCATGCAGGAGCTTTTGTGATCGTGA CGGAAGAAGCCATTGCCAAGGGTATCCGG AGGATTGTGGCTGTCACA	

Table 1B AlaRS1 <sup>N1</sup> Mass spec peptides detected and inferred linking peptides		
Type / species	Sequence	SEQ.ID . NO.
Protein / mouse	LYVTYFGGDEAAGLEPDLECR	SEQ.ID. NO. 16
Protein / mouse	QIWQNLGLDEARILPGNMKDNFWEMGDTGPCGPCSEIHY DRIGGRDAAHLVNQDDPNVLEIWNLVFIQYNRES DGVLK PLPKKSIDTGMGLERLVSVLQNKMSNYD TDLFMPYFEAI QKGTGARPYTGKVGAEADADGIDMAYRVLADHAR	SEQ.ID. NO. 17
Protein / mouse	TITVALADGGRPDNTGR	SEQ.ID. NO. 18
Protein / mouse	GYVLRRLRRAVRYSHEKLNASRGFFATLVDV VVQSLGD AFPELKKDPEMVK	SEQ.ID. NO. 19
Protein / mouse	DIINEEEVQFLK	SEQ.ID. NO. 20

Table 1C AlaRS1 <sup>N1</sup> Concatenated sequences based on mass spec peptides detected		
Type / species	Sequence	SEQ.ID . NO.
Protein / mouse	LYVTYFGGDEAAGLEPDLECRQIWQNLGLDEARILPGNM KDNFWEMGDTGPCGPCSEIHYDRIGGRDAAHLVNQDDP NVLEIWNLVFIQYNRES DGVLKPLPKKSIDTGMGLERLVS VLQNKMSNYD TDLFMPYFEAIQKGTGARPYTGKVGAE ADGIDMAYRVLADHARTITVALADGGRPDNTGRGYVLR RILRRAVRYSHEKLNASRGFFATLVDV VVQSLGD AFPELKKDPEMVKDIINEEEVQFLK	SEQ.ID. NO. 21

<b>Table 1D</b> <b>AlaRS1<sup>N5</sup></b> <b>Mass spec peptides detected and inferred linking peptides</b>		
<b>Type / species</b>	<b>Sequence</b>	<b>SEQ.ID . NO.</b>
Protein / mouse	MALELLTQEFGIPVER	SEQ.ID. NO. 22
Protein / mouse	LYVTYFGGDEAAGLEPDLECRQIWQNLGLDEAR	SEQ.ID. NO. 23
Protein / mouse	ILPGNMKDNFWEMGDTGPCGPCSEIHYDRIGGRDAAHLV NQDDPNVLEIWNLVFIQYNRES DGVLKPLPKKSIDTGMG LERLVSVLQNK	SEQ.ID. NO. 24
Protein / mouse	MSNYDTDLFMPYFEAIQK	SEQ.ID. NO. 25
Protein / mouse	GTGARPYTGKVGAEADADGIDMAYRVLADHAR	SEQ.ID. NO. 26
Protein / mouse	TITVALADGGRPDNTGR	SEQ.ID. NO. 27
Protein / mouse	GYVLRRILRRAVRYSHEKLNASRGFFATLVDVVVQSLGD AFPELKKDPEMVK	SEQ.ID. NO. 28
Protein / mouse	DIINEEEVQFLK	SEQ.ID. NO. 29
Protein / mouse	TLSRGRRILDRKIQSLGDCK	SEQ.ID. NO. 30
Protein / mouse	TIPGDTAWLLYDTYGFPVDLTGLIAEEK	SEQ.ID. NO. 31
Protein / mouse	GLVVDMMNGFEEERRLAQLKSQ GK	SEQ.ID. NO. 32
Protein / mouse	GAGDEDLIMLDIYAIEELR	SEQ.ID. NO. 33
Protein / mouse	AKGLEATDDSPKYNYQSDSSGSYVFECTVATVLALRREK	SEQ.ID. NO. 34
Protein / mouse	MFVDEVVTGQECGVVLDK	SEQ.ID. NO. 35
Protein / mouse	TCFYAEQGGQIYDEGYLVK	SEQ.ID. NO. 36
Protein / mouse	VDDSSSEDKTEFTVK	SEQ.ID. NO. 37
Protein / mouse	NAQVRGGYVLHIGTIYGNLKVGDQVRLFIDEPRRRPVMS NHTATHILNFALRSVLGEADQKGS LVAPDRLRFDF TAKG AMSTQQIK	SEQ.ID. NO. 38
Protein / mouse	KAEEIVNGMIEAAKPVYTQDCPLAAAK	SEQ.ID. NO. 39

<b>Table 1E</b> <b>AlaRS1<sup>N5</sup></b> <b>Concatenated sequences based on mass spec peptides detected</b>		
Type / species	Sequence	SEQ.ID . NO.
Protein / mouse	MALELLTQEF GIPVERLYVTYFGGDEAAGLEPDLECRQI WQNLGLDEARILPGNMKDNFWEMGDTGPCGPCSEIHYD RIGGRDAAHLVNQDDPNVLEIWNLVFIQYNRES DGVLKP LPKKSIDTGMGLERLVSVLQNKMSNYD TDLFMPYFEAIQ KGTGARPYTGKVGAEADAGIDMAYRVLADHARTITVAL ADGGRPDNTGRGYVLRRLRRAVRYSHEKLNASRGFFAT LVDVVVQSLGDAFPELKKDPEMVKDIINEEEVQFLKTL SR GRRILDRKIQSLGDCKTIPGDTAWLLYDTYGFV DLTGLI AEEKGLVVD MNGFEEERRLAQLKSQKGAGDEDLIMLDI YAIEELRAKGLEATDDSPKYN YQSDSSGSYVFECTVATV LALRREKMFVDEVVTGQECGVVLDKTCFYAEQGGQIYD EGYLVKVDDSSDKTEFTVKNAQVRGGYVLHIGTIYGNL KVG DQVRLFIDEPRRRPVMSNHTATHILNFALRSVLGEA DQKGSLVAPDRLRFDF TAKGAMSTQQIKKAEI VNGMIE AAKPVYTQDCPLAAAK	SEQ.ID. NO. 40

<b>Table 2</b> <b>N-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID . NO.
AlaRS1 <sup>N</sup> <sub>4</sub>	Protein / Human / 1-224+51	MDSTLTASEIRQRFIDFFKRNEHTYVHSSATIP LDDPTLLFANAGMNQFKPIFLNTIDPSHPMAK LSRAANTQKCIRAGGKHNDLDDVGKDVYHH TFFEMLGWSWFGDYFKELACKMALELLTQEF GIPIERLYVTYFGGDEAAGLEADLECKQIWQN LGLDDTKILPGNMKDNFWEMGDTGPCGPCSE IHYDRIGGRDAAHLVNQDDPNVLEIWNLVFI QYNRAQVPDHTLGKLVLRMPMGLTWPTGC WLTTLGPSLWHWLMVAGLTTQGVD MC	SEQ.ID. NO.41
AlaRS1 <sup>N</sup> <sub>4</sub>	DNA / Human /	ATGGACTCTACTCTAACAGCAAGTGAAATC CGGCAGCGATTTATAGATTTCTTCAAGAGG AACGAGCATACGTATGTTCACTCGTCTGCC ACCATCCCATTGGATGACCCCACTTTGCTCT TTGCCAATGCAGGCATGAACCAGTTTAAAC CCATTTTCCTGAACACAATTGACCCATCTCA CCCCATGGCAAAGCTGAGCAGAGCTGCCAA TACCCAGAAGTGCATCCGGGCTGGGGGCAA ACATAATGACCTGGACGATGTGGGCAAGGA TGTCTATCATCACACCTTCTTCGAGATGCTG GGCTCTTGGTCTTTTGGAGATTACTTTAAGG	SEQ.ID. NO.42

<b>Table 2</b> <b>N-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID . NO.</b>
		AATTGGCATGTAAGATGGCTCTGGA ACTCC TCACCCAAGAGTTTGGCATTCCCATTGAAA GACTTTATGTTACTTACTTTGGCGGGGATGA AGCAGCTGGCTTAGAAGCAGATCTGGAATG CAAACAGATCTGGCAAAATTTGGGGCTGGA TGACACCAAAATCCTCCCAGGCAACATGAA GGATAACTTCTGGGAGATGGGTGACACGGG CCCCTGTGGTCCTTGCAGTGAGATCCACTA CGACCGGATTGGTGGTCGGGACGCCGCACA TCTTGTCAACCAGGACGACCCTAATGTGCT GGAGATCTGGAACCTTGTGTTCATCCAGTA TAACAGGGCACAGGTGCCCCGACCATACACT GGGAAAGTTGGTGCTGAGGATGCCGATGGG ATTGACATGGCCTACCGGGTGCTGGCTGAC CACGCTCGGACCATCACTGTGGCACTGGCT GATGGTGGCCGGCCTGACAACACAGGGCGT GGATATGTGTTGA	
AlaRS1 <sup>N</sup> <sub>6</sub>	Protein / Human / 1-497+24	MDSTLTASEIRQRFIDFFKRNEHTYVHSSATIP LDDPTLLFANAGMNQFKPIFLNTIDPSHPMAK LSRAANTQKCIRAGGKHNDLDDVGKDVYHH TFFEMLGWSWFGDYFKELACKMALELLTQEF GIPIERLYVTYFGGDEAAGLEADLECKQIWQN LGLDDTKILPGNMKDNFWEMGDTGPCGPCSE IHYDRIGGRDAAHLVNQDDPNVLEIWNLVFI QYNREADGILKPLPKSIDTGMGLERLVSVLQ NKMSNYDITDLFVPYFEAIQKGTGARPYTGKV GAEDADGIDMAYRVLADHARTITVALADGG RPDNTGRGYVLRRLRRVRYAHEKLNASRG FFATLVDVVVQSLGDAFPELKKDPDMVKDII NEEEVQFLKTLRGRRLDRKIQSLGDSKTIPT DTAWLLYDTYGFPVDLTGLIAEEKGLVDM DGFEEERKLAQLKSQKGAGGEDLIMLDIYAI EELRARGLEVTDSPKYNHYLDSSGSYENRV YSEECSGPRRVCATHWNHLR	SEQ.ID. NO.43
AlaRS1 <sup>N</sup> <sub>6</sub>	DNA / Human /	ATGGACTCTACTCTAACAGCAAGTGAAATC CGGCAGCGATTTATAGATTTCTTCAAGAGG AACGAGCATACGTATGTTCACTCGTCTGCC ACCATCCCATTGGATGACCCCACTTTGCTCT TTGCCAATGCAGGCATGAACCAGTTTAAAC CCATTTTCCTGAACACAATTGACCCATCTCA CCCCATGGCAAAGCTGAGCAGAGCTGCCAA TACCCAGAAGTGCATCCGGGCTGGGGGCAA ACATAATGACCTGGACGATGTGGGCAAGGA	SEQ.ID. NO.44

**Table 2**  
**N-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing**

Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID . NO.
		TGTCTATCATCACACCTTCTTCGAGATGCTG GGCTCTTGGTCTTTTGGAGATTACTTTAAGG AATTGGCATGTAAGATGGCTCTGGAACCTCC TCACCCAAGAGTTTGGCATTCCCATTGAAA GACTTTATGTTACTTACTTTGGCGGGGATGA AGCAGCTGGCTTAGAAGCAGATCTGGAATG CAAACAGATCTGGCAAAATTTGGGGCTGGA TGACACCAAAATCCTCCCAGGCAACATGAA GGATAACTTCTGGGAGATGGGTGACACGGG CCCCTGTGGTCCTTGCAGTGAGATCCACTA CGACCGGATTGGTGGTCGGGACGCCGCACA TCTTGTCAACCAGGACGACCCTAATGTGCT GGAGATCTGGAACCTTGTGTTTCATCCAGTA TAACAGGGAAGCTGATGGCATTCTGAAACC TCTTCCCAAGAAAAGCATTGACACAGGGAT GGGCCTGGAACGACTGGTATCTGTGCTGCA GAATAAGATGTCCAACCTATGACACTGACCT TTTTGTCCCTTACTTTGAAGCCATTCAGAAG GGCACAGGTGCCCCGACCATACTGGGAAA GTTGGTGCTGAGGATGCCGATGGGATTGAC ATGGCCTACCGGGTGCTGGCTGACCACGCT CGGACCATCACTGTGGCACTGGCTGATGGT GGCCGGCCTGACAACACAGGGCGTGGATAT GTGTTGAGACGGATTCTCCGCCGAGCTGTC CGATACGCCCATGAAAAGCTCAATGCCAGC AGGGGCTTCTTTGCTACGTTAGTGGATGTTG TCGTCCAGTCCCTGGGAGATGCATTTCTTG AGCTGAAGAAGGACCCAGACATGGTGAAG GACATCATTAATGAAGAAGAGGTGCAGTTT CTCAAGACTCTCAGCAGAGGGCGTCGCATC CTGGACAGGAAAATTCAGAGCCTGGGAGA CAGCAAGACCATTCCCGGAGACACTGCTTG GCTCCTCTATGACACCTATGGGTTTCCAGTG GATCTGACTGGACTGATTGCTGAAGAGAAG GGCCTGGTGGTAGACATGGATGGCTTTGAA GAGGAGAGGAAACTGGCCCAGCTGAAATC ACAGGGCAAGGGAGCTGGTGGGGAAGACC TCATTATGCTGGACATTTACGCTATCGAAG AGCTCCGGGCACGGGGTCTGGAGGTCACAG ATGATTCCCCAAAGTACAATTACCATTTGG ACTCCAGTGGTAGCTATGAAAACAGAGTTT ACAGTGAAGAATGCTCAGGTCCGAGGAGG GTATGTGCTACACATTGGAACCATCTACGG	

<b>Table 2</b> <b>N-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID . NO.</b>
		TGA	
AlaRS1 <sup>N</sup> <sub>7</sub>	Protein / Human / 1-595+70	MDSTLTASEIRQRFIDFFKRNEHTYVHSSATIP LDDPTLLFANAGMNQFKPIFLNTIDPSHPMAK LSRAANTQKCIRAGGKHNDLDDVGKDVYHH TFFEMLGSWSF GDYFKELACKMALELLTQEF GIPIERLYVTYFGGDEAAGLEADLECKQIWQN LGLDDTKILPGNMKDNFWEMGDTGPCGPCSE IHYDRIGGRDAAHLVNQDDPNVLEIWNLVFI QYNREADGILKPLPKKSIDTGMGLERLVSVLQ NKMSNYDTDLFVPYFEAIQKGTGARPYTGKV GAEDADGIDMAYRVLADHARTITVALADGG RPDNTGRGYVLRRLRAVRYAHEKLNASRG FFATLVDVVVQSLGDAFPELKKDPDMVKDII NEEEVQFLKTL SRGRRILDRKIQSLGDSKTI PG DTAWLLYDTYGFPVDLTGLIAEEKGLVDM DGFEEERKLAQLKSQKGAGGEDLIMLDIYAI EELRARGLEVTD DSPKYNHLDSSGSYVFEN TVATVMALRREKMFVEEVSTGQECGVVLDK TCFYAEQGGQIYDEGYLVKVDDSSDKTEFT VKNAQVRGGYVLHIGTIYGDLKVGDQVWLF I DEAPAELESCRSFCDRDGRSHCQGYPEDCGC HRCRGPEGPQESRELEEMSLCHGSQSEGSDCS KQGCAEGDR	SEQ.ID. NO.45
AlaRS1 <sup>N</sup> <sub>7</sub>	DNA / Human /	ATGGACTCTACTCTAACAGCAAGTGAAATC CGGCAGCGATTTATAGATTTCTTCAAGAGG AACGAGCATACGTATGTTCACTCGTCTGCC ACCATCCCATTGGATGACCCCACTTTGCTCT TTGCCAATGCAGGCATGAACCAGTTTAAAC CCATTTTCCTGAACACAATTGACCCATCTCA CCCCATGGCAAAGCTGAGCAGAGCTGCCAA TACCCAGAAGTGCATCCGGGCTGGGGGCAA ACATAATGACCTGGACGATGTGGGCAAGGA TGTCTATCATCACACCTTCTTCGAGATGCTG GGCTCTTGGTCTTTTGGAGATTACTTTAAGG AATTGGCATGTAAGATGGCTCTGGAAC TCC TCACCCAAGAGTTTGGCATTCCCAT TGAAA GACTTTATGTTACTTACTTTGGCGGGGATGA AGCAGCTGGCTTAGAAGCAGATCTGGAATG CAAACAGATCTGGCAAAATTTGGGGCTGGA TGACACCAAAATCCTCCCAGGCAACATGAA GGATAACTTCTGGGAGATGGGTGACACGGG CCCCTGTGGTCCTTGCAGTGAGATCCACTA CGACCGGATTGGTGGTCGGGACGCCGCACA	SEQ.ID. NO.46

**Table 2**  
**N-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing**

Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID . NO.
		TCTTGTC AACCAGGACGACCCTAATGTGCT GGAGATCTGGAACCTTGTGTT CATCCAGTA TAACAGGGAAGCTGATGGCATTCTGAAACC TCTTCCCAAGAAAAGCATTGACACAGGGAT GGGCCTGGAACGACTGGTATCTGTGCTGCA GAATAAGATGTCCA ACTATGACACTGACCT TTTTGTCCCTTACTTTGAAGCCATT CAGAAG GGCACAGGTGCCCCGACCATACTGGGAAA GTTGGTGCTGAGGATGCCGATGGGATTGAC ATGGCCTACCGGGTGCTGGCTGACCACGCT CGGACCATCACTGTGGCACTGGCTGATGGT GGCCGGCCTGACAACACAGGGCGTGGATAT GTGTTGAGACGGATTCTCCGCCGAGCTGTC CGATACGCCCATGAAAAGCTCAATGCCAGC AGGGGCTTCTTTGCTACGTTAGTGGATGTTG TCGTCCAGTCCCTGGGAGATGCATTTCTCTG AGCTGAAGAAGGACCCAGACATGGTGAAG GACATCATTAATGAAGAAGAGGTGCAGTTT CTCAAGACTCTCAGCAGAGGGCGTCGCATC CTGGACAGGAAAATT CAGAGCCTGGGAGA CAGCAAGACCATTC CCGGAGACACTGCTTG GCTCCTCTATGACACCTATGGGTTTCCAGTG GATCTGACTGGACTGATTGCTGAAGAGAAG GGCCTGGTGGTAGACATGGATGGCTTTGAA GAGGAGAGGAAACTGGCCCCAGCTGAAATC ACAGGGCAAGGGAGCTGGTGGGGAAGACC TCATTATGCTGGACATTTACGCTATCGAAG AGCTCCGGGCACGGGGTCTGGAGGTCACAG ATGATTCCCCAAAGTACAATTACCATTTGG ACTCCAGTGGTAGCTATGTATTTGAGAACA CAGTGGCTACGGTGATGGCTCTGCGCAGGG AGAAGATGTTCTGTGGAAGAGGTGTCCACAG GCCAGGAGTGTGGAGTGGTGCTGGACAAG ACCTGTTTCTATGCTGAGCAAGGAGGCCAG ATCTATGACGAAGGCTACCTGGTGAAGGTG GATGACAGCAGTGAAGATAAAACAGAGTTT ACAGTGAAGAATGCTCAGGTCCGAGGAGG GTATGTGCTACACATTGGAACCATCTACGG TGACCTGAAAGTGGGGGATCAGGTCTGGCT GTTTATTGATGAGGCACCTGCGGA ACTCGA GTCATGCAGGAGCTTTTGTGATCGTGACGG AAGAAGCCATTGCCAAGGGTATCCGGAGG ATTGTGGCTGTCACAGGTGCCGAGGCCAG	

<b>Table 2</b> <b>N-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID . NO.</b>
		AAGGCCCTCAGGAAAGCAGAGAGCTTGAA GAAATGTCTCTCTGTTCATGGAAGCCAAAGT GAAGGCTCAGACTGCTCCAAACAAGGATGT GCAGAGGGGAGATCGCTGA	
AlaRS1 <sup>N</sup> <sub>8</sub>	Protein / Human / 1-321 + 6	MDSTLTASEIRQRFIDFFKRNEHTYVHSSATIP LDDPTLLFANAGMNQFKPIFLNTIDPSHPMAK LSRAANTQKCIRAGGKHNDLDDVGKDVYHH TFFEMLGWSWFGDYFKELACKMALELLTQEF GIPIERLYVTYFGGDEAAGLEADLECKQIWQN LGLDDTKILPGNMKDNFWEMGDTGPCGPCSE IHYDRIGGRDAAHLVNQDDPNVLEIWNLVFI QYNREADGILKPLPKKSIDTGMGLERLVSVLQ NKMSNYDITDLFVPYFEAIQKGTGARPYTGKV GAEDADGIDMAYRVLADHARTITVALADGG RPDNTGRGEMHFLS	SEQ.ID. NO.47
AlaRS1 <sup>N</sup> <sub>8</sub>	DNA / Human /	ATGGACTCTACTCTAACAGCAAGTGAAATC CGGCAGCGATTTATAGATTTCTTCAAGAGG AACGAGCATACTATGTTCACTCGTCTGCC ACCATCCCATTGGATGACCCCACTTTGCTCT TTGCCAATGCAGGCATGAACCAGTTTAAAC CCATTTTCCTGAACACAATTGACCCATCTCA CCCCATGGCAAAGCTGAGCAGAGCTGCCAA TACCCAGAAGTGCATCCGGGCTGGGGGCAA ACATAATGACCTGGACGATGTGGGCAAGGA TGTCTATCATCACACCTTCTTCGAGATGCTG GGCTCTTGGTCTTTTGGAGATTACTTTAAGG AATTGGCATGTAAGATGGCTCTGGAACCTC TCACCCAAGAGTTTGGCATTCCCATTGAAA GACTTTATGTTACTTACTTTGGCGGGGATGA AGCAGCTGGCTTAGAAGCAGATCTGGAATG CAAACAGATCTGGCAAAATTTGGGGCTGGA TGACACCAAAATCCTCCCAGGCAACATGAA GGATAACTTCTGGGAGATGGGTGACACGGG CCCCTGTGGTCTTGCAGTGAGATCCACTA CGACCGGATTGGTGGTCTGGGACGCCGCACA TCTTGTCAACCAGGACGACCTAATGTGCT GGAGATCTGGAACCTTGTGTTTCATCCAGTA TAACAGGGAAGCTGATGGCATTCTGAAACC TCTTCCCAAGAAAAGCATTGACACAGGGAT GGGCCTGGAACGACTGGTATCTGTGCTGCA GAATAAGATGTCCAACATGACACTGACCT TTTTGTCCCTTACTTTGAAGCCATTGAGAAG GGCACAGGTGCCCCGACCATACTGGGAAA	SEQ.ID. NO.48



<b>Table 2</b> <b>N-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID . NO.</b>
		GTTGGTGCTGAGGATGCCGATGGGATTGAC ATGGCCTACCGGGTGCTGGCTGACCACGCT CGGACCATCACTGTGGCACTGGCTGATGGT GGCCGGCCTGACAACACAGGGCGTGGGGA GATGCATTTCTGAGCTGA	

<b>Table 2B</b> <b>AARS polypeptides unique splice junctions</b>					
<b>Name</b>	<b>Type / species</b>	<b>Amino acid and Nucleic Acid Sequences in the vicinity of the unique splice junction</b>	<b>Human fetal brain</b>	<b>Human Adult brain</b>	<b>SEQ.ID. NO.</b>
A1-AS01	DNA / Human /	ACCTTGTGTTTCATCCAGTATA ACAG GGCACAGGTGCCCGAC CATACTG	2	0	SEQ.ID. NO.49
	Protein / Human /	LVFIQYNRAQVPDHTL			SEQ.ID. NO.50
A1-AS04	DNA / Human /	CATTTGGACTCCAGTGGTAGC TATG AAAACAGAGTTTACAG TGAAGAATG	650	86	SEQ.ID. NO.51
	Protein / Human /	HLDSSGSYENRVYSEE			SEQ.ID. NO.52
A1-AS06	DNA / Human /	TCAGGTCTGGCTGTTTATTGA TGAG GCACCTGCGGAACCTCG AGTCATGCA	0	15	SEQ.ID. NO.53
	Protein / Human /	QVWLFIDEAPAELESC			SEQ.ID. NO.54
A1-AS07	DNA / Human /	GCCGGCCTGACAACACAGGG CGTGG GGAGATGCATTTCT GAGCTGAAGA			SEQ.ID. NO.55
	Protein / Human /	RPDNTGRGEMHFLS			SEQ.ID. NO.56

<b>Table 3</b> <b>N-terminal AARS polypeptides and nucleic acids identified by Bioinformatics</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID. NO.</b>
AlaRS1 <sup>N</sup> <sub>2</sub>	Protein / Human / 1-286	MDSTLTASEIRQRFIDFFKRNEHTYVHSSATI PLDDPTLLFANAGMNQFKPIFLNTIDPSHPM AKLSRAANTQKCIRAGGKHNDLDDVGKDV	SEQ.ID. NO.57

Table 3 N-terminal AARS polypeptides and nucleic acids identified by Bioinformatics			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
		YHHTFFEMLGSWSF GDYFKELACKMALEL LTQEFGIPIERLYVTYFGGDEAAGLEADLEC KQIWQNLGLDDTKILPGNMKDNFWEMGDT GPCGPCSEIHYDRIGGRDAAHLVNQDDPNV LEIWNLVFIQYNREADGILKPLPKKSIDTGM GLERLVSVLQNKMSNYD TDLFVPYFEAIQK GTGARPYTGKVGAE	
AlaRS1 <sup>N</sup> 2	DNA / Human /	ATGGACTCTACTCTAACAGCAAGTGAAAT CCGGCAGCGATTTATAGATTTCTTCAAGA GGAACGAGCATACGTATGTTCACTCGTCT GCCACCATCCCATTGGATGACCCCACTTT GCTCTTTGCCAATGCAGGCATGAACCACT TTAAACCCATTTTCCTGAACACAATTGAC CCATCTCACCCCATGGCAAAGCTGAGCAG AGCTGCCAATACCCAGAAGTGCATCCGGG CTGGGGGCAAACATAATGACCTGGACGAT GTGGGCAAGGATGTCTATCATCACACCTT CTTCGAGATGCTGGGCTCTTGGTCTTTTGG AGATTACTTTAAGGAATTGGCATGTAAGA TGGCTCTGGAACCTCCTACCCAAGAGTTT GGCATTCCCATTGAAAGACTTTATGTTAC TTACTTTGGCGGGGATGAAGCAGCTGGCT TAGAAGCAGATCTGGAATGCAAACAGAT CTGGCAAAATTTGGGGCTGGATGACACCA AAATCCTCCCAGGCAACATGAAGGATAAC TTCTGGGAGATGGGTGACACGGGCCCTTG TGGTCCTTGCAGTGAGATCCACTACGACC GGATTGGTGGTCGGGACGCCGCACATCTT GTCAACCAGGACGACCCTAATGTGCTGGA GATCTGGAACCTTGTGTTTCATCCAGTATA ACAGGGAAGCTGATGGCATTCTGAAACCT CTTCCAAGAAAAGCATTGACACAGGGAT GGGCCTGGAACGACTGGTATCTGTGCTGC AGAATAAGATGTCCAACATGACACTGAC CTTTTGTCCCTTACTTTGAAGCCATTGAG AAGGGCACAGGTGCCCGACCATACTG GGAAAGTTGGTGCTGAG	SEQ.ID. NO.58
AlaRS1 <sup>N</sup> 3	Protein / Human / 1-488	MDSTLTASEIRQRFIDFFKRNEHTYVHSSATI PLDDPTLLFANAGMNQFKPIFLNTIDPSHPM AKLSRAANTQK CIRAGGKHNDLDDVGKDV YHHTFFEMLGSWSF GDYFKELACKMALEL LTQEFGIPIERLYVTYFGGDEAAGLEADLEC KQIWQNLGLDDTKILPGNMKDNFWEMGDT	SEQ.ID. NO.59

Table 3 N-terminal AARS polypeptides and nucleic acids identified by Bioinformatics			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
		GPCGPCSEIHYDRIGGRDAAHLVNQDDPNV LEIWNLVFIQYNREADGILKPLPKKSIDTGM GLERLVSVLQNKMSNYDSDLFVPYFEAIQK GTGARPYTGKVGAEADGIDMAYRVLADH ARTITVALADGGRPDNTGRGYVLRRLRA VRYAHEKLNASRGFFATLVDVQVQSLGDA FPELKKDPDMVKDIINEEEVQFLKTLRGR ILDRKIQSLGDSKTIPGDTAWLLYDTYGFV DLTGLIAEEKGLVVDMDGFEEERKLAQLKS QGKGAGGEDLIMLDIYAIEELRARGLEVTD DSPKYN	
AlaRS1 <sup>N</sup> <sub>3</sub>	DNA / Human /	ATGGACTCTACTCTAACAGCAAGTGAAAT CCGGCAGCGATTTATAGATTTCTTCAAGA GGAACGAGCATACTGATGTTCACTCGTCT GCCACCATCCCATTGGATGACCCCACTTT GCTCTTTGCCAATGCAGGCATGAACCAGT TTAAACCCATTTTCCTGAACACAATTGAC CCATCTCACCCCATGGCAAAGCTGAGCAG AGCTGCCAATACCCAGAAGTGCATCCGGG CTGGGGGCAAACATAATGACCTGGACGAT GTGGGCAAGGATGTCTATCATCACACCTT CTTCGAGATGCTGGGCTCTTGGTCTTTTG AGATTACTTTAAGGAATTGGCATGTAAGA TGGCTCTGGAACCTCACCACAAGAGTTT GGCATTCCCATTGAAAGACTTTATGTTAC TTACTTTGGCGGGGATGAAGCAGCTGGCT TAGAAGCAGATCTGGAATGCAAACAGAT CTGGCAAAATTTGGGGCTGGATGACACCA AAATCCTCCCAGGCAACATGAAGGATAAC TTCTGGGAGATGGGTGACACGGGCCCTG TGGTCCTTGCAGTGAGATCCACTACGACC GGATTGGTGGTCGGGACGCCGCACATCTT GTCAACCAGGACGACCCTAATGTGCTGGA GATCTGGAACCTTGTGTTTCATCCAGTATA ACAGGGAAGCTGATGGCATTCTGAAACCT CTTCCCAAGAAAAGCATTGACACAGGGAT GGGCCTGGAACGACTGGTATCTGTGCTGC AGAATAAGATGTCCAACATGACACTGAC CTTTTGTCCCTTACTTTGAAGCCATTCAG AAGGGCACAGGTGCCCGACCATACTG GGAAAGTTGGTGCTGAGGATGCCGATGG GATTGACATGGCCTACCGGGTGCTGGCTG ACCACGCTCGGACCATCACTGTGGCACTG	SEQ.ID. NO.60

Table 3 N-terminal AARS polypeptides and nucleic acids identified by Bioinformatics			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
		GCTGATGGTGGCCGGCCTGACAACACAGG GCGTGGATATGTGTTGAGACGGATTCTCC GCCGAGCTGTCCGATACGCCCATGAAAAG CTCAATGCCAGCAGGGGCTTCTTTGCTAC GTTAGTGGATGTTGTCGTCCAGTCCCTGG GAGATGCATTTCTGAGCTGAAGAAGGAC CCAGACATGGTGAAGGACATCATTAATGA AGAAGAGGTGCAGTTTCTCAAGACTCTCA GCAGAGGGCGTCGCATCCTGGACAGGAA AATTCAGAGCCTGGGAGACAGCAAGACC ATTCCCGGAGACACTGCTTGGCTCCTCTA TGACACCTATGGGTTTCCAGTGGATCTGA CTGGACTGATTGCTGAAGAGAAGGGCCTG GTGGTAGACATGGATGGCTTTGAAGAGGA GAGGAAACTGGCCCAGCTGAAATCACAG GGCAAGGGAGCTGGTGGGGAAGACCTCA TTATGCTGGACATTTACGCTATCGAAGAG CTCCGGGCACGGGGTCTGGAGGTCACAGA TGATTCCCCAAAGTACAAT	

## C-terminal AARS Polypeptides: (Tables 4, 5 &amp; 6)

Table 4A C-terminal AARS polypeptides identified by MS			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
AlaRS1 <sup>CI</sup>	Protein / Human / 476-968	RGLEVTDDSPKYNHYHLDSSGSYVFENTVAT VMALRREKMFVEEVSTGQECGVVLDKTCF YAEQGGQIYDEGYLVKVDDSSSEDKTEFTVK NAQVRGGYVLHIGTIYGDLKVG DQVWLFID EPRRRPIMSNHTATHILNFALRSVLGEADQK GSLVAPDRLRFDF TAKGAMSTQQIKKAE EI ANEMIEAAKAVYTQDCPLAAAKAIQGLRA VFDETYPDPVRVVSIGVPVSELLDDPSGPAG SLTSVEFCGGTHLRNSSHAGAFVIVTEEAIA KGIRRIVAVTGAE AQKALRKAESLKKCLSV MEAKVKAQTAPNKDVQREIADLGEALATA VIPQWQKDELRETLKSLKKVMDDLDRASK ADVQKRVLEKTKQFIDSNPNQPLVILEMES GASAKALNEALKL FKMHSPT SAMLFTVD NEAGKITCLCQVPQNAANRGLKASEWVQQ VSGLMDGKGGGKDVSAQATGKNVGCLQE	SEQ.ID. NO.79

Table 4A C-terminal AARS polypeptides identified by MS			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
		ALQLATSFAQLRLGDVKN	
AlaRS1 <sup>CI</sup>	DNA / Human	CGGGGTCTGGAGGTCACAGATGATTCCCC AAAGTACAATTACCATTTGGACTCCAGTG GTAGCTATGTATTTGAGAACACAGTGGCT ACGGTGATGGCTCTGCGCAGGGAGAAGA TGTTTCGTGGAAGAGGTGTCCACAGGCCAG GAGTGTGGAGTGGTGCTGGACAAGACCTG TTTCTATGCTGAGCAAGGAGGCCAGATCT ATGACGAAGGCTACCTGGTGAAGGTGGAT GACAGCAGTGAAGATAAAACAGAGTTTA CAGTGAAGAATGCTCAGGTCCGAGGAGG GTATGTGCTACACATTGGAACCATCTACG GTGACCTGAAAGTGGGGGATCAGGTCTGG CTGTTTATTGATGAGCCCCGACGAAGACC CATCATGAGCAACCACACAGCTACGCACA TTCTGAACTTCGCCCTGCGCTCAGTGCTTG GGGAAGCTGACCAGAAAGGCTCATTGGTT GCTCCTGACCGCCTCAGATTTGACTTTACT GCCAAGGGAGCCATGTCCACCCAACAGAT CAAGAAGGCTGAAGAGATTGCTAATGAG ATGATTGAGGCAGCCAAGGCCGTCTATAC CCAGGATTGCCCCCTGGCAGCAGCGAAAG CCATCCAGGGCCTACGGGCTGTGTTTGAT GAGACCTATCCTGACCCTGTGCGAGTCGT CTCCATTGGGGTCCCGGTGTCCGAGTTGC TGGATGACCCCTCTGGGCCTGCTGGCTCC CTGACTTCTGTTGAGTTCTGTGGGGGAAC GCACCTGCGGAACTCGAGTCATGCAGGAG CTTTTGTGATCGTGACGGAAGAAGCCATT GCCAAGGGTATCCGGAGGATTGTGGCTGT CACAGGTGCCGAGGCCCAAGGCCCTC AGGAAAGCAGAGAGCTTGAAGAAATGTC TCTCTGTCATGGAAGCCAAAGTGAAGGCT CAGACTGCTCCAAACAAGGATGTGCAGA GGGAGATCGCTGACCTTGGAGAGGCCCTG GCCACTGCAGTCATCCCCCAGTGGCAGAA GGATGAATTGCGGGAGACTCTCAAATCCC TAAAGAAGGTCATGGATGACTTGGACCGA GCCAGCAAAGCCGATGTCCAGAAACGAG TGTTAGAGAAGACGAAGCAGTTCATCGAC AGCAACCCCAACCAGCCTCTTGTCATCCT GGAGATGGAGAGCGGCGCCTCAGCCAAG GCCCTGAATGAAGCCTTGAAGCTCTTCAA GATGCACTCCCCTCAGACTTCTGCCATGC	SEQ.ID. NO.80

<b>Table 4A</b> <b>C-terminal AARS polypeptides identified by MS</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID. NO.</b>
		TCTTCACGGTGGACAATGAGGCTGGCAAG ATCACGTGCCTGTGTCAAGTCCCCAGAA TGCAGCCAATCGGGGCTTAAAAGCCAGCG AGTGGGTGCAGCAGGTGTCAGGCTTGATG GACGGTAAAGGTGGTGGCAAGGATGTGT CTGCACAGGCCACAGGCAAGAACGTTGG CTGCCTGCAGGAGGCGCTGCAGCTGGCCA CTTCCTTCGCCCAGCTGCGCCTCGGGGAT GTAAAGAACTGA	
AlaRS1 <sup>C2</sup>	Protein / Human / 758-968	AEAQKALRKAESLKKCLSVMEAKVKAQTA PNKDVQREIADLGEALATAVIPQWQKDEL ETLKSLKKVMDDLDRASKADVQKRVLEKT KQFIDSNPNQPLVILEMESGASAKALNEALK LFKMHSPTQTSAMLFTVDNEAGKITCLCQVP QNAANRGLKASEWVQQVSGLMDGKGGGK DVSAQATGKNVGCLQEALQLATSFAQLRL GDVKN	SEQ.ID. NO.81
AlaRS1 <sup>C2</sup>	DNA / Human	GCCGAGGCCCAGAAGGCCCTCAGGAAAG CAGAGAGCTTGAAGAAATGTCTCTGTGTC ATGGAAGCCAAAGTGAAGGCTCAGACTG CTCCAAACAAGGATGTGCAGAGGGAGAT CGCTGACCTTGGAGAGGCCCTGGCCACTG CAGTCATCCCCAGTGGCAGAAGGATGAA TTGCGGGAGACTCTCAAATCCCTAAAGAA GGTCATGGATGACTTGGACCGAGCCAGCA AAGCCGATGTCCAGAAACGAGTGTTAGA GAAGACGAAGCAGTTCATCGACAGCAAC CCCAACCAGCCTCTTGTATCCTGGAGAT GGAGAGCGGCGCCTCAGCCAAGGCCCTG AATGAAGCCTTGAAGCTCTTCAAGATGCA CTCCCCTCAGACTTCTGCCATGCTCTTCAC GGTGGACAATGAGGCTGGCAAGATCACG TGCCTGTGTCAAGTTCCCCAGAATGCAGC CAATCGGGGCTTAAAAGCCAGCGAGTGG GTGCAGCAGGTGTCAGGCTTGATGGACGG TAAAGGTGGTGGCAAGGATGTGTCTGCAC AGGCCACAGGCAAGAACGTTGGCTGCCTG CAGGAGGCGCTGCAGCTGGCCACTTCCTT CGCCCAGCTGCGCCTCGGGGATGTAAAGA ACTGA	SEQ.ID. NO.82
AlaRS1 <sup>C3</sup>	Protein / Human / 588-968	QVWLFIDEPRRRPIMSNHTATHILNFALRSV LGEADQKGSLVAPDRLRFDF TAKGAMSTQ QIKKAEEIANEMIEAAKAVYTQDCPLAAAK AIQGLRAVFDETY PDPVRVVSIGVPVSELLD	SEQ.ID. NO.83

<b>Table 4A</b> <b>C-terminal AARS polypeptides identified by MS</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID. NO.</b>
		DPSGPAGSLTSVEFCGGTHLRNSSHAGAFVI VTEEAIAKGIRRIVAVTGAEAQKALRKAESL KKCLSVMEAKVKAQTAPNKDVQREIADLG EALATAVIPQWQKDELRETLKSLKKVMDD LDRASKADVQKRVLEKTKQFIDSNPNQPLV ILEMESGASAKALNEALKLFKMHSPTSAM LFTVDNEAGKITCLCQVPQNAANRGLKASE WVQQVSGLMDGKGGGKDVSAQATGKNVG CLQEALQLATSFAQLRLGDVKN	
AlaRS1 <sup>C3</sup>	DNA / Human	CAGGTCTGGCTGTTTATTGATGAGCCCCG ACGAAGACCCATCATGAGCAACCACACA GCTACGCACATTCTGAACTTCGCCCTGCG CTCAGTGCTTGGGGAAGCTGACCAGAAAG GCTCATTGGTTGCTCCTGACCGCCTCAGA TTTGACTTTACTGCCAAGGGAGCCATGTC CACCCAACAGATCAAGAAGGCTGAAGAG ATTGCTAATGAGATGATTGAGGCAGCCAA GGCCGTCTATACCCAGGATTGCCCCCTGG CAGCAGCGAAAGCCATCCAGGGCCTACG GGCTGTGTTTGATGAGACCTATCCTGACC CTGTGCGAGTCGTCTCCATTGGGGTCCCG GTGTCCGAGTTGCTGGATGACCCCTCTGG GCCTGCTGGCTCCCTGACTTCTGTTGAGTT CTGTGGGGGAACGCACCTGCGGAACTCGA GTCATGCAGGAGCTTTTGTGATCGTGACG GAAGAAGCCATTGCCAAGGGTATCCGGA GGATTGTGGCTGTCACAGGTGCCGAGGCC CAGAAGGCCCTCAGGAAAGCAGAGAGCT TGAAGAAATGTCTCTCTGTCATGGAAGCC AAAGTGAAGGCTCAGACTGCTCCAAACA AGGATGTGCAGAGGGAGATCGCTGACCTT GGAGAGGCCCTGGCCACTGCAGTCATCCC CCAGTGGCAGAAGGATGAATTGCGGGAG ACTCTCAAATCCCTAAAGAAGGTCATGGA TGACTTGGACCGAGCCAGCAAAGCCGATG TCCAGAAACGAGTGTTAGAGAAGACGAA GCAGTTCATCGACAGCAACCCCAACCAGC CTCTTGTCATCCTGGAGATGGAGAGCGGC GCCTCAGCCAAGGCCCTGAATGAAGCCTT GAAGCTCTTCAAGATGCACTCCCCTCAGA CTTCTGCCATGCTCTTCACGGTGGACAAT GAGGCTGGCAAGATCACGTGCCTGTGTCA AGTTCCCCAGAATGCAGCCAATCGGGGCT TAAAAGCCAGCGAGTGGGTGCAGCAGGT	SEQ.ID. NO.84

<b>Table 4A</b> <b>C-terminal AARS polypeptides identified by MS</b>			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
		GTCAGGCTTGATGGACGGTAAAGGTGGTG GCAAGGATGTGTCTGCACAGGCCACAGGC AAGAACGTTGGCTGCCTGCAGGAGGCGCT GCAGCTGGCCACTTCCTTCGCCCAGCTGC GCCTCGGGGATGTAAAGAACTGA	

<b>Table 4B</b> <b>AlaRS1<sup>C1</sup></b> <b>Mass spec peptides detected and inferred linking peptides</b>		
Type / species	Sequence	SEQ.ID. NO.
Protein / mouse	TCFYAEQGGQIYDEGYLVK	SEQ.ID. NO.85
Protein / mouse	VDDSSDKTEFTVKNAQVRGGYVLHIGTIYGNLKVGDQ VRLFIDEPRRRPVMSNHTATHILNFALRSVLGEADQKGS LVAPDRLRFDFATAKGAMSTQQIKKAAEEIVNGMIEAAKPV YTQDCPLAAAKAIQGLRAVFDETYPDVVRVVSIGVPVSE LLDDPCGPAGSLTSVEFCGGTHLRNSSHAGAFVIVTEEAI AKGIRRVAVTGAEAQKALRKSETLKKSLSAMEAKVKA QTAPNKDVQR	SEQ.ID. NO.86
Protein / mouse	EIADLGEALATAVIPQWQK	SEQ.ID. NO.87
Protein / mouse	MFVDEVVTGQECGVVLDK	SEQ.ID. NO.88
Protein / mouse	TCFYAEQGGQIYDEGYLVK	SEQ.ID. NO.89
Protein / mouse	VDDSSDKTEFTVK	SEQ.ID. NO.90
Protein / mouse	YNYQSDSSGSYVFECTVATVLALR	SEQ.ID. NO.91
Protein / mouse	REKMFVDEVVTGQECGVVLDK	SEQ.ID. NO.92
Protein / mouse	TCFYAEQGGQIYDEGYLVK	SEQ.ID. NO.93
Protein / mouse	VDDSSDKTEFTVKNAQVRGGYVLHIGTIYGNLKVGDQ VRLFIDEPRRRPVMSNHTATHILNFALRSVLGEADQKGS LVAPDRLRFDFATAKGAMSTQQIKKAAEEIVNGMIEAAKPV YTQDCPLAAAKAIQGLRAVFDETYPDVVRVVSIGVPVSE LLDDPCGPAGSLTSVEFCGGTHLRNSSHAGAFVIVTEEAI AKGIRRVAVTGAEAQKALRKSETLKKSLSAMEAKVKA QTAPNKDVQREIADLGEALATAVIPQWQKDEQRETLKS LKKVMDDLDRASKADVQKRVLEKTKQLIDSNPNQPLVI	SEQ.ID. NO.94



<b>Table 4B</b> <b>AlaRS1<sup>C1</sup></b> <b>Mass spec peptides detected and inferred linking peptides</b>		
Type / species	Sequence	SEQ.ID. NO.
	LEMESGASAKALNEALKLKFTHSPQTSAMLFTVDNEAG KITCLCQVPQNAANRGLK	
Protein / mouse	ASEWVQQVSGGLMDGK	SEQ.ID. NO.95
Protein / mouse	VDDSSDKTEFTVK	SEQ.ID. NO.96
Protein / mouse	NAQVRGGYVLHIGTIYGNLKVGDQVRLFIDEPRRRPVMS NHTATHILNFALRSVLGEADQKGSLVAPDRLRFDFATAKG AMSTQQIKKAAEIVNGMIEAAKPVYTQDCPLAAAKAIQG LRAVFDETYPDVVRVVSIGVPVSELLDDPCGPAGSLTSVE FCGGTHLRNSSHAGAFVIVTEEAIAKGIRRVAVTGAEAAQ KALRKSETLKKSLSAMEAKVKAQTAPNKDVQR	SEQ.ID. NO.97
Protein / mouse	EIADLGEALATAVIPQWQK	SEQ.ID. NO.98
Protein / mouse	DEQRETLKSLKKVMDLDRASKADVQKRVLEKTKQLID SNPNQPLVILEMESGASAKALNEALKLKF	SEQ.ID. NO.99
Protein / mouse	THSPQTSAMLFTVDNEAGK	SEQ.ID. NO.100
Protein / mouse	ITCLCQVPQNAANR	SEQ.ID. NO.101
Protein / mouse	GLK	SEQ.ID. NO.102
Protein / mouse	ASEWVQQVSGGLMDGK	SEQ.ID. NO.103

<b>Table 4C</b> <b>AlaRS1<sup>C1</sup></b> <b>Concatenated sequences based on mass spec peptides detected</b>		
Type / species	Sequence	SEQ.ID. NO.
Protein / mouse	<u>TCFYAEQGGQIYDEGYLVK</u> VDDSSDKTEFTVKNAQVR GGYVLHIGTIYGNLKVGDQVRLFIDEPRRRPVMSNHTAT HILNFALRSVLGEADQKGSLVAPDRLRFDFATAKGAMSTQ QIKKAAEIVNGMIEAAKPVYTQDCPLAAAKAIQGLRAVF DETYPDVVRVVSIGVPVSELLDDPCGPAGSLTSVEFCGGT HLRNSSHAGAFVIVTEEAIAKGIRRVAVTGAEAAQKALR KSETLKKSLSAMEAKVKAQTAPNKDVQREIADLGEALA TAVIPQWQK	SEQ.ID. NO.104
Protein / mouse	<u>MFVDEVVTGQECGVVLDK</u> TCFYAEQGGQIYDEGYLVK <u>VDDSSDKTEFTVK</u>	SEQ.ID. NO.105
Protein / mouse	<u>YNYQSDSSGSYVFECTVATVLALRREKMFVDEVVTGQE</u> <u>CGVVLDKTCFYAEQGGQIYDEGYLVK</u> VDDSSDKTEFT	SEQ.ID. NO.106

<b>Table 4C</b> <b>AlaRS1<sup>C1</sup></b> <b>Concatenated sequences based on mass spec peptides detected</b>		
Type / species	Sequence	SEQ.ID. NO.
	VKNAQVRGGYVLHIGTIYGNLKVGDQVRLFIDEPRRRPV MSNHTATHILNFALRSVLGEADQKGS LVAPDRLRFDFTA KGAMSTQQIKKAEIIVNGMIEAAKPVYTQDCPLAAAKAI QGLRAVFDETYDPVVRVVSIGVPVSELLDDPCGPAGSLT SVEFCGGTHLRNSSHAGAFVIVTEEAIAKGIRRIVAVTGA EAQKALRKSETLKKSLSAMEAKVKAQTAPNKDVQREIA DLGEALATAVIPQWQKDEQRETLKSLKKVMDDLDRASK ADVQKRVLEKTKQLIDSNPNQPLVILEMESGASAKALNE ALKLFKTHSPQTSAMLFTVDNEAGKITCLCQVPQNAAN RGLKASEWVQQVSGLMDGK	
Protein / mouse	<u>VDDSS</u> EDKTEFTVKNAQVRGGYVLHIGTIYGNLKVGDQ VRLFIDEPRRRPVMSNHTATHILNFALRSVLGEADQKGS LVAPDRLRFDFTA KGAMSTQQIKKAEIIVNGMIEAAKPV YTQDCPLAAAKAIQGLRAVFDETYDPVVRVVSIGVPVSE LLDDPCGPAGSLTSVEFCGGTHLRNSSHAGAFVIVTEEAIA AKGIRRIVAVTGAEAQKALRKSETLKKSLSAMEAKVKA QTAPNKDVQREIADLGEALATAVIPQWQKDEQRETLKS LKKVMDDLDRASKADVQKRVLEKTKQLIDSNPNQPLVI LEMESGASAKALNEALKLFKTHSPQTSAMLFTVDNEAG <u>KITCLCQVPQNAANRGLKASEWVQQVSGLMDGK</u>	SEQ.ID. NO.107

<b>Table 4D</b> <b>AlaRS1<sup>C2</sup></b> <b>Mass spec peptides detected and inferred linking peptides</b>		
Type / species	Sequence	SEQ.ID. NO.
Protein / mouse	DMSAQATGKNVGCLQEALQLATSFAQLR	SEQ.ID. NO.108
Protein / mouse	DMSAQATGK	SEQ.ID. NO.109
Protein / mouse	NVGCLQEALQLATSFAQLR	SEQ.ID. NO.110

<b>Table 4E</b> <b>AlaRS1<sup>C2</sup></b> <b>Concatenated sequences based on mass spec peptides detected</b>		
Type / species	Sequence	SEQ.ID. NO.
Protein / mouse	<u>MFVDEVVTGQECGVVLDK</u> TCFYAEQGGQIYDEGYLVK <u>VDDSS</u> EDKTEFTVK	SEQ.ID. NO.111

<b>Table 4F</b> <b>AlaRS1<sup>C3</sup></b> <b>Mass spec peptides detected and inferred linking peptides</b>		
Type / species	Sequence	SEQ.ID. NO.
Protein / mouse	EIADLGEALATAVIPQWQK	SEQ.ID. NO.112
Protein / mouse	DEQRETLKSLKKVMDDLDRASKADVQKRVLEKTKQLID SNPNQPLVILEMESGASAKALNEALKLFKTHSPQTSAML FTVDNEAGK	SEQ.ID. NO.113
Protein / mouse	ITCLCQVPQNAANR	SEQ.ID. NO.114

<b>Table 4G</b> <b>AlaRS1<sup>C3</sup></b> <b>Concatenated sequences based on mass spec peptides detected</b>		
Type / species	Sequence	SEQ.ID. NO.
Protein / mouse	<u>EIADLGEALATAVIPQWQK</u> DEQRETLKSLKKVMDDLDR ASKADVQKRVLEKTKQLIDSNPNQPLVILEMESGASAKA LNEALKLFKTHSPQTSAMLFTVDNEAGK <u>ITCLCQVPQNA</u> <u>ANR</u>	SEQ.ID. NO.115

<b>Table 5A</b> <b>C-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
AlaRS1 <sup>C8</sup>	Protein / Human / 293-968	MAYRVLADHARTITVALADGGRPDNTGRG YVLRRILRRAVRYAHEKLNASRGFFATLVD VVVQSLGDAFPELKKDPDMVKDIINEEEVQ FLKTLRGRRIIDRDKIQLGDSKTIPGDTAW LLYDITYGFPVDLTGLIAEEKGLVVDMDGFE EERKLAQLKSQKGAGGEDLIMLDIYAIEEL RARGLEVTDSPKYNHYHLDSSGSYVFENTV ATVMALRREKMFVEEVSTGQECGVVLDKT CFYAEQGGQIYDEGYLVKVDDSSSEDKTEFT VKNAQVRGGYVLHIGTIYGDLKVG DQVWL FIDEPRRRPIMSNHTATHILNFALRSVLGEA DQKGSVLAPDRLRFDFTAKGAMSTQQIKKA EEIANEMIEAAKAVYTQDCPLAAAKAIQGL RAVFDETYDPVVRVVSIGVPVSELDDPSGP AGSLTSVEFCGGTHLRNSSHAGAFVIVTEEA	SEQ.ID. NO.116

<b>Table 5A</b> <b>C-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID. NO.</b>
		IAKGIRRIVAVTGAEAQKALRKAESLKKCLS VMEAKVKAQTAPNKDVQREIADLGEALAT AVIPQWQKDELRETLKSLKKVMDLDRAS KADVQKRVLEKTKQFIDSNPNQPLVILEME SGASAKALNEALKLFLKMHSPQTSAMLFTVD NEAGKITCLCQVPQNAANRGLKASEWVQQ VSLMDGKGGGKDVSAQATGKNVGCLQE ALQLATSFAQLRLGDVKN	
AlaRS1 <sup>C8</sup>	DNA / Human /	ATGGCCTACCGGGTGCTGGCTGACCACGC TCGGACCATCACTGTGGCACTGGCTGATG GTGGCCGGCCTGACAACACAGGGCGTGG ATATGTGTTGAGACGGATTCTCCGCCGAG CTGTCCGATACGCCCATGAAAAGCTCAAT GCCAGCAGGGGCTTCTTTGCTACGTTAGT GGATGTTGTCGTCCAGTCCCTGGGAGATG CATTCCTGAGCTGAAGAAGGACCCAGAC ATGGTGAAGGACATCATTAATGAAGAAG AGGTGCAGTTTCTCAAGACTCTCAGCAGA GGGCGTCGCATCCTGGACAGGAAAATTCA GAGCCTGGGAGACAGCAAGACCATTCCC GGAGACACTGCTTGGCTCCTCTATGACAC CTATGGGTTTCCAGTGGATCTGACTGGAC TGATTGCTGAAGAGAAGGGCCTGGTGGTA GACATGGATGGCTTTGAAGAGGAGAGGA AACTGGCCCAGCTGAAATCACAGGGCAA GGGAGCTGGTGGGGAAGACCTCATTATGC TGGACATTTACGCTATCGAAGAGCTCCGG GCACGGGGTCTGGAGGTCACAGATGATTC CCCAAAGTACAATTACCATTTGGACTCCA GTGGTAGCTATGTATTTGAGAACACAGTG GCTACGGTGATGGCTCTGCGCAGGGAGAA GATGTTTCGTGGAAGAGGTGTCCACAGGCC AGGAGTGTGGAGTGGTGCTGGACAAGAC CTGTTTCTATGCTGAGCAAGGAGGCCAGA TCTATGACGAAGGCTACCTGGTGAAGGTG GATGACAGCAGTGAAGATAAAACAGAGT TTACAGTGAAGAATGCTCAGGTCCGAGGA GGGTATGTGCTACACATTGGAACCATCTA CGGTGACCTGAAAGTGGGGGATCAGGTCT GGCTGTTTATTGATGAGCCCCGACGAAGA CCCATCATGAGCAACCACACAGCTACGCA CATTCTGAACTTCGCCCTGCGCTCAGTGCT TGGGGAAGCTGACCAGAAAGGCTCATTG	SEQ.ID. NO.117

<b>Table 5A</b> <b>C-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID. NO.</b>
		GTTGCTCCTGACCGCCTCAGATTTGACTTT ACTGCCAAGGGAGCCATGTCCACCCAACA GATCAAGAAGGCTGAAGAGATTGCTAAT GAGATGATTGAGGCAGCCAAGGCCGTCTA TACCCAGGATTGCCCCCTGGCAGCAGCGA AAGCCATCCAGGGCCTACGGGCTGTGTTT GATGAGACCTATCCTGACCCTGTGCGAGT CGTCTCCATTGGGGTCCCGGTGTCCGAGT TGCTGGATGACCCCTCTGGGCCTGCTGGC TCCCTGACTTCTGTTGAGTTCTGTGGGGG AACGCACCTGCGGAACTCGAGTCATGCAG GAGCTTTTGTGATCGTGACGGAAGAAGCC ATTGCCAAGGGTATCCGGAGGATTGTGGC TGTCACAGGTGCCGAGGCCCCAGAAGGCC TCAGGAAAGCAGAGAGCTTGAAGAAATG TCTCTCTGTCATGGAAGCCAAAGTGAAGG CTCAGACTGCTCCAAACAAGGATGTGCAG AGGGAGATCGCTGACCTTGGAGAGGCCCT GGCCACTGCAGTCATCCCCCAGTGGCAGA AGGATGAATTGCGGGAGACTCTCAAATCC CTAAAGAAGGTCATGGATGACTTGGACCG AGCCAGCAAAGCCGATGTCCAGAAACGA GTGTTAGAGAAGACGAAGCAGTTCATCGA CAGCAACCCCAACCAGCCTCTTGTCATCC TGGAGATGGAGAGCGGCGCCTCAGCCAA GGCCCTGAATGAAGCCTTGAAGCTCTTCA AGATGCACTCCCCTCAGACTTCTGCCATG CTCTTACGGTGGACAATGAGGCTGGCAA GATCACGTGCCTGTGTCAAGTTCCCCAGA ATGCAGCCAATCGGGGCTTAAAAGCCAGC GAGTGGGTGCAGCAGGTGTCAGGCTTGAT GGACGGTAAAGGTGGTGGCAAGGATGTG TCTGCACAGGCCACAGGCAAGAACGTTGG CTGCCTGCAGGAGGCGCTGCAGCTGGCCA CTTCTTCGCCCAGCTGCGCCTCGGGGAT GTAAAGAAGTGA	
AlaRS1 <sup>C9</sup>	Protein / Human / 1-48 + 450-968	MDSTLTASEIRQRFIDFFKRNEHTYVHSSATI PLDDPTLLFANAGMNQLKSQKGAGGEDLI MLDIYAIEELRARGLEVTDSPKYNHYHLS SGSYVFENTVATVMALRREKMFVEEVSTG QECGVVLDKTCFYAEQGGQIYDEGYLVKV DDSSDKTEFTVKNAQVRGGYVLHIGTIYG DLKVGDDQVWLFIDEPRRRPIMSNHTATHIL	SEQ.ID. NO.118

<b>Table 5A</b> <b>C-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
		NFALRSVLGEADQKGSLVAPDRLRFDF TAK GAMSTQQIKKAEEIANEMIEAAKAVYTQDC PLAAAKAIQGLRAVFDETYDPVRVVSIGVP VSELLDDPSGPAGSLTSVEFCGGTHLRNSSH AGAFVIVTEEAIAKGIRRIVAVTGAEAQKAL RKAESLKKCLSVMEAKVKAQTAPNKDVQR EIADLGEALATAVIPQWQKDELRETLKSLK KVMDDLDRASKADVQKRVLEKTKQFIDSN PNQPLVILEMESGASAKALNEALKLFKMHS PQTSAMLFTVDNEAGKITCLCQVPQNAANR GLKASEWVQQVSGLMDGKGGGKDVSAQA TGKNVGCLQEALQLATSFAQLRLGDVKN	
AlaRS1 <sup>C9</sup>	DNA / Human /	ATGGACTCTACTCTAACAGCAAGTGAAAT CCGGCAGCGATTTATAGATTTCTTCAAGA GGAACGAGCATACGTATGTTCACTCGTCT GCCACCATCCCATTGGATGACCCCACTTT GCTCTTTGCCAATGCAGGCATGAACCAGC TGAAATCACAGGGCAAGGGAGCTGGTGG GGAAGACCTCATTATGCTGGACATTTACG CTATCGAAGAGCTCCGGGCACGGGGTCTG GAGGTACAGATGATTCCCCAAAGTACAA TTACCATTTGGACTCCAGTGGTAGCTATG TATTTGAGAACACAGTGGCTACGGTGATG GCTCTGCGCAGGGAGAAGATGTTCGTGGA AGAGGTGTCCACAGGCCAGGAGTGTGGA GTGGTGCTGGACAAGACCTGTTTCTATGC TGAGCAAGGAGGCCAGATCTATGACGAA GGCTACCTGGTGAAGGTGGATGACAGCA GTGAAGATAAAACAGAGTTTACAGTGAA GAATGCTCAGGTCCGAGGAGGGTATGTGC TACACATTGGAACCATCTACGGTGACCTG AAAGTGGGGGATCAGGTCTGGCTGTTTAT TGATGAGCCCCGACGAAGACCCATCATGA GCAACCACACAGCTACGCACATTCTGAAC TTCGCCCTGCGCTCAGTGCTTGGGGAAGC TGACCAGAAAGGCTCATTTGGTTGCTCCTG ACCGCCTCAGATTTGACTTTACTGCCAAG GGAGCCATGTCCACCCAACAGATCAAGA AGGCTGAAGAGATTGCTAATGAGATGATT GAGGCAGCCAAGGCCGTCTATACCCAGG ATTGCCCCCTGGCAGCAGCGAAAGCCATC CAGGGCCTACGGGCTGTGTTTGATGAGAC CTATCCTGACCCTGTGCGAGTCGTCTCCAT	SEQ.ID. NO.119

<b>Table 5A</b> <b>C-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID. NO.</b>
		TGGGGTCCCCGGTGTCCGAGTTGCTGGATG ACCCCTCTGGGCCTGCTGGCTCCCTGACTT CTGTTGAGTTCTGTGGGGGAACGCACCTG CGGAACCTCGAGTCATGCAGGAGCTTTTGT GATCGTGACGGAAGAAGCCATTGCCAAG GGTATCCGGAGGATTGTGGCTGTCACAGG TGCCGAGGCCCAGAAGGCCCTCAGGAAA GCAGAGAGCTTGAAGAAATGTCTCTCTGT CATGGAAGCCAAAGTGAAGGCTCAGACT GCTCCAAACAAGGATGTGCAGAGGGAGA TCGCTGACCTTGGAGAGGCCCTGGCCACT GCAGTCATCCCCCAGTGGCAGAAGGATGA ATTGCGGGAGACTCTCAAATCCCTAAAGA AGGTCATGGATGACTTGGACCGAGCCAGC AAAGCCGATGTCCAGAAACGAGTGTTAG AGAAGACGAAGCAGTTCATCGACAGCAA CCCCAACCAGCCTCTTGTTCATCCTGGAGA TGGAGAGCGGCGCCTCAGCCAAGGCCCTG AATGAAGCCTTGAAGCTCTTCAAGATGCA CTCCCCCTCAGACTTCTGCCATGCTCTTCAC GGTGGACAATGAGGCTGGCAAGATCACG TGCCTGTGTCAAGTTCCCCAGAATGCAGC CAATCGGGGCTTAAAAGCCAGCGAGTGG GTGCAGCAGGTGTCAGGCTTGATGGACGG TAAAGGTGGTGGCAAGGATGTGTCTGCAC AGGCCACAGGCAAGAACGTTGGCTGCCTG CAGGAGGCGCTGCAGCTGGCCACTTCCTT CGCCAGCTGCGCCTCGGGGATGTAAAGA ACTGA	
AlaRS1 <sup>C1</sup> <sub>0</sub>	Protein / Human / 1-407 + 498-968	MDSTLTASEIRQRFIDFFKRNEHTYVHSSATI PLDDPTLLFANAGMNQFKPIFLNTIDPSHPM AKLSRAANTQKCIRAGGKHNDLDDVGKDV YHHTFFEMLGWSFSGDYFKELACKMALEL LTQEFGIPIERLYVTYFGGDEAAGLEADLEC KQIWQNLGLDDTKILPGNMKDNFWEAGDT GPCGPCSEIHYDRIGGRDAAHLVNQDDPNV LEIWNLVFIQYNREADGILKPLPKKSIDTGM GLERLVSFLQNKMSNYDSDLFVPYFEAIQK GTGARPYTGKVGAEADADGIDMAYRVLADH ARTITVALADGGRPDNTGRGYVLRRIARRA VRYAHEKLNASRGFFATLVDVVVQSLGDA FPELKKDPDMVKDIINEEEVQFLKTLRGRR ILDRKIQSLGDSKTIPVFENTVATVMALRRE	SEQ.ID. NO.120

<b>Table 5A</b> <b>C-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
		KMFVEEVSTGQECGVVLDKTCFYAEQGGQI YDEGYLVKVDDSSSEDKTEFTVKNAQVRGG YVLHIGTIYGDLKVGQVWLFIDEPRRRPIM SNHTATHILNFALRSVLGEADQKGS�VAPD RLRFDFTAKGAMSTQQIKKAEIEMIEA AKAVYTQDCPLAAAKAIQGLRAVFDETY DPVRVVSIGVPVSELLDDPSGPAGSLTSVEF CGGTHLRNSSHAGAFVIVTEEAIAKGIRIV AVTGAEAQKALRKAESLKKCLSVMEAKVK AQTAPNKDVQREIADLGEALATAVIPQWQK DELRETLKSLKKVMDDLDRASKADVQKRV LEKTKQFIDSNPNQPLVILEMESGASAKALN EALKLFKMHSPTSAMLFVDNEAGKITCL CQVPQNAANRGLKASEWVQQVSGLMGK GGGKDVSAQATGKNVGCLQEALQLATSFA QLRLGDVKN	
AlaRS1 <sup>CT</sup> <sub>0</sub>	DNA / Human /	ATGGACTCTACTCTAACAGCAAGTGAAAT CCGGCAGCGATTTATAGATTTCTTCAAGA GGAACGAGCATACTGATGTTCACTCGTCT GCCACCATCCCATTGGATGACCCCACTTT GCTCTTTGCCAATGCAGGCATGAACCAGT TTAAACCCATTTTCCTGAACACAATTGAC CCATCTCACCCCATGGCAAAGCTGAGCAG AGCTGCCAATACCCAGAAGTGCATCCGGG CTGGGGGCAAACATAATGACCTGGACGAT GTGGGCAAGGATGTCTATCATCACACCTT CTTGAGATGCTGGGCTCTGGTCTTTTGG AGATTACTTTAAGGAATTGGCATGTAAGA TGGCTCTGGAACCTCACCACAAGAGTTT GGCATTCCCATTGAAAGACTTTATGTTAC TTACTTTGGCGGGGATGAAGCAGCTGGCT TAGAAGCAGATCTGGAATGCAAACAGAT CTGGCAAAATTTGGGGCTGGATGACACCA AAATCCTCCCAGGCAACATGAAGGATAAC TTCTGGGAGATGGGTGACACGGGCCCTG TGGTCCTTGCAGTGAGATCCACTACGACC GGATTGGTGGTCGGGACGCCGCACATCTT GTCAACCAGGACGACCCTAATGTGCTGGA GATCTGGAACCTTGTGTTTATCCAGTATA ACAGGGAAGCTGATGGCATTCTGAAACCT CTTCCCAAGAAAAGCATTGACACAGGGAT GGGCCTGGAACGACTGGTATCTGTGCTGC AGAATAAGATGTCCAACCTATGACACTGAC	SEQ.ID. NO.121



<b>Table 5A</b> <b>C-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID. NO.</b>
		CTTTTGTCCCTTACTTTGAAGCCATTCAG AAGGGCACAGGTGCCCCGACCATACTG GGAAAGTTGGTGCTGAGGATGCCGATGG GATTGACATGGCCTACCGGGTGCTGGCTG ACCACGCTCGGACCATCACTGTGGCACTG GCTGATGGTGGCCGGCCTGACAACACAGG GCGTGGATATGTGTTGAGACGGATTCTCC GCCGAGCTGTCCGATACGCCCATGAAAAG CTCAATGCCAGCAGGGGCTTCTTTGCTAC GTTAGTGGATGTTGTCGTCCAGTCCCTGG GAGATGCATTCCTGAGCTGAAGAAGGAC CCAGACATGGTGAAGGACATCATTAATGA AGAAGAGGTGCAGTTTCTCAAGACTCTCA GCAGAGGGCGTCGCATCCTGGACAGGAA AATTCAGAGCCTGGGAGACAGCAAGACC ATTCCCGTATTTGAGAACACAGTGGCTAC GGTGATGGCTCTGCGCAGGGAGAAGATGT TCGTGGAAGAGGTGTCCACAGGCCAGGA GTGTGGAGTGGTGCTGGACAAGACCTGTT TCTATGCTGAGCAAGGAGGCCAGATCTAT GACGAAGGCTACCTGGTGAAGGTGGATG ACAGCAGTGAAGATAAAACAGAGTTTAC AGTGAAGAATGCTCAGGTCCGAGGAGGG TATGTGCTACACATTGGAACCATCTACGG TGACCTGAAAGTGGGGGATCAGGTCTGGC TGTTTATTGATGAGCCCCGACGAAGACCC ATCATGAGCAACCACACAGCTACGCACAT TCTGAACTTCGCCCTGCGCTCAGTGCTTG GGGAAGCTGACCAGAAAGGCTCATTGGTT GCTCCTGACCGCCTCAGATTTGACTTTACT GCCAAGGGAGCCATGTCCACCCAACAGAT CAAGAAGGCTGAAGAGATTGCTAATGAG ATGATTGAGGCAGCCAAGGCCGTCTATAC CCAGGATTGCCCCCTGGCAGCAGCGAAAG CCATCCAGGGCCTACGGGCTGTGTTTGAT GAGACCTATCCTGACCCTGTGCGAGTCGT CTCCATTGGGGTCCCGGTGTCCGAGTTGC TGGATGACCCCTCTGGGCCTGCTGGCTCC CTGACTTCTGTTGAGTTCTGTGGGGGAAC GCACCTGCGGAACTCGAGTCATGCAGGAG CTTTTGTGATCGTGACGGAAGAAGCCATT GCCAAGGGTATCCGGAGGATTGTGGCTGT CACAGGTGCCGAGGCCCAAGAAGGCCCTC	

<b>Table 5A</b> <b>C-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
		AGGAAAGCAGAGAGCTTGAAGAAATGTC TCTCTGTCATGGAAGCCAAAGTGAAGGCT CAGACTGCTCCAAACAAGGATGTGCAGA GGGAGATCGCTGACCTTGGAGAGGCCCTG GCCACTGCAGTCATCCCCAGTGGCAGAA GGATGAATTGCGGGAGACTCTCAAATCCC TAAAGAAGGTCATGGATGACTTGGACCGA GCCAGCAAAGCCGATGTCCAGAAACGAG TGTTAGAGAAGACGAAGCAGTTCATCGAC AGCAACCCCAACCAGCCTCTTGTATCCT GGAGATGGAGAGCGGCGCCTCAGCCAAG GCCCTGAATGAAGCCTTGAAGCTCTTCAA GATGCACTCCCCTCAGACTTCTGCCATGC TCTTCACGGTGGACAATGAGGCTGGCAAG ATCACGTGCCTGTGTCAAGTTCCCCAGAA TGCAGCCAATCGGGGCTTAAAAGCCAGCG AGTGGGTGCAGCAGGTGTCAGGCTTGATG GACGGTAAAGGTGGTGGCAAGGATGTGT CTGCACAGGCCACAGGCAAGAACGTTGG CTGCCTGCAGGAGGCGCTGCAGCTGGCCA CTTCCTTCGCCCAGCTGCGCCTCGGGGAT GTAAAGAACTGA	
AlaRS1 <sup>CI</sup> <sub>1</sub>	Protein / Human / 16 aa + 558-968	MIPQSTITIWTTPVVAMKTEFTVKNAQVRGG YVLHIGTIYGDLKVGDVWLFIDEPRRRPIM SNHTATHILNFALRSVLGEADQKGSVLAPD RLRFDFATAKGAMSTQQIKKAEIENEMIEA AKAVYTQDCPLAAAKAIQGLRAVFDETY DPVRVVSIGVPVSELLDDPSGPAGSLTSVEF CGGTHLRNSSHAGAFVIVTEEAIAKGIRRV AVTGAEAQKALRKAESLKKCLSVMEAKVK AQAPNKDVQREIADLGEALATAVIPQWQK DELRETLKSLKKVMDDLDRASKADVQKRV LEKTKQFIDSNPNQPLVILEMESGASAKALN EALKLFBMHSPQTSAMLFTVDNEAGKITCL CQVPQNAANRGLKASEWVQQVSGLMDGK GGGKDVSAQATGKNVGCLQEALQLATSFA QLRLGDVKN	SEQ.ID. NO.122
AlaRS1 <sup>CI</sup> <sub>1</sub>	DNA / Human /	ATGATTCCCCAAAGTACAATTACCATTTG GACTCCAGTGGTAGCTATGAAAACAGAGT TTACAGTGAAGAATGCTCAGGTCCGAGGA GGGTATGTGCTACACATTGGAACCATCTA CGGTGACCTGAAAGTGGGGGATCAGGTCT GGCTGTTTATTGATGAGCCCCGACGAAGA	SEQ.ID. NO.123

<b>Table 5A</b> <b>C-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID. NO.</b>
		CCCATCATGAGCAACCACACAGCTACGCA CATTCTGAACTTCGCCCTGCGCTCAGTGCT TGGGGAAGCTGACCAGAAAGGCTCATTG GTTGCTCCTGACCGCCTCAGATTTGACTTT ACTGCCAAGGGAGCCATGTCCACCCAACA GATCAAGAAGGCTGAAGAGATTGCTAAT GAGATGATTGAGGCAGCCAAGGCCGTCTA TACCCAGGATTGCCCCCTGGCAGCAGCGA AAGCCATCCAGGGCCTACGGGCTGTGTTT GATGAGACCTATCCTGACCCTGTGCGAGT CGTCTCCATTGGGGTCCCGGTGTCCGAGT TGCTGGATGACCCCTCTGGGCCTGCTGGC TCCCTGACTTCTGTTGAGTTCTGTGGGGG AACGCACCTGCGGAACTCGAGTCATGCAG GAGCTTTTGTGATCGTGACGGAAGAAGCC ATTGCCAAGGGTATCCGGAGGATTGTGGC TGTCACAGGTGCCGAGGCCCCAGAAGGCC TCAGGAAAGCAGAGAGCTTGAAGAAATG TCTCTCTGTCATGGAAGCCAAAGTGAAGG CTCAGACTGCTCCAAACAAGGATGTGCAG AGGGAGATCGCTGACCTTGGAGAGGCCCT GGCCACTGCAGTCATCCCCCAGTGGCAGA AGGATGAATTGCGGGAGACTCTCAAATCC CTAAAGAAGGTCATGGATGACTTGGACCG AGCCAGCAAAGCCGATGTCCAGAAACGA GTGTTAGAGAAGACGAAGCAGTTCATCGA CAGCAACCCCAACCAGCCTCTTGTTCATCC TGGAGATGGAGAGCGGCGCCTCAGCCAA GGCCCTGAATGAAGCCTTGAAGCTCTTCA AGATGCACTCCCCTCAGACTTCTGCCATG CTCTTACGGTGGACAATGAGGCTGGCAA GATCACGTGCCTGTGTCAAGTTCCCCAGA ATGCAGCCAATCGGGGCTTAAAAGCCAGC GAGTGGGTGCAGCAGGTGTCAGGCTTGAT GGACGGTAAAGGTGGTGGCAAGGATGTG TCTGCACAGGCCACAGGCAAGAACGTTGG CTGCCTGCAGGAGGCGCTGCAGCTGGCCA CTTCTTCGCCCAGCTGCGCCTCGGGGAT GTAAAGAACTGA	
AlaRS1 <sup>C1</sup> <sub>2</sub>	Protein / Human / 2 aa + 727-968	MRHLRNSSHAGAFVIVTEEAIAKGIRRIVAV TGAEAQKALRKAESLKKCLSVMEAKVKAQ TAPNKDVQREIADLGEALATAVIPQWQKDE LRETLKSLKKVMDDLDRASKADVQKRVLE	SEQ.ID. NO.124

<b>Table 5A</b> <b>C-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
		KTKQFIDSNPNQPLVILEMESGASAKALNEA LKLFKMHSPTQTSAMLFTVDNEAGKITCLCQ VPQNAANRGLKASEWVQQVSGLMDGKGG GKDVSAQATGKNVGCLQEALQLATSFAQL RLGDVKN	
AlaRS1 <sup>CI</sup> <sub>2</sub>	DNA / Human /	ATGAGGCACCTGCGGAACCTCGAGTCATGC AGGAGCTTTTGTGATCGTGACGGAAGAAG CCATTGCCAAGGGTATCCGGAGGATTGTG GCTGTACAGGTGCCGAGGCCCAGAAGG CCCTCAGGAAAGCAGAGAGCTTGAAGAA ATGTCTCTCTGTCATGGAAGCCAAAGTGA AGGCTCAGACTGCTCCAAACAAGGATGTG CAGAGGGAGATCGCTGACCTTGGAGAGG CCCTGGCCACTGCAGTCATCCCCAGTGG CAGAAGGATGAATTGCGGGAGACTCTCA AATCCCTAAAGAAGGTCATGGATGACTTG GACCGAGCCAGCAAAGCCGATGTCCAGA AACGAGTGTTAGAGAAGACGAAGCAGTT CATCGACAGCAACCCCAACCAGCCTCTTG TCATCCTGGAGATGGAGAGCGGCGCCTCA GCCAAGGCCCTGAATGAAGCCTTGAAGCT CTTCAAGATGCACTCCCCTCAGACTTCTG CCATGCTCTTACGGTGGACAATGAGGCT GGCAAGATCACGTGCCTGTGTCAAGTTCC CCAGAATGCAGCCAATCGGGGCTTAAAA GCCAGCGAGTGGGTGCAGCAGGTGTCAG GCTTGATGGACGGTAAAGGTGGTGGCAA GGATGTGTCTGCACAGGCCACAGGCAAG AACGTTGGCTGCCTGCAGGAGGCGCTGCA GCTGGCCACTTCCTTCGCCCAGCTGCGCC TCGGGGATGTAAAGAACTGA	SEQ.ID. NO.125
AlaRS1 <sup>CI</sup> <sub>3</sub>	Protein / Human / 370-968	MVKDIINEEEVQFLKTLRGRRLDRKIQSL GDSKTIPGDTAWLLYDTYGFVDTLGLIAEE KGLVVDMDGFEEERKLAQLKSQKGAGGE DLIMLDIYAIEELRARGLEVTD DSPKYNH L DSSGSYVFENTVATVMALRREKMFVEEVST GQECGVVLDKTCFYAEQGGQIYDEGYLVK VDDSSSEDKTEFTVKNAQVRGGYVLHIGTIY GDLKVG DQVWLFIDEPRRRPIMSNHTATHI LNFALRSVLGEADQKGS LVAPDRLRFDFTA KGAMSTQQIKKAEI ANEMIEAAKAVYTQD CPLAAAKAIQGLRAVFDETYPD PVRVVSIG VPVSELLDDPSGPAGSLTSVEFCGGTHLRNS	SEQ.ID. NO.126

<b>Table 5A</b> <b>C-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID. NO.</b>
		SHAGAFVIVTEEAIAKGIRRVAVTGAEQK ALRKAESLKKCLSVMEAKVKAQTAPNKDV QREIADLGEALATAVIPQWQKDELRETLKS LKKVMDDLDRASKADVQKRVLEKTKQFID SNPNQPLVILEMESGASAKALNEALKLFKM HSPQTSAMLFTVDNEAGKITCLCQVPQNAA NRGLKASEWVQQVSGLMDGKGGGKDVSA QATGKNVGCLQEALQLATSFAQLRLGDVK N	
AlaRS1 <sup>CT</sup> <sub>3</sub>	DNA / Human /	ATGGTGAAGGACATCATTAATGAAGAAG AGGTGCAGTTTCTCAAGACTCTCAGCAGA GGGCGTCGCATCCTGGACAGGAAAATTCA GAGCCTGGGAGACAGCAAGACCATTCCC GGAGACACTGCTTGGCTCCTCTATGACAC CTATGGGTTTCCAGTGGATCTGACTGGAC TGATTGCTGAAGAGAAGGGCCTGGTGGTA GACATGGATGGCTTTGAAGAGGAGAGGA AACTGGCCCAGCTGAAATCACAGGGCAA GGGAGCTGGTGGGGAAGACCTCATTATGC TGGACATTTACGCTATCGAAGAGCTCCGG GCACGGGGTCTGGAGGTCACAGATGATTC CCCAAAGTACAATTACCATTTGGACTCCA GTGGTAGCTATGTATTTGAGAACACAGTG GCTACGGTGATGGCTCTGCGCAGGGAGAA GATGTTTCGTGGAAGAGGTGTCCACAGGCC AGGAGTGTGGAGTGGTGCTGGACAAGAC CTGTTTCTATGCTGAGCAAGGAGGCCAGA TCTATGACGAAGGCTACCTGGTGAAGGTG GATGACAGCAGTGAAGATAAAACAGAGT TTACAGTGAAGAATGCTCAGGTCCGAGGA GGGTATGTGCTACACATTGGAACCATCTA CGGTGACCTGAAAGTGGGGGATCAGGTCT GGCTGTTTATTGATGAGCCCCGACGAAGA CCCATCATGAGCAACCACACAGCTACGCA CATTCTGAACTTCGCCCTGCGCTCAGTGCT TGGGGAAGCTGACCAGAAAGGCTCATTG GTTGCTCCTGACCGCCTCAGATTTGACTTT ACTGCCAAGGGAGCCATGTCCACCCAACA GATCAAGAAGGCTGAAGAGATTGCTAAT GAGATGATTGAGGCAGCCAAGGCCGTCTA TACCCAGGATTGCCCCCTGGCAGCAGCGA AAGCCATCCAGGGCCTACGGGCTGTGTTT GATGAGACCTATCCTGACCCTGTGCGAGT	SEQ.ID. NO.127

<b>Table 5A</b> <b>C-terminal AARS polypeptides and alternative transcripts identified by Deep Sequencing</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID. NO.</b>
		CGTCTCCATTGGGGTCCCGGTGTCCGAGT TGCTGGATGACCCCTCTGGGCCTGCTGGC TCCCTGACTTCTGTTGAGTTCTGTGGGGG AACGCACCTGCGGAACCTCGAGTCATGCAG GAGCTTTTGTGATCGTGACGGAAGAAGCC ATTGCCAAGGGTATCCGGAGGATTGTGGC TGTCACAGGTGCCGAGGCCCAAGAAGCCCC TCAGGAAAGCAGAGAGCCTTGAAGAAATG TCTCTCTGTCATGGAAGCCAAAGTGAAGG CTCAGACTGCTCCAAACAAGGATGTGCAG AGGGAGATCGCTGACCTTGGAGAGGCCCT GGCCACTGCAGTCATCCCCCAGTGGCAGA AGGATGAATTGCGGGAGACTCTCAAATCC CTAAAGAAGGTCATGGATGACTTGGACCG AGCCAGCAAAGCCGATGTCCAGAAACGA GTGTTAGAGAAGACGAAGCAGTTCATCGA CAGCAACCCCAACCAGCCTCTTGTCATCC TGGAGATGGAGAGCGGCGCCTCAGCCAA GGCCCTGAATGAAGCCTTGAAGCTCTTCA AGATGCACTCCCCTCAGACTTCTGCCATG CTCTTCACGGTGGACAATGAGGCTGGCAA GATCACGTGCCTGTGTCAAGTTCCCCAGA ATGCAGCCAATCGGGGCTTAAAAGCCAGC GAGTGGGTGCAGCAGGTGTCAGGCTTGAT GGACGGTAAAGGTGGTGGCAAGGATGTG TCTGCACAGGCCACAGGCAAGAACGTTGG CTGCCTGCAGGAGGCGCTGCAGCTGGCCA CTTCTTCGCCCAGCTGCGCCTCGGGGAT GTAAAGAACTGA	

<b>Table 5B</b> <b>AARS polypeptides unique splice junctions</b>					
<b>Name</b>	<b>Type / species</b>	<b>Amino acid and Nucleic Acid Sequences in the vicinity of the unique splice junction</b>	<b>Human fetal brain</b>	<b>Human Adult brain</b>	<b>SEQ.ID. NO.</b>
A1-AS01	DNA / Human /	ACCTTGTGTTTCATCCAGTAT AACAG GGCACAGGTGCCC GACCATACTG	1	0	SEQ.ID. NO.128
	Protein / Human /	N/A			
A1-	DNA /	CTTTGCCAATGCAGGCATG			SEQ.ID.

<b>Table 5B</b> <b>AARS polypeptides unique splice junctions</b>					
<b>Name</b>	<b>Type / species</b>	<b>Amino acid and Nucleic Acid Sequences in the vicinity of the unique splice junction</b>	<b>Human fetal brain</b>	<b>Human Adult brain</b>	<b>SEQ.ID. NO.</b>
AS02	Human /	AACCAG CTGAAATCACAG GGCAAGGGAGCTG			NO.129
	Protein / Human /	FANAGMNQLKSQKGA			SEQ.ID. NO.130
A1-AS03	DNA / Human /	CTGGGAGACAGCAAGACCA TTCCCG TATTTGAGAACAC AGTGGCTACGGT	1,415	397	SEQ.ID. NO.131
	Protein / Human /	LGDSKTIPVFENTVAT			SEQ.ID. NO.132
A1-AS04	DNA / Human /	CATTTGGACTCCAGTGGTA GCTATG AAAACAGAGTTTA CAGTGAAGAATG			SEQ.ID. NO.133
	Protein / Human /	IWTPVVAMKTEFTVKN			SEQ.ID. NO.134
A1-AS06	DNA / Human /	TCAGGTCTGGCTGTTTATTG ATGAG GCACCTGCGGAAC CGAGTCATGCA			SEQ.ID. NO.135
	Protein / Human /	MRHLRNSSHA			SEQ.ID. NO.136
A1-AS07	DNA / Human /	GCCGGCCTGACAACACAGG GCGTGG GGAGATGCATTC CTGAGCTGAAGA			SEQ.ID. NO.137
	Protein / Human /	N/A			

<b>Table 6</b> <b>C-terminal AARS polypeptides identified by Bioinformatics</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID. NO.</b>
AlaRS1 <sub>C4</sub>	Protein / Human / 659-968	MIEAAKAVYTQDCPLAAAKAIQGLRAVFDE TYPDPVRVVSIGVPVSELLDDPSGPAGSLTSV EFCGGTHLRNSSHAGAFVIVTEEAIAKGIRRI VAVTGAEAQKALRKAESLKKCLSVMEAKV KAQTAPNKDVQREIADLGEALATAVIPQWQ KDELRETLKSLKKVMDDLDRASKADVQKR VLEKTKQFIDSNPNQPLVILEMESGASAKAL NEALKLFKMHSPQTSAMLFTVDNEAGKITCL CQVPQNAANRGLKASEWVQQVSGGLMDGKG GGKDVSAQATGKNVGCLQEALQLATSFAQL RLGDVKN	SEQ.ID. NO.138
AlaRS1	DNA /	ATGATTGAGGCAGCCAAGGCCGTCTATAC	SEQ.ID.

<b>Table 6</b> <b>C-terminal AARS polypeptides identified by Bioinformatics</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID. NO.</b>
C4	Human /	CCAGGATTGCCCCCTGGCAGCAGCGAAAG CCATCCAGGGCCTACGGGCTGTGTTTGATG AGACCTATCCTGACCCTGTGCGAGTCGTCT CCATTGGGGTCCCGGTGTCCGAGTTGCTGG ATGACCCCTCTGGGCCTGCTGGCTCCCTGA CTTCTGTTGAGTTCTGTGGGGGAACGCACC TGCGGAACTCGAGTCATGCAGGAGCTTTTG TGATCGTGACGGAAGAAGCCATTGCCAAG GGTATCCGGAGGATTGTGGCTGTCACAGG TGCCGAGGCCCAGAAGGCCCTCAGGAAAG CAGAGAGCTTGAAGAAATGTCTCTCTGTCA TGGAAGCCAAAGTGAAGGCTCAGACTGCT CCAAACAAGGATGTGCAGAGGGAGATCGC TGACCTTGAGAGAGGCCCTGGCCACTGCAG TCATCCCCCAGTGGCAGAAGGATGAATTG CGGGAGACTCTCAAATCCCTAAAGAAGGT CATGGATGACTTGGACCGAGCCAGCAAAG CCGATGTCCAGAAACGAGTGTTAGAGAAG ACGAAGCAGTTCATCGACAGCAACCCCAA CCAGCCTCTTGTATCCTGGAGATGGAGAG CGGCGCCTCAGCCAAGGCCCTGAATGAAG CCTTGAAGCTCTTCAAGATGCACTCCCCTC AGACTTCTGCCATGCTCTTCACGGTGGACA ATGAGGCTGGCAAGATCACGTGCCTGTGT CAAGTTCCCCAGAATGCAGCCAATCGGGG CTTAAAGCCAGCGAGTGGGTGCAGCAGG TGTCAGGCTTGATGGACGGTAAAGGTGGT GGCAAGGATGTGTCTGCACAGGCCACAGG CAAGAACGTTGGCTGCCTGCAGGAGGCGC TGCAGCTGGCCACTTCCTTCGCCCAGCTGC GCCTCGGGGATGTAAAGAACTGA	NO.139
AlaRS1 C5	Protein / Human / 691-968	PDPVRVVSIGVPVSELDDPSGPAGSLTSVEF CGGTHLRNSSHAGAFVIVTEEAIAKGIRRIVA VTGAEAQKALRKAESLKKCLSVMEAKVKA QTAPNKDVQREIADLGEALATAVIPQWQKD ELRETLKSLKKVMDDLDRASKADVQKRVLE KTKQFIDSNPNQPLVILEMESGASAKALNEA LKLFKMHSPQTSAMLFTVDNEAGKITCLCQ VPQNAANRGLKASEWVQQVSGGLMDGKGGG KDVSAAQATGKNVGCLQEALQLATSFAQLRL GDVKN	SEQ.ID. NO.140
AlaRS1 C5	DNA / Human /	CCTGACCCTGTGCGAGTCGTCTCCATTGGG GTCCCCGGTGTCCGAGTTGCTGGATGACCCC TCTGGGCCTGCTGGCTCCCTGACTTCTGTT	SEQ.ID. NO.141



<b>Table 6</b> <b>C-terminal AARS polypeptides identified by Bioinformatics</b>			
<b>Name</b>	<b>Type / species / Residues</b>	<b>Amino acid and Nucleic Acid Sequences</b>	<b>SEQ.ID. NO.</b>
		GAGTTCTGTGGGGGAACGCACCTGCGGAA CTCGAGTCATGCAGGAGCTTTTGTGATCGT GACGGAAGAAGCCATTGCCAAGGGTATCC GGAGGATTGTGGCTGTACAGGTGCCGAG GCCCAGAAGGCCCTCAGGAAAGCAGAGAG CTTGAAGAAATGTCTCTCTGTCATGGAAGC CAAAGTGAAGGCTCAGACTGCTCCAAACA AGGATGTGCAGAGGGAGATCGCTGACCTT GGAGAGGCCCTGGCCACTGCAGTCATCCC CCAGTGGCAGAAGGATGAATTGCGGGAGA CTCTCAAATCCCTAAAGAAGGTCATGGAT GACTTGGACCGAGCCAGCAAAGCCGATGT CCAGAAACGAGTGTTAGAGAAGACGAAGC AGTTCATCGACAGCAACCCCAACCAGCCT CTTGTCATCCTGGAGATGGAGAGCGGCGC CTCAGCCAAGGCCCTGAATGAAGCCTTGA AGCTCTTCAAGATGCACTCCCCTCAGACTT CTGCCATGCTCTTCACGGTGGACAATGAGG CTGGCAAGATCACGTGCCTGTGTCAAGTTC CCCAGAATGCAGCCAATCGGGGCTTAAAA GCCAGCGAGTGGGTGCAGCAGGTGTCAGG CTTGATGGACGGTAAAGGTGGTGGCAAGG ATGTGTCTGCACAGGCCACAGGCAAGAAC GTTGGCTGCCTGCAGGAGGCGCTGCAGCT GGCCACTTCCTTCGCCCAGCTGCGCCTCGG GGATGTAAAGAACTGA	
AlaRS1 C6	Protein / Human / 747-968	KGIRRIVAVTGAEAQKALRKAESLKKCLSV MEAKVKAQTAPNKDVQREIADLGEALATAV IPQWQKDELRETLKSLKKVMDDLDRASKAD VQKRVLEKTKQFIDSNPNQPLVILEMESGAS AKALNEALKLFKMHSPQTSAMLFTVDNEAG KITCLCQVPQNAANRGLKASEWVQQVSGLM DGKGGGKDVSAQATGKNVGCLQEALQLAT SFAQLRLGDVKN	SEQ.ID. NO.142
AlaRS1 C6	DNA / Human /	AAGGGTATCCGGAGGATTGTGGCTGTCAC AGGTGCCGAGGCCCAGAAGGCCCTCAGGA AAGCAGAGAGCTTGAAGAAATGTCTCTCT GTCATGGAAGCCAAAGTGAAGGCTCAGAC TGCTCCAAACAAGGATGTGCAGAGGGAGA TCGCTGACCTTGGAGAGGCCCTGGCCACTG CAGTCATCCCCCAGTGGCAGAAGGATGAA TTGCGGGAGACTCTCAAATCCCTAAAGAA GGTCATGGATGACTTGGACCGAGCCAGCA AAGCCGATGTCCAGAAACGAGTGTTAGAG	SEQ.ID. NO.143

<b>Table 6</b> <b>C-terminal AARS polypeptides identified by Bioinformatics</b>			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
		AAGACGAAGCAGTTCATCGACAGCAACCC CAACCAGCCTCTTGTTCATCCTGGAGATGGA GAGCGGCGCCTCAGCCAAGGCCCTGAATG AAGCCTTGAAGCTCTTCAAGATGCACTCCC CTCAGACTTCTGCCATGCTCTTCACGGTGG ACAATGAGGCTGGCAAGATCACGTGCCTG TGTCAAGTTCCCCAGAATGCAGCCAATCG GGGCTTAAAAGCCAGCGAGTGGGTGCAGC AGGTGTCAGGCTTGATGGACGGTAAAGGT GGTGGCAAGGATGTGTCTGCACAGGCCAC AGGCAAGAACGTTGGCTGCCTGCAGGAGG CGCTGCAGCTGGCCACTTCCTTCGCCCAGC TGCGCCTCGGGGATGTAAAGAACTGA	
AlaRS1 C7	Protein / Human / 856-968	PLVILEMESGASAKALNEALKLFKMHSPTQS AMLFTVDNEAGKITCLCQVPQNAANRGLKA SEWVQQVSGLMMDGKGGGKDVSAQATGKN VGCLQEALQLATSFAQLRLGDVKN	SEQ.ID. NO.144
AlaRS1 C7	DNA / Human /	CCTCTTGTTCATCCTGGAGATGGAGAGCGGC GCCTCAGCCAAGGCCCTGAATGAAGCCTT GAAGCTCTTCAAGATGCACTCCCCTCAGAC TTCTGCCATGCTCTTCACGGTGGACAATGA GGCTGGCAAGATCACGTGCCTGTGTCAAG TTCCCCAGAATGCAGCCAATCGGGGCTTA AAAGCCAGCGAGTGGGTGCAGCAGGTGTC AGGCTTGATGGACGGTAAAGGTGGTGGCA AGGATGTGTCTGCACAGGCCACAGGCAAG AACGTTGGCTGCCTGCAGGAGGCGCTGCA GCTGGCCACTTCCTTCGCCCAGCTGCGCCT CGGGGATGTAAAGAACTGA	SEQ.ID. NO.145

**Internal AARS Polypeptides: (Tables 7, 8 & 9)**

<b>Table 7A</b> <b>AARS polypeptides identified by MS</b>			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.

<b>Table 7B</b> <b>Mass spec peptides detected and inferred linking peptides</b>		
Type / species	Sequence	SEQ.ID. NO.

Table 7C Concatenated sequences based on mass spec peptides detected		
Type / species	Sequence	SEQ.ID. NO.

Table 8 AARS polypeptides and alternative transcripts identified by Deep Sequencing			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.

Table 8B AARS polypeptides unique splice junctions			
Name	Type / species	Amino acid and Nucleic Acid Sequences in the vicinity of the unique splice junction	SEQ.ID. NO.

Table 9 AARS polypeptides identified by Bioinformatics			
Name	Type / species / Residues	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
AlaRS1 <sup>1</sup>	Protein / Human / 476-749	RGLEVTDDSPKYNHYHLDSSGSYVFENTVATVMALRR EKMFEVEEVSTGQECGVVLDKTCFYAEQGGQIYDEGY LVKVDDSSDKTEFTVKNAQVRGGYVLHIGTIYGDLK VGDQVWLFIDEPRRRPIMSNHTATHILNFALRSVLGEA DQKGS LVAPDRLRFDTAKGAMSTQQIKKAEI ANEM IEAAKAVYTQDCPLAAAKAIQGLRAVFDETYDPVVRV VSGVPVSELLDDPSGPAGSLTSVEFCGGTHLRNSSHA GAFVIVTEEAIAKGI	SEQ.ID. NO.177
AlaRS1 <sup>1</sup>	DNA / Human /	CGGGGTCTGGAGGTCACAGATGATTCCCC AAAGTACAATTACCATTTGGACTCCAGTGG TAGCTATGTATTTGAGAACACAGTGGCTAC GGTGATGGCTCTGCGCAGGGAGAAAGATGT TCGTGGAAGAGGTGTCCACAGGCCAGGAG TGTGGAGTGGTGCTGGACAAGACCTGTTTC TATGCTGAGCAAGGAGGCCAGATCTATGA CGAAGGCTACCTGGTGAAGGTGGATGACA GCAGTGAAGATAAAACAGAGTTTACAGTG AAGAATGCTCAGGTCCGAGGAGGGTATGT GCTACACATTGGAACCATCTACGGTGACCT GAAAGTGGGGGATCAGGTCTGGCTGTTTA TTGATGAGCCCCGACGAAGACCCATCATG AGCAACCACACAGCTACGCACATTCTGAA	SEQ.ID. NO.178

<b>Table 9</b> <b>AARS polypeptides identified by Bioinformatics</b>			
Name	Type / species /Residue s	Amino acid and Nucleic Acid Sequences	SEQ.ID. NO.
		CTTCGCCCTGCGCTCAGTGCTTGGGGAAGC TGACCAGAAAGGCTCATTGGTTGCTCCTGA CCGCCTCAGATTTGACTTTACTGCCAAGGG AGCCATGTCCACCCAACAGATCAAGAAGG CTGAAGAGATTGCTAATGAGATGATTGAG GCAGCCAAGGCCGTCTATACCCAGGATTG CCCCCTGGCAGCAGCGAAAGCCATCCAGG GCCTACGGGCTGTGTTTGATGAGACCTATC CTGACCCTGTGCGAGTCGTCTCCATTGGGG TCCCGGTGTCCGAGTTGCTGGATGACCCCT CTGGGCCTGCTGGCTCCCTGACTTCTGTTG AGTTCTGTGGGGGAACGCACCTGCGGAAC TCGAGTCATGCAGGAGCTTTTGTGATCGTG ACGGAAGAAGCCATTGCCAAGGGTATC	

[00117] “Protein fragments,” or the amino acid sequence of protein fragments, such as proteolytic fragments or splice variant fragments, can be characterized, identified, or derived according to a variety of techniques. For instance, splice variants can be identified by techniques such as deep sequencing (*see, e.g., Xing et al., RNA*. 14:1470-1479, 2008; and Zhang *et al., Genome Research*. 17:503-509, 2007). As a further example, protein fragments such as proteolytic fragments can be identified *in vitro*, such as by incubating full-length or other AARS polypeptides with selected proteases, or they can be identified endogenously (*e.g., in vivo*). In certain embodiments, protein fragments such as endogenous proteolytic fragments can be generated or identified, for instance, by recombinantly expressing full-length or other AARS polypeptides in a selected microorganism or eukaryotic cell that has been either modified to contain one or more selected proteases, or that naturally contains one or more proteases that are capable of acting on a selected AARS polypeptide, and isolating and characterizing the endogenously produced protein fragments therefrom.

[00118] In certain embodiments, protein fragments such as endogenous (*e.g., naturally-occurring*) proteolytic fragments can be generated or identified, for instance, from various cellular fractions (*e.g., cytosolic, membrane, nuclear*) and/or growth medium of various cell-types, including, for example, immune cells such as monocytes, dendritic cells, macrophages (*e.g., RAW 264.7 macrophages*), neutrophils, eosinophils, basophils, and lymphocytes, such as B-cells and T-cells (*e.g., CD4+ helper and CD8+*

killer cells), including primary T-cells and T-cell lines such as Jurkat T-cells, as well as natural killer (NK) cells.

[00119] In certain embodiments, protein fragments such as endogenous proteolytic fragments, however generated, can be identified by techniques such as mass-spectrometry, or equivalent techniques. Once an *in vitro* or endogenously identified protein fragment has been generated or identified, it can be mapped or sequenced, and, for example, cloned into an expression vector for recombinant production, or produced synthetically.

[00120] A wide variety of proteases can be used to produce, identify, derive, or characterize the sequence of AARS protein fragments such as proteolytic fragments. Generally, proteases are usually classified according to three major criteria: (i) the reaction catalyzed, (ii) the chemical nature of the catalytic site, and (iii) the evolutionary relationship, as revealed by the structure. General examples of proteases or proteinases, as classified by mechanism of catalysis, include aspartic proteases, serine proteases, cysteine proteases, and metalloproteases.

[00121] Most aspartic proteases belong to the pepsin family. This family includes digestive enzymes, such as pepsin and chymosin, as well as lysosomal cathepsins D and processing enzymes such as renin, and certain fungal proteases (*e.g.*, penicillopepsin, rhizopuspepsin, endothiapepsin). A second family of aspartic proteases includes viral proteinases such as the protease from the AIDS virus (HIV), also called retropepsin.

[00122] Serine proteases include two distinct families. First, the chymotrypsin family, which includes the mammalian enzymes such as chymotrypsin, trypsin, elastase, and kallikrein, and second, the subtilisin family, which includes the bacterial enzymes such as subtilisin. The general 3D structure between these two families is different, but they have the same active site geometry, and catalysis proceeds via the same mechanism. The serine proteases exhibit different substrate specificities, differences which relate mainly to amino acid substitutions in the various enzyme subsites (substrate residue interacting sites). Some serine proteases have an extended interaction site with the substrate whereas others have a specificity that is restricted to the P1 substrate residue.

[00123] The cysteine protease family includes the plant proteases such as papain, actinidin, and bromelain, several mammalian lysosomal cathepsins, the cytosolic calpains (calcium-activated), as well as several parasitic proteases (*e.g.*, *Trypanosoma*, *Schistosoma*). Papain is the archetype and the best studied member of the family. Recent elucidation of the X-ray structure of the Interleukin-1-beta Converting Enzyme has revealed a novel type of fold for cysteine proteinases.

[00124] The metalloproteases are one of the older classes of proteases, found in bacteria, fungi, and higher organisms. They differ widely in their sequences and their 3D structures, but the great majority of enzymes contain a zinc atom that is catalytically active. In some cases, zinc may be replaced by another metal such as cobalt or nickel without loss of proteolytic activity. Bacterial thermolysin has been well characterized and its crystallographic structure indicates that zinc is bound by two histidines and one glutamic acid. Many metalloproteases contain the sequence motif HEXXH, which provides two histidine ligands for the zinc. The third ligand is either a glutamic acid (thermolysin, neprilysin, alanyl aminopeptidase) or a histidine (astacin, serralsin).

[00125] Illustrative proteases include, for example, achromopeptidase, aminopeptidase, ancrod, angiotensin converting enzyme, bromelain, calpain, calpain I, calpain II, carboxypeptidase A, carboxypeptidase B, carboxypeptidase G, carboxypeptidase P, carboxypeptidase W, carboxypeptidase Y, caspase 1, caspase 2, caspase 3, caspase 4, caspase 5, caspase 6, caspase 7, caspase 8, caspase 9, caspase 10, caspase 11, caspase 12, caspase 13, cathepsin B, cathepsin C, cathepsin D, cathepsin E, cathepsin G, cathepsin H, cathepsin L, chymopapain, chymase, chymotrypsin, clostripain, collagenase, complement C1r, complement C1s, complement Factor D, complement factor I, cucumisin, dipeptidyl peptidase IV, elastase (leukocyte), elastase (pancreatic), endoproteinase Arg-C, endoproteinase Asp-N, endoproteinase Glu-C, endoproteinase Lys-C, enterokinase, factor Xa, ficin, furin, granzyme A, granzyme B, HIV Protease, IGase, kallikrein tissue, leucine aminopeptidase (general), leucine aminopeptidase (cytosol), leucine aminopeptidase (microsomal), matrix metalloprotease, methionine aminopeptidase, neutrase, papain, pepsin, plasmin, prolidase, pronase E, prostate specific antigen, protease alkalophilic from *Streptomyces griseus*, protease from *Aspergillus*, protease from *Aspergillus saitoi*, protease from *Aspergillus sojae*, protease (*B. licheniformis*) (alkaline or alcalase), protease from *Bacillus polymyxa*, protease from *Bacillus* sp., protease from *Rhizopus* sp., protease S, proteasomes, proteinase from *Aspergillus oryzae*, proteinase 3, proteinase A, proteinase K, protein C, pyroglutamate aminopeptidase, rennin, rennin, streptokinase, subtilisin, thermolysin, thrombin, tissue plasminogen activator, trypsin, tryptase and urokinase.

[00126] Certain embodiments relate to isolated AARS polypeptides, comprising, consisting essentially of, or consisting of amino acid sequences that have been derived from endogenous, naturally-occurring AARS polypeptide fragments, and pharmaceutical compositions comprising said fragments, and methods of use thereof. These and related embodiments can be generated or identified *in vivo*, *ex vivo*, and/or *in vitro*. In certain preferred *in vitro* embodiments, AARS proteolytic fragments are generated or identified by incubating an AARS polypeptide, such as a full-length

AARS polypeptide, with one or more isolated human proteases, mainly those proteases that are endogenous or natural to humans, such as elastase and others described herein and known in the art. Other embodiments relate to isolated AARS polypeptides, comprising, consisting essentially of, or consisting of amino acid sequences that have been derived from endogenous, naturally-occurring AARS splice variants, and pharmaceutical compositions comprising said fragments, and methods of use thereof. Essentially, AARS protein fragment can be isolated from samples that have been exposed to proteases, whether *in vivo* or *in vitro*.

[00127] In certain embodiments, AARS protein fragments can be identified by techniques such as mass-spectrometry, or equivalent techniques. Merely by way of illustration and not limitation, in certain embodiments the proteomes from various cell types, tissues, or body fluids from a variety of physiological states (*e.g.*, hypoxia, diet, age, disease) or fractions thereof may be separated by 1D SDS-PAGE and the gel lanes cut into bands at fixed intervals; after which the bands may be optionally digested with an appropriate protease, such as trypsin, to release the peptides, which may then be analyzed by 1D reverse phase LC-MS/MS. The resulting proteomic data may be integrated into so-called peptographs, which plot, in the left panel, sequence coverage for a given protein in the horizontal dimension (N to C terminus, left to right) versus SDS-PAGE migration in the vertical dimension (high to low molecular weight, top to bottom). The specific peptide fragments can then be sequenced or mapped. In certain embodiments, the AARS reference fragment may be characterized by its unique molecular weight, as compared, for example, to the molecular weight of the corresponding full-length AARS.

[00128] As noted above, embodiments of the present invention include the AARS polypeptides set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9. Also included are “variants” of the AARS reference polypeptides. The recitation polypeptide “variant” refers to polypeptides that are distinguished from a reference AARS polypeptide by the addition, deletion, and/or substitution of at least one amino acid residue, and which typically retain (*e.g.*, mimic) or modulate (*e.g.*, antagonize) one or more non-canonical activities of a reference AARS polypeptide.

[00129] Moreover human Alanyl tRNA synthetases include several hundred highly related polymorphic forms, and these are known in the art to be at least partially functionally interchangeable. It would thus be a routine matter to select a naturally occurring variant of Alanyl tRNA synthetase, including, for example the single nucleotide polymorphic forms listed in **Table A** to create an AARS polypeptide containing one or more amino acid changes based on the sequence of any of the

homologues, orthologs, and naturally-occurring isoforms of human as well as other species of Alanyl tRNA synthetase.

<b>Table A</b> <b>Human Alanyl tRNA synthetase SNPs</b>			
<b>Gene Bank Accession Number</b>	<b>Nucleotide change</b>	<b>Gene Bank Accession Number</b>	<b>Nucleotide change</b>
rs118124510	A/G	rs11385686	-/T
rs118089442	C/G	rs11379998	-/A
rs118024420	G/T	rs10536509	-/CT
rs118023366	A/G	rs9931779	C/T
rs117953939	A/C	rs9922969	A/G
rs117946875	C/G	rs9673840	G/T
rs117874802	C/T	rs8063749	C/T
rs117857165	A/G	rs8062865	A/G
rs117758610	C/T	rs8062097	C/T
rs117724995	C/T	rs8060812	A/G
rs117721378	A/G	rs8057463	A/T
rs117690614	C/T	rs8056196	A/C
rs117626545	A/T	rs8056048	A/G
rs117598688	A/G	rs8054995	G/T
rs117478137	A/C	rs8046786	C/T
rs117391661	A/G	rs8046662	A/G
rs117349252	C/T	rs8045571	A/G
rs117345088	A/C	rs7205918	C/T
rs117312086	A/C	rs7202056	A/C
rs117253324	C/T	rs7198871	A/G
rs117204433	C/T	rs7193598	C/T
rs117174836	A/C	rs7192000	C/T
rs117108016	C/T	rs7190921	C/G
rs117057627	C/T	rs7186104	C/T
rs117030089	A/G	rs7186024	A/G
rs116941916	G/T	rs7185665	A/C
rs116861081	C/T	rs7184495	C/T
rs116764273	G/T	rs6499309	A/G
rs116762869	G/T	rs4985408	A/C
rs116712553	A/G	rs4985407	A/G
rs116553521	G/T	rs4081753	A/C/G
rs116544164	A/G	rs2070203	C/T
rs116399415	C/T	rs1183292	A/G
rs116369869	A/G	rs1179964	A/G
rs116309532	C/G	rs1179963	A/G
rs116297318	A/G	rs1179962	C/T
rs115882953	G/T	rs1143506	A/G
rs115433323	C/T	rs34538071	A/G
rs115307561	C/G	rs34511567	-/C



<b>Table A</b> <b>Human Alanyl tRNA synthetase SNPs</b>			
<b>Gene Bank Accession Number</b>	<b>Nucleotide change</b>	<b>Gene Bank Accession Number</b>	<b>Nucleotide change</b>
rs115160127	G/T	rs34472677	C/T
rs115134263	C/G	rs34369477	-/A
rs114823713	C/T	rs34363830	-/G
rs114598544	A/C	rs34306553	C/G
rs114595840	A/T	rs34299014	-/A
rs114572751	C/T	rs34296479	-/G
rs114525231	A/T	rs34268796	-/A
rs114484755	A/G	rs34129241	-/T
rs114349195	A/G	rs34087264	A/G
rs114327666	A/G	rs28579865	A/G
rs114266345	A/G	rs28515500	A/G
rs114248390	C/T	rs16970361	C/G
rs114213309	A/C	rs16970307	C/T
rs114121156	A/G	rs16970302	G/T
rs114103198	C/G	rs13338059	A/G
rs114025877	A/C	rs13332867	A/G
rs113960057	A/C	rs13332726	C/G
rs113902056	G/T	rs12924911	A/C
rs113824201	A/G	rs12598140	A/C
rs113812226	A/G	rs12445696	C/T
rs113737430	C/T	rs12445694	C/T
rs113726987	A/G	rs12149776	A/G
rs113691826	A/G	rs12149660	A/G
rs113645277	A/G	rs12149282	C/T
rs113610820	A/G	rs12149011	A/G
rs113518696	G/T	rs11862188	C/T
rs113516046	C/T	rs11862150	C/T
rs113473235	A/T	rs11862053	G/T
rs113447635	A/C	rs11648189	A/G
rs113436178	C/T	rs11643367	A/G
rs113435697	C/T	rs11537667	A/G
rs113403017	C/G	rs11537665	C/T
rs113366691	A/G	rs11537664	C/T
rs113339156	A/G	rs11537663	A/T
rs113285988	A/G	rs11432931	-/T
rs113279975	A/G	rs11432930	-/C
rs113242213	A/G	rs11418378	-/G
rs113207374	C/T	rs57438636	-/A/AA
rs113142906	-/T	rs57281906	C/G
rs113041194	-/CTGGGC	rs57241744	C/T
rs113030826	A/G	rs56833219	C/T
rs112984536	C/T	rs56734965	-/A

**Table A**  
**Human Alanine tRNA synthetase SNPs**

<b>Gene Bank Accession Number</b>	<b>Nucleotide change</b>	<b>Gene Bank Accession Number</b>	<b>Nucleotide change</b>
rs112790341	C/G	rs56394253	-
rs112778140	-/G	rs56389123	G/T
rs112717253	C/T	rs56115411	A/C
rs112620320	A/T	rs55936739	C/T
rs112579905	A/G	rs55819743	A/G
rs112509050	C/T	rs55805938	C/T
rs112411203	A/G	rs55790133	A/C
rs112393101	C/T	rs55683684	C/T
rs112303399	A/G	rs55657328	A/C
rs112174725	A/T	rs55649889	-/AA
rs112146705	C/T	rs36046810	-/A
rs112129984	G/T	rs36016919	-/T
rs112061695	C/T	rs36016621	-/C
rs112055531	A/G	rs35986765	G/T
rs112053092	C/T	rs35769308	C/G
rs112043250	A/G	rs35744709	A/T
rs111980438	C/T	rs35734224	-/A
rs111942451	A/G	rs35687595	-/G
rs111938682	C/G	rs35628195	-/TT
rs111906601	A/G	rs35564585	-/A
rs111864362	A/G	rs35442003	-/T
rs111848470	C/T	rs35421372	-/A
rs111783220	G/T	rs35286903	-/G
rs111776118	-/A	rs35259849	-/T
rs111756183	A/G	rs35168298	-/A
rs111719142	G/T	rs35133764	G/T
rs111698902	A/G	rs35093895	-/T
rs111695190	C/T	rs35052297	-/G
rs111624368	C/T	rs34979457	-/G
rs111620465	A/G	rs34927148	-/A
rs111602240	G/T	rs34846578	-/CT
rs111598897	A/G	rs34829782	-/C
rs111583584	A/G	rs34656219	-/G
rs111567777	G/T	rs34642380	-/AA
rs111539662	-/A	rs71151180	-/TTTTTTTTTTTTTTT
rs111461416	C/T	rs71151179	-/A
rs111435738	A/G	rs67463183	-/A
rs111426337	A/T	rs67413523	-/TG/TGTGTGTGTG
rs111421609	A/G	rs67270571	(LARGEDELETION)/-
rs111412304	A/T	rs67219164	- /GCTGGATTTTCTGATGG
rs111338842	A/G	rs67112563	-/AG

<b>Table A</b> <b>Human Alanine tRNA synthetase SNPs</b>			
<b>Gene Bank Accession Number</b>	<b>Nucleotide change</b>	<b>Gene Bank Accession Number</b>	<b>Nucleotide change</b>
rs111333360	C/T	rs66627085	A/C
rs111315764	A/G	rs66625764	-/CA
rs80257731	C/G	rs62049422	A/C
rs80256286	A/G	rs62049421	A/T
rs80228853	A/C	rs62049420	A/C
rs80222640	A/G	rs62049419	G/T
rs80200401	A/C	rs62049418	A/T
rs80166718	A/G	rs61594809	C/T
rs80140241	G/T	rs61172065	C/T
rs80062047	C/G	rs60903849	A/G
rs79944605	G/T	rs60853195	A/G
rs79879166	C/T	rs60567469	A/G
rs79809458	A/G	rs60502803	-/TTTTTT
rs79785372	A/G	rs60445375	-/CTGGGC
rs79355112	- /AGAAAAAAAAAA AAAAAAAAAA	rs60426209	C/T
rs79296187	C/T	rs60103946	A/G
rs79277862	A/C	rs59869320	C/T
rs79100623	A/T	rs59612351	A/G
rs111906601	A/G	rs59384897	G/T
rs111864362	A/G	rs59002209	-/AA
rs111848470	C/T	rs58575154	A/G
rs111783220	G/T	rs58467558	C/G
rs111776118	-/A	rs58131227	A/G
rs111756183	A/G	rs57976512	-/A
rs111719142	G/T	rs57940453	A/G
rs111698902	A/G	rs57936835	-/CACACACA
rs111695190	C/T	rs57607171	C/T
rs111624368	C/T	rs57444625	-/A
rs111620465	A/G	rs1049384	C/T
rs111602240	G/T	rs775208	C/T
rs111598897	A/G	rs775205	G/T
rs111583584	A/G	rs775204	C/T
rs111567777	G/T	rs74570457	A/C
rs111539662	-/A	rs74508832	C/T
rs111461416	C/T	rs74446936	A/G
rs111435738	A/G	rs74218028	A/G
rs111426337	A/T	rs74024189	A/G
rs111421609	A/G	rs74024188	A/C
rs111412304	A/T	rs74024187	C/T
rs111338842	A/G	rs74024186	C/T

<b>Table A</b> <b>Human Alanyl tRNA synthetase SNPs</b>			
<b>Gene Bank Accession Number</b>	<b>Nucleotide change</b>	<b>Gene Bank Accession Number</b>	<b>Nucleotide change</b>
rs111333360	C/T	rs74024185	C/T
rs111315764	A/G	rs74024184	A/G
rs80257731	C/G	rs74024183	C/T
rs80256286	A/G	rs74024182	A/G
rs80228853	A/C	rs74024180	C/T
rs80222640	A/G	rs73575193	C/T
rs80200401	A/C	rs73575188	C/T
rs80166718	A/G	rs72790626	C/G
rs80140241	G/T	rs72790624	G/T
rs80062047	C/G	rs72427279	-/G
rs79944605	G/T	rs72289034	- /GCCCATCAGAAAATCC A
rs79879166	C/T	rs72275659	- /AAAAAAAAAAAAAAAAAA AA
rs79809458	A/G	rs72253030	-/TTTTTT
rs79785372	A/G	rs72251314	-/CACACACA
rs79355112	- /AGAAAAAAAAAAAA AAAAAAAAAAAA	rs72185281	-/T
rs79296187	C/T	rs72183811	-/ACACACAC
rs79277862	A/C	rs72172097	-/ACACACAC
rs79100623	A/T	rs72095566	-/T
rs79053418	A/G	rs71943340	-/ACACACAC
rs79022911	C/T	rs71935559	-/A
rs79010592	G/T	rs71928705	-/AC
rs78893985	A/G	rs71912498	-/TT
rs78862566	A/G	rs71847937	-/AA
rs78835148	A/G	rs71455236	C/T
rs78771134	A/T	rs71455235	C/G
rs78570051	A/C	rs71385652	CA/TG
rs78523270	A/G	rs71385651	-/AA
rs78479085	C/T	rs71227275	
rs78344434	A/G	rs71151181	-/T
rs78325403	A/C	rs775203	A/G
rs78298397	A/C	rs775202	C/T
rs78271061	A/C	rs775201	C/T
rs78093603	A/C	rs775200	C/T
rs77982431	A/C	rs775199	A/G
rs77814513	A/G	rs775198	A/G
rs77790607	C/T	rs775197	C/G

<b>Table A</b> <b>Human Alanyl tRNA synthetase SNPs</b>			
<b>Gene Bank Accession Number</b>	<b>Nucleotide change</b>	<b>Gene Bank Accession Number</b>	<b>Nucleotide change</b>
rs77753666	G/T		
rs77749002	G/T		
rs77543430	G/T		
rs77279473	C/T		
rs77154298	A/C		
rs77145843	A/G		
rs76884167	A/G		
rs76856995	A/T		
rs76775711	C/T		
rs76596965	G/T		
rs76446904	C/G		
rs76379088	A/G		
rs76305777	A/C		
rs76247508	G/T		
rs76088863	C/G		
rs75989036	C/G		
rs75949713	A/G		
rs75942139	A/C		
rs75881159	A/C		
rs75655605	C/G		
rs75604742	A/C		
rs75500409	A/C		
rs75414419	A/G		
rs75209821	A/G		
rs75117375	A/G		
rs75106906	A/C		
rs75055059	A/C		
rs75033663	G/T		
rs75029357	C/G		
rs74845494	C/T		
rs74835675	C/T		
rs74817799	C/T		
rs74733646	A/T		
rs74732192	A/G		

**[00130]** In certain embodiments, a polypeptide variant is distinguished from a reference polypeptide by one or more substitutions, which may be conservative or non-conservative, as described herein and known in the art. In certain embodiments, the polypeptide variant comprises conservative substitutions and, in this regard, it is well understood in the art that some amino acids may be changed to others with broadly similar properties without changing the nature of the activity of the polypeptide.

**[00131]** In certain embodiments, a variant polypeptide includes an amino acid sequence having at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or more sequence identity or similarity to a corresponding sequence of an AARS reference polypeptide, as described herein, and substantially retains the non-canonical activity of that reference polypeptide. Also included are sequences differing from the reference AARS sequences by the addition, deletion, or substitution of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150 or more amino acids but which retain the properties of the reference AARS polypeptide. In certain embodiments, the amino acid additions or deletions occur at the C-terminal end and/or the N-terminal end of the AARS reference polypeptide. In certain embodiments, the amino acid additions include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50 or more wild-type residues (*i.e.*, from the corresponding full-length AARS polypeptide) that are proximal to the C-terminal end and/or the N-terminal end of the AARS reference polypeptide.

**[00132]** In certain embodiments, variant polypeptides differ from the corresponding AARS reference sequences by at least 1% but less than 20%, 15%, 10% or 5% of the residues. (If this comparison requires alignment, the sequences should be aligned for maximum similarity. “Looped” out sequences from deletions or insertions, or mismatches, are considered differences.) The differences are, suitably, differences or changes at a non-essential residue or a conservative substitution. In certain embodiments, the molecular weight of a variant AARS polypeptide differs from that of the AARS reference polypeptide by about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, or more.

**[00133]** Also included are biologically active “fragments” of the AARS reference polypeptides, *i.e.*, biologically active fragments of the AARS protein fragments. Representative biologically active fragments generally participate in an interaction, *e.g.*, an intramolecular or an inter-molecular interaction. An inter-molecular interaction can be a specific binding interaction or an enzymatic interaction. An inter-molecular interaction can be between an AARS polypeptide and a cellular binding partner, such as a cellular receptor or other host molecule that participates in the non-canonical activity of the AARS polypeptide. In some embodiments, AARS proteins, variants, and biologically active fragments thereof, bind to one or more cellular binding partners with an affinity of at least about 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, or 50 nM. The binding affinity of an AARS protein fragment for a selected cellular binding partner, particularly a binding partner that participates in a non-

canonical activity, is typically stronger than that of the AARS protein fragment's corresponding full-length AARS polypeptide, by at least about 1.5x, 2x, 2.5x, 3x, 3.5x, 4x, 4.5x, 5x, 6x, 7x, 8x, 9x, 10x, 15x, 20x, 25x, 30x, 40x, 50x, 60x, 70x, 80x, 90x, 100x, 200x, 300x, 400x, 500x, 600x, 700x, 800x, 900x, 1000x or more (including all integers in between). The binding affinity of an AARS protein fragment for a binding partner that participates in at least one canonical activity of an AARS is typically weaker than that of the AARS protein fragment's corresponding full-length AARS polypeptide, by at least about 1.5x, 2x, 2.5x, 3x, 3.5x, 4x, 4.5x, 5x, 6x, 7x, 8x, 9x, 10x, 15x, 20x, 25x, 30x, 40x, 50x, 60x, 70x, 80x, 90x, 100x, 200x, 300x, 400x, 500x, 600x, 700x, 800x, 900x, 1000x or more.

**[00134]** Typically, biologically active fragments comprise a domain or motif with at least one activity of an AARS reference polypeptide and may include one or more (and in some cases all) of the various active domains, and include fragments having a non-canonical activity. In some cases, biologically active fragments of an AARS polypeptide have a biological activity that is unique to the particular, truncated fragment, such that the full-length AARS polypeptide may not have that activity. In certain cases, the biological activity may be revealed by separating the biologically active AARS polypeptide fragment from the other full-length AARS polypeptide sequences, or by altering certain residues of the full-length AARS wild-type polypeptide sequence to unmask the biologically active domains.

**[00135]** A biologically active fragment of an AARS reference polypeptide can be a polypeptide fragment which is, for example, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 450, 500, 550, 600, 650, 700, 750 or more contiguous or non-contiguous amino acids, including all integers (*e.g.*, 101, 102, 103) and ranges (*e.g.*, 50-100, 50-150, 50-200) in between, of the amino acid sequences set forth in any one of the AARS reference polypeptides described herein, but typically exclude the full-length AARS. In certain embodiments, a biologically active fragment comprises a non-canonical activity-related sequence, domain, or motif. In certain embodiments, the C-terminal or N-terminal region of any AARS reference polypeptide may be truncated by about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, or 700 or more amino acids, or by about 10-50, 20-50, 50-100, 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, 450-500, 500-550, 550-600, 600-650, 650-700 or more amino acids, including all integers and ranges in between (*e.g.*, 101, 102, 103, 104, 105), so long as the truncated AARS polypeptide retains the non-

canonical activity of the reference polypeptide. Typically, the biologically-active fragment has no less than about 1%, about 5 %, about 10%, about 25%, or about 50% of an activity of the biologically-active (*i.e.*, non-canonical activity) AARS reference polypeptide from which it is derived. Exemplary methods for measuring such non-canonical activities are described in the Examples.

[00136] As noted above, an AARS polypeptide may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of an AARS reference polypeptide can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, Kunkel (1985, *Proc. Natl. Acad. Sci. USA* 82: 488-492), Kunkel *et al.*, (1987, *Methods in Enzymol*, 154: 367-382), U.S. Pat. No. 4,873,192, Watson, J. D. *et al.*, (“Molecular Biology of the Gene”, Fourth Edition, Benjamin/Cummings, Menlo Park, Calif., 1987) and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff *et al.*, (1978) Atlas of Protein Sequence and Structure (Natl. Biomed. Res. Found., Washington, D.C.).

[00137] Similarly it is within the skill in the art to address and / or mitigate immunogenicity concerns if they arise using an AARS polypeptide, *e.g.*, by the use of automated computer recognition programs to identify potential T cell epitopes, and directed evolution approaches to identify less immunogenic forms.

[00138] Methods for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property are known in the art. Such methods are adaptable for rapid screening of the gene libraries generated by combinatorial mutagenesis of AARS polypeptides. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify AARS polypeptide variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave *et al.*, (1993) *Protein Engineering*, 6: 327-331). Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be desirable as discussed in more detail below.

[00139] Biologically active truncated and/or variant AARS polypeptides may contain conservative amino acid substitutions at various locations along their sequence, as compared to a reference AARS amino acid residue. Additionally, naturally occurring variants of AARS proteins have been sequenced, and are known in the art to be at least partially functionally interchangeable. It would thus be a routine matter to select an amino acid position to introduce a conservative, or non conservative mutation into an



AARS polypeptide based on naturally occurring sequence variation among the known AARS protein homologues, orthologs, and naturally-occurring isoforms of human as well as other species of an AARS protein.

[00140] A “conservative amino acid substitution” is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art, which can be generally sub-classified as follows:

[00141] Acidic: The residue has a negative charge due to loss of H ion at physiological pH and the residue is attracted by aqueous solution so as to seek the surface positions in the conformation of a peptide in which it is contained when the peptide is in aqueous medium at physiological pH. Amino acids having an acidic side chain include glutamic acid and aspartic acid.

[00142] Basic: The residue has a positive charge due to association with H ion at physiological pH or within one or two pH units thereof (*e.g.*, histidine) and the residue is attracted by aqueous solution so as to seek the surface positions in the conformation of a peptide in which it is contained when the peptide is in aqueous medium at physiological pH. Amino acids having a basic side chain include arginine, lysine and histidine.

[00143] Charged: The residues are charged at physiological pH and, therefore, include amino acids having acidic or basic side chains (*i.e.*, glutamic acid, aspartic acid, arginine, lysine and histidine).

[00144] Hydrophobic: The residues are not charged at physiological pH and the residue is repelled by aqueous solution so as to seek the inner positions in the conformation of a peptide in which it is contained when the peptide is in aqueous medium. Amino acids having a hydrophobic side chain include tyrosine, valine, isoleucine, leucine, methionine, phenylalanine and tryptophan.

[00145] Neutral/polar: The residues are not charged at physiological pH, but the residue is not sufficiently repelled by aqueous solutions so that it would seek inner positions in the conformation of a peptide in which it is contained when the peptide is in aqueous medium. Amino acids having a neutral/polar side chain include asparagine, glutamine, cysteine, histidine, serine and threonine.

[00146] This description also characterizes certain amino acids as “small” since their side chains are not sufficiently large, even if polar groups are lacking, to confer hydrophobicity. With the exception of proline, “small” amino acids are those with four carbons or less when at least one polar group is on the side chain and three carbons or less when not. Amino acids having a small side chain include glycine, serine, alanine and threonine. The gene-encoded secondary amino acid proline is a special case due to

its known effects on the secondary conformation of peptide chains. The structure of proline differs from all the other naturally-occurring amino acids in that its side chain is bonded to the nitrogen of the  $\alpha$ -amino group, as well as the  $\alpha$ -carbon. Several amino acid similarity matrices are known in the art (*see e.g.*, PAM120 matrix and PAM250 matrix as disclosed for example by Dayhoff *et al.*, 1978, A model of evolutionary change in proteins). Matrices for determining distance relationships In M. O. Dayhoff, (ed.), Atlas of protein sequence and structure, Vol. 5, pp. 345-358, National Biomedical Research Foundation, Washington DC; and by Gonnet *et al.*, (*Science*, 256: 14430-1445, 1992), however, include proline in the same group as glycine, serine, alanine and threonine. Accordingly, for the purposes of the present invention, proline is classified as a “small” amino acid.

[00147] The degree of attraction or repulsion required for classification as polar or nonpolar is arbitrary and, therefore, amino acids specifically contemplated by the invention have been classified as one or the other. Most amino acids not specifically named can be classified on the basis of known behavior.

[00148] Amino acid residues can be further sub-classified as cyclic or non-cyclic, and aromatic or non-aromatic, self-explanatory classifications with respect to the side-chain substituent groups of the residues, and as small or large. The residue is considered small if it contains a total of four carbon atoms or less, inclusive of the carboxyl carbon, provided an additional polar substituent is present; three or less if not. Small residues are, of course, always non-aromatic. Dependent on their structural properties, amino acid residues may fall in two or more classes. For the naturally-occurring protein amino acids, sub-classification according to this scheme is presented in **Table B**.

**Table B: Amino acid sub-classification**

Sub-classes	Amino acids
Acidic	Aspartic acid, Glutamic acid
Basic	Noncyclic: Arginine, Lysine; Cyclic: Histidine
Charged	Aspartic acid, Glutamic acid, Arginine, Lysine, Histidine
Small	Glycine, Serine, Alanine, Threonine, Proline
Polar/neutral	Asparagine, Histidine, Glutamine, Cysteine, Serine, Threonine
Polar/large	Asparagine, Glutamine
Hydrophobic	Tyrosine, Valine, Isoleucine, Leucine, Methionine, Phenylalanine, Tryptophan
Aromatic	Tryptophan, Tyrosine, Phenylalanine
Residues that influence	Glycine and Proline

Sub-classes	Amino acids
chain orientation	

[00149] Conservative amino acid substitution also includes groupings based on side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulphur-containing side chains is cysteine and methionine. For example, it is reasonable to expect that replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid will not have a major effect on the properties of the resulting variant polypeptide. Whether an amino acid change results in a functional truncated and/or variant AARS polypeptide can readily be determined by assaying its non-canonical activity, as described herein. Conservative substitutions are shown in Table C under the heading of exemplary substitutions. Amino acid substitutions falling within the scope of the invention, are, in general, accomplished by selecting substitutions that do not differ significantly in their effect on maintaining (a) the structure of the peptide backbone in the area of the substitution, (b) the charge or hydrophobicity of the molecule at the target site, (c) the bulk of the side chain, or (d) the biological function. After the substitutions are introduced, the variants are screened for biological activity.

**Table C: Exemplary Amino Acid Substitutions**

Original Residue	Exemplary Substitutions	Preferred Substitutions
Ala	Val, Leu, Ile	Val
Arg	Lys, Gln, Asn	Lys
Asn	Gln, His, Lys, Arg	Gln
Asp	Glu	Glu
Cys	Ser	Ser
Gln	Asn, His, Lys,	Asn
Glu	Asp, Lys	Asp
Gly	Pro	Pro
His	Asn, Gln, Lys, Arg	Arg
Ile	Leu, Val, Met, Ala, Phe, Norleu	Leu

Original Residue	Exemplary Substitutions	Preferred Substitutions
Leu	Norleu, Ile, Val, Met, Ala, Phe	Ile
Lys	Arg, Gln, Asn	Arg
Met	Leu, Ile, Phe	Leu
Phe	Leu, Val, Ile, Ala	Leu
Pro	Gly	Gly
Ser	Thr	Thr
Thr	Ser	Ser
Trp	Tyr	Tyr
Tyr	Trp, Phe, Thr, Ser	Phe
Val	Ile, Leu, Met, Phe, Ala, Norleu	Leu

[00150] Alternatively, similar amino acids for making conservative substitutions can be grouped into three categories based on the identity of the side chains. The first group includes glutamic acid, aspartic acid, arginine, lysine, histidine, which all have charged side chains; the second group includes glycine, serine, threonine, cysteine, tyrosine, glutamine, asparagine; and the third group includes leucine, isoleucine, valine, alanine, proline, phenylalanine, tryptophan, methionine, as described in Zubay, G., *Biochemistry*, third edition, Wm.C. Brown Publishers (1993).

[00151] Thus, a predicted non-essential amino acid residue in a truncated and/or variant AARS polypeptide is typically replaced with another amino acid residue from the same side chain family. Alternatively, mutations can be introduced randomly along all or part of an AARS coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for an activity of the parent polypeptide to identify mutants which retain that activity. Following mutagenesis of the coding sequences, the encoded peptide can be expressed recombinantly and the activity of the peptide can be determined. A “non-essential” amino acid residue is a residue that can be altered from the reference sequence of an embodiment polypeptide without abolishing or substantially altering one or more of its activities. Suitably, the alteration does not substantially abolish one of these activities, for example, the activity is at least 20%, 40%, 60%, 70% or 80% 100%, 500%, 1000% or more of the reference AARS sequence. An “essential” amino acid residue is a residue that, when altered from the reference sequence of an AARS polypeptide, results in abolition of an activity of the parent molecule such that less than 20% of the reference activity is present. For example, such essential amino acid residues include those that are conserved in AARS polypeptides across different species, including those sequences that are conserved in the active binding site(s) or motif(s) of AARS polypeptides from various sources.

[00152] In general, polypeptides and fusion polypeptides (as well as their encoding polynucleotides) are isolated. An “isolated” polypeptide or polynucleotide is one that is removed from its original environment. For example, a naturally-occurring protein is isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably, such polypeptides are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure. A polynucleotide is considered to be isolated if, for example, it is cloned into a vector that is not a part of the natural environment.

[00153] Certain embodiments also encompass dimers of AARS polypeptides. Dimers may include, for example, homodimers between two identical AARS polypeptides, heterodimers between two different AARS polypeptides (*e.g.*, a full-length YRS polypeptide and a truncated YRS polypeptide; a truncated YRS polypeptide and a truncated WRS polypeptide), and/or heterodimers between an AARS polypeptide and a heterologous polypeptide. Certain heterodimers, such as those between an AARS polypeptide and a heterologous polypeptide, may be bi-functional, as described herein.

[00154] Also included are monomers of AARS polypeptides, including isolated AARS polypeptides monomers that do not substantially dimerize with a second AARS polypeptide, whether due to one or more substitutions, truncations, deletions, additions, chemical modifications, or a combination of these alterations. In certain embodiments, monomeric AARS polypeptides possess biological activities, including non-canonical activities, which are not possessed by dimeric or multimeric AARS polypeptide complexes.

[00155] Certain embodiments of the present invention also contemplate the use of modified AARS polypeptides, including modifications that improved the desired characteristics of an AARS polypeptide, as described herein. Modifications of AARS polypeptides of the invention include chemical and/or enzymatic derivatizations at one or more constituent amino acid, including side chain modifications, backbone modifications, and N- and C-terminal modifications including acetylation, hydroxylation, methylation, amidation, and the attachment of carbohydrate or lipid moieties, cofactors, and the like. Exemplary modifications also include pegylation of an AARS polypeptide (*see, e.g.*, Veronese and Harris, *Advanced Drug Delivery Reviews* 54: 453-456, 2002; and Pasut *et al.*, *Expert Opinion. Ther. Patents* 14(6) 859-894 2004, both herein incorporated by reference).

[00156] PEG is a well-known polymer having the properties of solubility in water and in many organic solvents, lack of toxicity, and lack of immunogenicity. It is also clear, colorless, odorless, and chemically stable. For these reasons and others, PEG has been selected as the preferred polymer for attachment, but it has been employed solely

for purposes of illustration and not limitation. Similar products may be obtained with other water-soluble polymers, including without limitation; polyvinyl alcohol, other poly(alkylene oxides) such as poly(propylene glycol) and the like, poly(oxyethylated polyols) such as poly(oxyethylated glycerol) and the like, carboxymethylcellulose, dextran, polyvinyl alcohol, polyvinyl pyrrolidone, poly-1,3- dioxolane, poly-1,3,6-trioxane, ethylene/maleic anhydride, and polyaminoacids. One skilled in the art will be able to select the desired polymer based on the desired dosage, circulation time, resistance to proteolysis, and other considerations.

[00157] In particular a wide variety of PEG derivatives are both available and suitable for use in the preparation of PEG-conjugates. For example, NOF Corp.'s PEG reagents sold under the trademark SUNBRIGHT® Series provides numerous PEG derivatives, including methoxypolyethylene glycols and activated PEG derivatives such as methoxy-PEG amines, maleimides, N-hydroxysuccinimide esters, and carboxylic acids, for coupling by various methods to the N-terminal, C-terminal or any internal amino acid of the AARS polypeptide. Nektar Therapeutics' Advanced PEGylation technology also offers diverse PEG-coupling technologies to potentially improve the safety and efficacy of an AARS polypeptide based therapeutic.

[00158] A search of patents, published patent applications, and related publications will also provide those skilled in the art reading this disclosure with significant possible PEG-coupling technologies and PEG-derivatives. For example, US Pat. Nos. 6,436,386; 5,932,462; 5,900,461; 5,824,784; and 4,904,584; the contents of which are incorporated by reference in their entirety, describe such technologies and derivatives, and methods for their manufacture.

[00159] In certain aspects, chemoselective ligation technology may be utilized to modify AARS polypeptides of the invention, such as by attaching polymers in a site-specific and controlled manner. Such technology typically relies on the incorporation of chemoselective anchors into the protein backbone by either chemical or recombinant means, and subsequent modification with a polymer carrying a complementary linker. As a result, the assembly process and the covalent structure of the resulting protein-polymer conjugate may be controlled, enabling the rational optimization of drug properties, such as efficacy and pharmacokinetic properties (*see, e.g., Kochendoerfer, Current Opinion in Chemical Biology* 9:555-560, 2005).

[00160] In other embodiments, fusion proteins of AARS polypeptide to other proteins are also included, and these fusion proteins may increase the AARS polypeptide's biological activity, secretion, targeting, biological life, ability to penetrate cellular membranes, or the blood brain barrier, or pharmacokinetic properties. Examples of fusion proteins that improve pharmacokinetic properties ("PK modifiers")

include without limitation, fusions to human albumin (Osborn et al.: *Eur. J. Pharmacol.* 456(1-3): 149-158, (2002)), antibody Fc domains, poly Glu or poly Asp sequences, and transferrin. Additionally, fusion with conformationally disordered polypeptide sequences composed of the amino acids Pro, Ala, and Ser ('PASylation') or hydroxyethyl starch (sold under the trademark HESYLATION®) provides a simple way to increase the hydrodynamic volume of the AARS polypeptide. This additional extension adopts a bulky random structure, which significantly increases the size of the resulting fusion protein. By this means the typically rapid clearance of smaller AARS polypeptides via kidney filtration is retarded by several orders of magnitude. Additionally use of Ig G fusion proteins has also been shown to enable some fusion protein proteins to penetrate the blood brain barrier (Fu et al., (2010) *Brain Res.* 1352:208-13).

[00161] Examples of fusion proteins that improve penetration across cellular membranes include fusions to membrane translocating sequences. In this context, the term “membrane translocating sequences” refers to naturally occurring and synthetic amino acid sequences that are capable of membrane translocation across a cellular membrane. Representative membrane translocating sequences include those based on the naturally occurring membrane translocating sequences derived from the Tat protein, and homeotic transcription protein Antennapedia, as well as synthetic membrane translocating sequences based in whole or part on poly Arginine and Lysine residues. Representative membrane translocating sequences include for example those disclosed in the following patents, US5,652,122; US 5,670,617; US5,674,980; US5,747,641; US5,804,604; US6,316,003; US7,585,834; US7,312,244; US7,279,502; US7,229,961; US7,169,814; US7,453,011; US7,235,695; US6,982,351; US6,605,115; US7,306,784; US7,306,783; US6,589,503; US6,348,185; US6,881,825; US7,431,915; WO0074701A2; WO2007111993A2; WO2007106554A2; WO02069930A1; WO03049772A2; WO03106491A2; and WO2008063113A1.

[00162] It will be appreciated that a flexible molecular linker (or spacer) optionally may be interposed between, and covalently join, the AARS polypeptide and any of the fusion proteins disclosed herein.

[00163] Additionally in some embodiments, the AARS polypeptide can include synthetic, or naturally occurring secretion signal sequences, derived from other well characterized secreted proteins. In some embodiments such proteins, may be processed by proteolytic cleavage to form the AARS polypeptide in situ. Such fusions proteins include for example fusions of AARS polypeptide to ubiquitin to provide a new N-terminal amino acid, or the use of a secretion signal to mediate high level secretion of

the AARS polypeptide into the extracellular medium, or N, or C-terminal epitope tags to improve purification or detection.

[00164] The AARS polypeptides described herein may be prepared by any suitable procedure known to those of skill in the art, such as by recombinant techniques. In addition to recombinant production methods, polypeptides of the invention may be produced by direct peptide synthesis using solid-phase techniques (Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963)). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Alternatively, various fragments may be chemically synthesized separately and combined using chemical methods to produce the desired molecule.

#### IV. AARS POLYNUCLEOTIDES

[00165] Embodiments of the present invention include polynucleotides that encode one or more newly identified protein fragments of an aminoacyl-tRNA synthetase (AARS), in addition to complements, variants, and fragments thereof. In certain embodiments, an AARS polynucleotide encodes all or a portion of the AARS polypeptide reference sequence(s) as set forth in Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9, which represent splice variants, proteolytic fragments, or other type of fragments of Alanine tRNA synthetase. Certain embodiments include polynucleotides, encoding polypeptides or proteins that comprise the sequence of one or more splice junctions of those splice variants, in addition to complements, variants, and fragments thereof. In certain embodiments, typically due to the singular nature of a selected AARS splice variant, which combines exons in a new or exceptional way, the AARS polynucleotide references sequences comprise a unique or exceptional splice junction. Certain embodiments exclude a corresponding full-length AARS polynucleotide.

[00166] Also included within the AARS polynucleotides of the present invention are primers, probes, antisense oligonucleotides, and RNA interference agents that comprise all or a portion of these reference polynucleotides, which are complementary to all or a portion of these reference polynucleotides, or which specifically hybridize to these reference polynucleotides, as described herein.

[00167] The term “polynucleotide” or “nucleic acid” as used herein designates mRNA, RNA, cRNA, cDNA or DNA. The term typically refers to polymeric form of nucleotides of at least 10 bases in length, either ribonucleotides or deoxynucleotides or a modified form of either type of nucleotide. The term includes single and double stranded forms of DNA. The terms “DNA” and “polynucleotide” and “nucleic acid” refer to a DNA molecule that has been isolated free of total genomic DNA of a



particular species. Therefore, an isolated DNA segment encoding a polypeptide refers to a DNA segment that contains one or more coding sequences yet is substantially isolated away from, or purified free from, total genomic DNA of the species from which the DNA segment is obtained. Also included are non-coding polynucleotides (*e.g.*, primers, probes, oligonucleotides), which do not encode an AARS polypeptide. Included within the terms “DNA segment” and “polynucleotide” are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phagemids, phage, viruses, and the like.

**[00168]** Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials. Hence, the polynucleotides of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably.

**[00169]** It is therefore contemplated that a polynucleotide fragment of almost any length may be employed; with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. Included are polynucleotides of about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 41, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 220, 240, 260, 270, 280, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000 or more (including all integers in between) bases in length, including any portion or fragment (*e.g.*, greater than about 6, 7, 8, 9, or 10 nucleotides in length) of an AARS reference polynucleotide (*e.g.*, base number X-Y, in which X is about 1-3000 or more and Y is about 10-3000 or more), or its complement.

**[00170]** Embodiments of the present invention also include “variants” of the AARS reference polynucleotide sequences. Polynucleotide “variants” may contain one or more substitutions, additions, deletions and/or insertions in relation to a reference polynucleotide. Generally, variants of an AARS reference polynucleotide sequence may have at least about 30%, 40% 50%, 55%, 60%, 65%, 70%, generally at least about 75%, 80%, 85%, desirably about 90% to 95% or more, and more suitably about 98% or more sequence identity to that particular nucleotide sequence as determined by sequence alignment programs described elsewhere herein using default parameters. In certain embodiments, variants may differ from a reference sequence by about 1, 2, 3, 4,

5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 41, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100 (including all integers in between) or more bases. In certain embodiments, such as when the polynucleotide variant encodes an AARS polypeptide having a non-canonical activity, the desired activity of the encoded AARS polypeptide is not substantially diminished relative to the unmodified polypeptide. The effect on the activity of the encoded polypeptide may generally be assessed as described herein.

**[00171]** Certain embodiments include polynucleotides that hybridize to a reference AARS polynucleotide sequence, or to their complements, under stringency conditions described below. As used herein, the term “hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions” describes conditions for hybridization and washing. Guidance for performing hybridization reactions can be found in Ausubel *et al.*, (1998, *supra*), Sections 6.3.1-6.3.6. Aqueous and non-aqueous methods are described in that reference and either can be used.

**[00172]** Reference herein to low stringency conditions include and encompass from at least about 1% v/v to at least about 15% v/v formamide and from at least about 1 M to at least about 2 M salt for hybridization at 42° C, and at least about 1 M to at least about 2 M salt for washing at 42° C. Low stringency conditions also may include 1% Bovine Serum Albumin (BSA), 1 mM EDTA, 0.5 M NaHPO<sub>4</sub> (pH 7.2), 7% SDS for hybridization at 65°C, and (i) 2 × SSC, 0.1% SDS; or (ii) 0.5% BSA, 1 mM EDTA, 40 mM NaHPO<sub>4</sub> (pH 7.2), 5% SDS for washing at room temperature. One embodiment of low stringency conditions includes hybridization in 6 × sodium chloride/sodium citrate (SSC) at about 45°C, followed by two washes in 0.2 × SSC, 0.1% SDS at least at 50° C (the temperature of the washes can be increased to 55° C for low stringency conditions).

**[00173]** Medium stringency conditions include and encompass from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5 M to at least about 0.9 M salt for hybridization at 42° C, and at least about 0.1 M to at least about 0.2 M salt for washing at 55° C. Medium stringency conditions also may include 1% Bovine Serum Albumin (BSA), 1 mM EDTA, 0.5 M NaHPO<sub>4</sub> (pH 7.2), 7% SDS for hybridization at 65° C, and (i) 2 × SSC, 0.1% SDS; or (ii) 0.5% BSA, 1 mM EDTA, 40 mM NaHPO<sub>4</sub> (pH 7.2), 5% SDS for washing at 60-65° C. One embodiment of medium stringency conditions includes hybridizing in 6 × SSC at about 45°C, followed by one or more washes in 0.2 × SSC, 0.1% SDS at 60°C. High stringency conditions include and encompass from at least about 31% v/v to at least about 50% v/v formamide and from about 0.01 M to about 0.15 M salt for hybridization at 42° C, and about 0.01 M to about 0.02 M salt for washing at 55° C.

[00174] High stringency conditions also may include 1% BSA, 1 mM EDTA, 0.5 M NaHPO<sub>4</sub> (pH 7.2), 7% SDS for hybridization at 65° C, and (i) 0.2 × SSC, 0.1% SDS; or (ii) 0.5% BSA, 1 mM EDTA, 40 mM NaHPO<sub>4</sub> (pH 7.2), 1% SDS for washing at a temperature in excess of 65° C. One embodiment of high stringency conditions includes hybridizing in 6 × SSC at about 45°C, followed by one or more washes in 0.2 × SSC, 0.1% SDS at 65° C. One embodiment of very high stringency conditions includes hybridizing in 0.5 M sodium phosphate, 7% SDS at 65° C, followed by one or more washes in 0.2 × SSC, 1% SDS at 65° C.

[00175] Other stringency conditions are well known in the art and a skilled artisan will recognize that various factors can be manipulated to optimize the specificity of the hybridization. Optimization of the stringency of the final washes can serve to ensure a high degree of hybridization. For detailed examples, see Ausubel *et al.*, *supra* at pages 2.10.1 to 2.10.16 and Sambrook *et al.* (1989, *supra*) at sections 1.101 to 1.104.

[00176] While stringent washes are typically carried out at temperatures from about 42° C to 68° C, one skilled in the art will appreciate that other temperatures may be suitable for stringent conditions. Maximum hybridization rate typically occurs at about 20° C to 25° C below the T<sub>m</sub> for formation of a DNA-DNA hybrid. It is well known in the art that the T<sub>m</sub> is the melting temperature, or temperature at which two complementary polynucleotide sequences dissociate. Methods for estimating T<sub>m</sub> are well known in the art (see Ausubel *et al.*, *supra* at page 2.10.8).

[00177] In general, the T<sub>m</sub> of a perfectly matched duplex of DNA may be predicted as an approximation by the formula:  $T_m = 81.5 + 16.6 (\log_{10} M) + 0.41 (\%G+C) - 0.63 (\% \text{ formamide}) - (600/\text{length})$  wherein: M is the concentration of Na<sup>+</sup>, preferably in the range of 0.01 molar to 0.4 molar; %G+C is the sum of guanosine and cytosine bases as a percentage of the total number of bases, within the range between 30% and 75% G+C; % formamide is the percent formamide concentration by volume; length is the number of base pairs in the DNA duplex. The T<sub>m</sub> of a duplex DNA decreases by approximately 1° C with every increase of 1% in the number of randomly mismatched base pairs. Washing is generally carried out at T<sub>m</sub> – 15° C for high stringency, or T<sub>m</sub> – 30° C for moderate stringency.

[00178] In one example of a hybridization procedure, a membrane (*e.g.*, a nitrocellulose membrane or a nylon membrane) containing immobilized DNA is hybridized overnight at 42° C in a hybridization buffer (50% deionized formamide, 5 × SSC, 5 × Denhardt's solution (0.1% ficoll, 0.1% polyvinylpyrrolidone and 0.1% bovine serum albumin), 0.1% SDS and 200 mg/mL denatured salmon sperm DNA) containing a labeled probe. The membrane is then subjected to two sequential medium stringency washes (*i.e.*, 2 × SSC, 0.1% SDS for 15 min at 45° C, followed by 2 × SSC, 0.1% SDS

for 15 min at 50° C), followed by two sequential higher stringency washes (*i.e.*, 0.2 × SSC, 0.1% SDS for 12 min at 55° C followed by 0.2 × SSC and 0.1% SDS solution for 12 min at 65-68° C).

[00179] As noted above, certain embodiments relate to AARS polynucleotides that encode an AARS polypeptide. Among other uses, these embodiments may be utilized to recombinantly produce a desired AARS polypeptide or variant thereof, or to express the AARS polypeptide in a selected cell or subject. It will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a polypeptide as described herein. Some of these polynucleotides may bear minimal homology to the nucleotide sequence of any native gene. Nonetheless, polynucleotides that vary due to differences in codon usage are specifically contemplated by the present invention, for example polynucleotides that are optimized for human and/or primate codon selection.

[00180] Therefore, multiple polynucleotides can encode the AARS polypeptides of the invention. Moreover, the polynucleotide sequence can be manipulated for various reasons. Examples include but are not limited to the incorporation of preferred codons to enhance the expression of the polynucleotide in various organisms (see generally Nakamura et al., Nuc. Acid. Res. (2000) 28 (1): 292). In addition, silent mutations can be incorporated in order to introduce, or eliminate restriction sites, decrease the density of CpG dinucleotide motifs (see for example, Kameda et al., Biochem. Biophys. Res. Commun. (2006) 349(4): 1269-1277) or reduce the ability of single stranded sequences to form stem-loop structures: (see, e.g., Zuker M., Nucl. Acid Res. (2003); 31(13): 3406-3415). In addition, mammalian expression can be further optimized by including a Kozak consensus sequence [*i.e.*, (a/g)cc(a/g)ccATGg] at the start codon. Kozak consensus sequences useful for this purpose are known in the art (Manty et al. PNAS 92: 2662-2666 (1995); Mantyh et al. Prot. Exp. & Purif. 6,124 (1995)).

[00181] The polynucleotides of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a polynucleotide fragment of almost any length may be employed; with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

[00182] Polynucleotides and fusions thereof may be prepared, manipulated and/or expressed using any of a variety of well established techniques known and available in the art. For example, polynucleotide sequences which encode polypeptides of the invention, or fusion proteins or functional equivalents thereof, may be used in

recombinant DNA molecules to direct expression of an AARS polypeptide in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express a given polypeptide.

[00183] As will be understood by those of skill in the art, it may be advantageous in some instances to produce polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence. Such polynucleotides are commonly referred to as “codon-optimized.” Any of the polynucleotides described herein may be utilized in a codon-optimized form. In certain embodiments, a polynucleotide can be codon optimized for use in specific bacteria such as *E. coli* or yeast such as *S. cerevisiae* (see, e.g., Burgess-Brown *et al.*, *Protein Expr Purif.* 59:94-102, 2008; Ermolaeva MD (2001) *Curr. Iss. Mol. Biol.* 3 (4) 91-7; Welch *et al.*, *PLoS ONE* 4(9): e7007 doi:10.1371/journal.pone.0007002).

[00184] Moreover, the polynucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter polypeptide encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, expression and/or activity of the gene product.

[00185] According to another aspect of the invention, polynucleotides encoding polypeptides of the invention may be delivered to a subject *in vivo*, e.g., using gene therapy techniques. Gene therapy refers generally to the transfer of heterologous nucleic acids to the certain cells, target cells, of a mammal, particularly a human, with a disorder or conditions for which such therapy is sought. The nucleic acid is introduced into the selected target cells in a manner such that the heterologous DNA is expressed and a therapeutic product encoded thereby is produced.

[00186] Various viral vectors that can be utilized for gene therapy as taught herein include adenovirus, herpes virus, vaccinia, adeno-associated virus (AAV), or, preferably, an RNA virus such as a retrovirus. Preferably, the retroviral vector is a derivative of a murine or avian retrovirus, or is a lentiviral vector. The preferred retroviral vector is a lentiviral vector. Examples of retroviral vectors in which a single foreign gene can be inserted include, but are not limited to: Moloney murine leukemia virus (MoMuLV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV), SIV, BIV, HIV and Rous Sarcoma Virus (RSV). A number of additional retroviral vectors can incorporate multiple genes. All of these vectors can

transfer or incorporate a gene for a selectable marker so that transduced cells can be identified and generated. By inserting a zinc finger derived-DNA binding polypeptide sequence of interest into the viral vector, along with another gene that encodes the ligand for a receptor on a specific target cell, for example, the vector may be made target specific. Retroviral vectors can be made target specific by inserting, for example, a polynucleotide encoding a protein (dimer). Illustrative targeting may be accomplished by using an antibody to target the retroviral vector. Those of skill in the art will know of, or can readily ascertain without undue experimentation, specific polynucleotide sequences which can be inserted into the retroviral genome to allow target specific delivery of the retroviral vector containing the zinc finger-nucleotide binding protein polynucleotide.

[00187] Since recombinant retroviruses are defective, they require assistance in order to produce infectious vector particles. This assistance can be provided, for example, by using helper cell lines that contain plasmids encoding all of the structural genes of the retrovirus under the control of regulatory sequences within the LTR. These plasmids are missing a nucleotide sequence which enables the packaging mechanism to recognize an RNA transcript for encapsulation. Helper cell lines which have deletions of the packaging signal include but are not limited to PSI.2, PA317 and PA12, for example. These cell lines produce empty virions, since no genome is packaged. If a retroviral vector is introduced into such cells in which the packaging signal is intact, but the structural genes are replaced by other genes of interest, the vector can be packaged and vector virion produced. The vector virions produced by this method can then be used to infect a tissue cell line, such as NIH 3T3 cells, to produce large quantities of chimeric retroviral virions.

[00188] “Non-viral” delivery techniques for gene therapy can also be used including, for example, DNA-ligand complexes, adenovirus-ligand-DNA complexes, direct injection of DNA,  $\text{CaPO}_4$  precipitation, gene gun techniques, electroporation, liposomes, lipofection, and the like. Any of these methods are widely available to one skilled in the art and would be suitable for use in the present invention. Other suitable methods are available to one skilled in the art, and it is to be understood that the present invention can be accomplished using any of the available methods of transfection. Lipofection can be accomplished by encapsulating an isolated DNA molecule within a liposomal particle and contacting the liposomal particle with the cell membrane of the target cell. Liposomes are self-assembling, colloidal particles in which a lipid bilayer, composed of amphiphilic molecules such as phosphatidyl serine or phosphatidyl choline, encapsulates a portion of the surrounding media such that the lipid bilayer surrounds a hydrophilic interior. Unilamellar or multilamellar liposomes can be

constructed such that the interior contains a desired chemical, drug, or, as in the instant invention, an isolated DNA molecule.

[00189] In another aspect, polynucleotides encoding polypeptides of the invention may be used to express and delivery an AARS polypeptide via cell therapy. Accordingly in another aspect, the current invention includes a cell therapy for treating a disease or disorder, comprising administering a host cell expressing, or capable of expressing, an AARS polypeptide.

[00190] Cell therapy involves the administration of cells which have been selected, multiplied and pharmacologically treated or altered (i.e. genetically modified) outside of the body (*Bordignon, C. et al, Cell Therapy: Achievements and Perspectives* (1999), *Haematologica*, 84, pp.1110-1149). Such host cells include for example, primary cells, including macrophages, and stem cells which have been genetically modified to express an AARS polypeptide. The aim of cell therapy is to replace, repair or enhance the biological function of damaged tissues or organs.

[00191] The use of transplanted cells has been investigated for the treatment of numerous endocrine disorders such as anemia and dwarfism, hematological disorders, kidney and liver failure, pituitary and CNS deficiencies and diabetes mellitus (*Uludag et al., Technology of Mammalian Cell Encapsulation* (2000), *Advanced Drug Delivery Reviews*, 42, pp. 29-64). Transplanted cells may function by releasing bioactive compounds such as an AARS polypeptide of the invention, to replace endogenous AARS polypeptides which are absent or produced in insufficient quantities in an effected system.

[00192] Embodiments of the present invention also include oligonucleotides, whether for detection, amplification, antisense therapies, or other purpose. For these and related purposes, the term “oligonucleotide” or “oligo” or “oligomer” is intended to encompass a singular “oligonucleotide” as well as plural “oligonucleotides,” and refers to any polymer of two or more of nucleotides, nucleosides, nucleobases or related compounds used as a reagent in the amplification methods of the present invention, as well as subsequent detection methods. The oligonucleotide may be DNA and/or RNA and/or analogs thereof.

[00193] The term oligonucleotide does not necessarily denote any particular function to the reagent, rather, it is used generically to cover all such reagents described herein. An oligonucleotide may serve various different functions, *e.g.*, it may function as a primer if it is capable of hybridizing to a complementary strand and can further be extended in the presence of a nucleic acid polymerase, it may provide a promoter if it contains a sequence recognized by an RNA polymerase and allows for transcription, and it may function to prevent hybridization or impede primer extension if

appropriately situated and/or modified. An oligonucleotide may also function as a probe, or an antisense agent. An oligonucleotide can be virtually any length, limited only by its specific function, *e.g.*, in an amplification reaction, in detecting an amplification product of the amplification reaction, or in an antisense or RNA interference application. Any of the oligonucleotides described herein can be used as a primer, a probe, an antisense oligomer, or an RNA interference agent.

**[00194]** The term “primer” as used herein refers to a single-stranded oligonucleotide capable of acting as a point of initiation for template-directed DNA synthesis under suitable conditions defined, for example, by buffer and temperature, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as a DNA or RNA polymerase or reverse transcriptase. The length of the primer, in any given case, depends on, for example, the intended use of the primer, and generally ranges from about 15 to 30 nucleotides, although shorter and longer primers may be used. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with such template. The primer site is the area of the template to which a primer hybridizes. The primer pair is a set of primers including a 5' upstream primer that hybridizes with the 5' end of the sequence to be amplified and a 3' downstream primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

**[00195]** The term “probe” as used herein includes a surface-immobilized or soluble but capable of being immobilized molecule that can be recognized by a particular target. *See, e.g.*, U.S. Patent No. 6,582,908 for an example of arrays having all possible combinations of probes with 10, 12, and more bases. Probes and primers as used herein typically comprise at least 10-15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 nucleotides of an AARS reference sequence or its complement. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the knowledge in the art and the specification, including the tables, figures, and Sequence Listing, may be used.

**[00196]** Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview N.Y.; Ausubel, F. M. *et al.* (1987) *Current Protocols in Molecular Biology*, Greene Publ. Assoc. & Wiley-Intersciences, New York N.Y.; Innis, M. *et al.* (1990) *PCR Protocols. A Guide to Methods and Applications*, Academic Press, San Diego Calif. PCR primer pairs can be



derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge Mass.).

[00197] Oligonucleotides for use as primers or probes may be selected using software known in the art. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas Tex.) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope.

[00198] The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge Mass.) allows the user to input a “mispriming library,” in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user’s specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described herein.

[00199] In certain embodiments, oligonucleotides can be prepared by stepwise solid-phase synthesis, employing methods detailed in the references cited above, and below with respect to the synthesis of oligonucleotides having a mixture or uncharged and cationic backbone linkages. In some cases, it may be desirable to add additional chemical moieties to the oligonucleotide, *e.g.*, to enhance pharmacokinetics or to facilitate capture or detection of the compound. Such a moiety may be covalently attached, typically to a terminus of the oligomer, according to standard synthetic

methods. For example, addition of a polyethyleneglycol moiety or other hydrophilic polymer, *e.g.*, one having 10-100 monomeric subunits, may be useful in enhancing solubility. One or more charged groups, *e.g.*, anionic charged groups such as an organic acid, may enhance cell uptake.

**[00200]** A variety of detectable molecules may be used to render an oligonucleotide, or protein detectable, such as a radioisotopes, fluorochromes, dyes, enzymes, nanoparticles, chemiluminescent markers, biotin, or other monomer known in the art that can be detected directly (*e.g.*, by light emission) or indirectly (*e.g.*, by binding of a fluorescently-labeled antibody).

**[00201]** Radioisotopes provide examples of detectable molecules that can be utilized in certain aspects of the present invention. Several radioisotopes can be used as detectable molecules for labeling nucleotides or proteins, including, for example,  $^{32}\text{P}$ ,  $^{33}\text{P}$ ,  $^{35}\text{S}$ ,  $^3\text{H}$ , and  $^{125}\text{I}$ . These radioisotopes have different half-lives, types of decay, and levels of energy which can be tailored to match the needs of a particular protocol. For example,  $^3\text{H}$  is a low energy emitter which results in low background levels, however this low energy also results in long time periods for autoradiography. Radioactively labeled ribonucleotides, deoxyribonucleotides and amino acids are commercially available. Nucleotides are available that are radioactively labeled at the first, or  $\alpha$ , phosphate group, or the third, or  $\gamma$ , phosphate group. For example, both [ $\alpha$  -  $^{32}\text{P}$ ] dATP and [ $\gamma$  -  $^{32}\text{P}$ ] dATP are commercially available. In addition, different specific activities for radioactively labeled nucleotides are also available commercially and can be tailored for different protocols.

**[00202]** Other examples of detectable molecules that can be utilized to detect an oligonucleotide include fluorophores. Several fluorophores can be used for labeling nucleotides including, for example, fluorescein, tetramethylrhodamine, Texas Red, and a number of others (*e.g.*, Haugland, *Handbook of Fluorescent Probes - 9th Ed.*, 2002, Molec. Probes, Inc., Eugene OR; Haugland, *The Handbook: A Guide to Fluorescent Probes and Labeling Technologies-10th Ed.*, 2005, Invitrogen, Carlsbad, CA).

**[00203]** As one example, oligonucleotides may be fluorescently labeled during chemical synthesis, since incorporation of amines or thiols during nucleotide synthesis permit addition of fluorophores. Fluorescently labeled nucleotides are commercially available. For example, uridine and deoxyuridine triphosphates are available that are conjugated to ten different fluorophores that cover the spectrum. Fluorescent dyes that can be bound directly to nucleotides can also be utilized as detectable molecules. For example, FAM, JOE, TAMRA, and ROX are amine reactive fluorescent dyes that have been attached to nucleotides and are used in automated DNA sequencing. These

fluorescently labeled nucleotides, for example, ROX-ddATP, ROX-ddCTP, ROX-ddGTP and ROX-ddUTP, are commercially available.

**[00204]** Non-radioactive and non-fluorescent detectable molecules are also available. As noted above, biotin can be attached directly to nucleotides and detected by specific and high affinity binding to avidin or streptavidin which has been chemically coupled to an enzyme catalyzing a colorimetric reaction (such as phosphatase, luciferase, or peroxidase). Digoxigenin labeled nucleotides can also similarly be used for non-isotopic detection of nucleic acids. Biotinylated and digoxigenin-labeled nucleotides are commercially available.

**[00205]** Very small particles, termed nanoparticles, also can be used to label oligonucleotide probes. These particles range from 1-1000 nm in size and include diverse chemical structures such as gold and silver particles and quantum dots. When irradiated with angled incident white light, silver or gold nanoparticles ranging from 40-120 nm will scatter monochromatic light with high intensity. The wavelength of the scattered light is dependent on the size of the particle. Four to five different particles in close proximity will each scatter monochromatic light, which when superimposed will give a specific, unique color. The particles are being manufactured by companies such as Genicon Sciences (Carlsbad, CA). Derivatized silver or gold particles can be attached to a broad array of molecules including, proteins, antibodies, small molecules, receptor ligands, and nucleic acids. For example, the surface of the particle can be chemically derivatized to allow attachment to a nucleotide.

**[00206]** Other types of nanoparticles that can be used for detection of a detectable molecule include quantum dots. Quantum dots are fluorescing crystals 1-5 nm in diameter that are excitable by light over a large range of wavelengths. Upon excitation by light having an appropriate wavelength, these crystals emit light, such as monochromatic light, with a wavelength dependent on their chemical composition and size. Quantum dots such as CdSe, ZnSe, InP, or InAs possess unique optical properties; these and similar quantum dots are available from a number of commercial sources (*e.g.*, NN-Labs, Fayetteville, AR; Ocean Nanotech, Fayetteville, AR; Nanoco Technologies, Manchester, UK; Sigma-Aldrich, St. Louis, MO).

**[00207]** Many dozens of classes of particles can be created according to the number of size classes of the quantum dot crystals. The size classes of the crystals are created either 1) by tight control of crystal formation parameters to create each desired size class of particle, or 2) by creation of batches of crystals under loosely controlled crystal formation parameters, followed by sorting according to desired size and/or emission wavelengths. Two examples of references in which quantum dots are embedded within

intrinsic silicon epitaxial layers of semiconductor light emitting/detecting devices are United States Patent Nos. 5,293,050 and 5,354,707 to Chapple Sokol, *et al.*

**[00208]** In certain embodiments, oligonucleotide primers or probes may be labeled with one or more light-emitting or otherwise detectable dyes. The light emitted by the dyes can be visible light or invisible light, such as ultraviolet or infrared light. In exemplary embodiments, the dye may be a fluorescence resonance energy transfer (FRET) dye; a xanthene dye, such as fluorescein and rhodamine; a dye that has an amino group in the alpha or beta position (such as a naphthylamine dye, 1-dimethylaminonaphthyl-5-sulfonate, 1-anilino-8-naphthalene sulfonate and 2-p-touidiny-6-naphthalene sulfonate); a dye that has 3-phenyl-7-isocyanatocoumarin; an acridine, such as 9-isothiocyanatoacridine and acridine orange; a pyrene, a benoxadiazole and a stilbene; a dye that has 3-( $\epsilon$ -carboxypentyl)-3'-ethyl-5,5'-dimethyloxacarboxyanine (CYA); 6-carboxy fluorescein (FAM); 5&6-carboxyrhodamine-110 (R110); 6-carboxyrhodamine-6G (R6G); N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA); 6-carboxy-X-rhodamine (ROX); 6-carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein (JOE); ALEXA FLUOR™; Cy2; Texas Red and Rhodamine Red; 6-carboxy-2',4,7,7'-tetrachlorofluorescein (TET); 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (HEX); 5-carboxy-2',4',5',7'-tetrachlorofluorescein (ZOE); NAN; NED; Cy3; Cy3.5; Cy5; Cy5.5; Cy7; and Cy7.5; IR800CW, ICG, Alexa Fluor 350; Alexa Fluor 488; Alexa Fluor 532; Alexa Fluor 546; Alexa Fluor 568; Alexa Fluor 594; Alexa Fluor 647; Alexa Fluor 680, or Alexa Fluor 750.

**[00209]** The AARS polynucleotides and oligonucleotides of the present invention can be used in any of the therapeutic, diagnostic, research, or drug discovery compositions and methods described herein.

#### V. ANTIBODIES

**[00210]** According to another aspect, the present invention further provides antibodies that exhibit binding specificity for an AARS polypeptide, or its native cellular binding partner (i.e. cellular receptor, lipid, carbohydrate, protein, or nucleic acid binding partner), or complex thereof, and methods of using same. The term antibody includes the various variations of the same, such as FABs, human antibodies, modified human antibodies, single chains, nonhuman antibodies, and other derivatives of the immunoglobulin fold that underlie immune system ligands for antigens, as described herein and known in the art. Antibodies can be used in any of the therapeutic, diagnostic, drug discovery, or protein expression/purification methods and compositions provided herein.

**[00211]** Certain antibodies of the present invention differ from certain previously made antibodies because they can distinguish between the AARS protein fragments of Table(s) 1-3, or Table(s) 4-6, or Table(s) 7-9 and their corresponding full-length AARS, typically by binding with greater affinity to the AARS protein fragments than to the corresponding full-length AARS. Generally, such antibodies may bind to unique sequences or structures generated or revealed by splice variations, proteolysis, or other cellular processing that generates an AARS protein fragment of the invention (*e.g.*, post translational processing, including but not limited to phosphorylation and other modifications that change protein structure). In some aspects the antibodies may bind to sequences around a unique splice junction (for example to one or more regions of at least 5 contiguous amino acids selected from the splice junction sequences listed in Tables 2B, 5B, or 8B, or alternatively to any amino acid sequence C-terminal of this splice site, for example as listed in Tables 2B, 5B, or 8B. For example, such antibodies may have binding specificity to one or more non-solvent exposed faces that are exposed in the AARS protein fragment but not in the full-length AARS, or sequences that are not found or are otherwise inaccessible in the full-length AARS. Antibodies may also bind to unique three-dimensional structures that result from differences in folding between the AARS protein fragment and the full-length AARS. Such differences in folding may be localized (*e.g.*, to a specific domain or region) or globalized. As one example, folding of AARS protein fragments may generate unique continuous or discontinuous epitopes that are not found in the corresponding or parent AARS. Examples also include antibodies that specifically bind to N- or C- termini generated by splice variations, proteolysis, or other cellular processing; such termini may be unique compared to the full-length AARS or may not be exposed for antibody binding in the full-length versions due to their termini being completely or partially buried in the overall structure of the larger AARS parent molecule.

**[00212]** In some embodiments, antibodies provided herein do not form aggregates, have a desired solubility, and/or have an immunogenicity profile that is suitable for use in humans, as described herein and known in the art. Also included are antibodies that are suitable for production work, such as to purify the AARS protein fragments described herein. Preferably, active antibodies can be concentrated to at least about 10mg/ml and optionally formulated for biotherapeutic uses.

**[00213]** In certain embodiments, antibodies are effective for modulating one or more of the non-canonical activities mediated by an AARS polypeptide of the invention. In certain embodiments, for example, the antibody is one that binds to an AARS polypeptide and/or its binding partner, inhibits their ability to interact with each other, and/or antagonizes the non-canonical activity of the AARS polypeptide. In certain

embodiments, for example, the antibody binds to the cellular binding partner of an AARS polypeptide, and mimics the AARS polypeptide activity, such as by increasing or agonizing the non-canonical activity mediated by the AARS polypeptide. Accordingly, antibodies may be used to diagnose, treat, or prevent diseases, disorders or other conditions that are mediated by an AARS polypeptide of the invention, such as by antagonizing or agonizing its activity partially or fully.

**[00214]** An antibody, or antigen-binding fragment thereof, is said to “specifically bind,” “immunologically bind,” and/or is “immunologically reactive” to a polypeptide of the invention if it reacts at a detectable level (within, for example, an ELISA assay) with the polypeptide, and does not react detectably in a statistically significant manner with unrelated polypeptides under similar conditions. In certain instances, a binding agent does not significantly interact with a full-length version of the AARS polypeptide.

**[00215]** Immunological binding, as used in this context, generally refers to the non-covalent interactions of the type which occur between an immunoglobulin molecule and an antigen for which the immunoglobulin is specific. The strength, or affinity of binding such as immunological binding interactions can be expressed in terms of the dissociation constant ( $K_d$ ) of the interaction, wherein a smaller  $K_d$  represents a greater affinity. Immunological binding properties of selected polypeptides can be quantified using methods well known in the art. *See, e.g., Davies et al. (1990) Annual Rev. Biochem. 59:439-473.* In certain illustrative embodiments, an antibody has an affinity for an AARS protein fragment of at least about 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, or 50 nM. In certain embodiments, the affinity of the antibody for an AARS protein fragment is stronger than its affinity for a corresponding full-length AARS polypeptide, typically by about 1.5x, 2x, 2.5x, 3x, 3.5x, 4x, 4.5x, 5x, 6x, 7x, 8x, 9x, 10x, 15x, 20x, 25x, 30x, 40x, 50x, 60x, 70x, 80x, 90x, 100x, 200x, 300x, 400x, 500x, 600x, 700x, 800x, 900x, 1000x or more (including all integers in between). In certain embodiments, an antibody has an affinity for a corresponding full-length AARS protein of at least about 0.05, 0.1, 0.25, 0.5, 0.75, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20  $\mu$ M. In certain embodiments, an antibody binds weakly or substantially undetectably to a full-length AARS protein.

**[00216]** An “antigen-binding site,” or “binding portion” of an antibody, refers to the part of the immunoglobulin molecule that participates in antigen binding. The antigen binding site is formed by amino acid residues of the N-terminal variable (“V”) regions of the heavy (“H”) and light (“L”) chains. Three highly divergent stretches within the V regions of the heavy and light chains are referred to as “hypervariable regions” which

are interposed between more conserved flanking stretches known as “framework regions,” or “FRs”. Thus the term “FR” refers to amino acid sequences which are naturally found between and adjacent to hypervariable regions in immunoglobulins. In an antibody molecule, the three hypervariable regions of a light chain and the three hypervariable regions of a heavy chain are disposed relative to each other in three dimensional space to form an antigen-binding surface. The antigen-binding surface is complementary to the three-dimensional surface of a bound antigen, and the three hypervariable regions of each of the heavy and light chains are referred to as “complementarity-determining regions,” or “CDRs.”

[00217] Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. See, e.g., Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. Monoclonal antibodies specific for a polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol.* 6:511-519, 1976, and improvements thereto. Also included are methods that utilize transgenic animals such as mice to express human antibodies. See, e.g., Neuberger *et al.*, *Nature Biotechnology* 14:826, 1996; Lonberg *et al.*, *Handbook of Experimental Pharmacology* 113:49-101, 1994; and Lonberg *et al.*, *Internal Review of Immunology* 13:65-93, 1995. Particular examples include the VELOCIMMUNE® platform by REGERNEREX® (see, e.g., U.S. Patent No. 6,596,541). Antibodies can also be generated or identified by the use of phage display or yeast display libraries (see, e.g., U.S. Patent No. 7,244,592; Chao *et al.*, *Nature Protocols*. 1:755-768, 2006). Non-limiting examples of available libraries include cloned or synthetic libraries, such as the Human Combinatorial Antibody Library (HuCAL), in which the structural diversity of the human antibody repertoire is represented by seven heavy chain and seven light chain variable region genes. The combination of these genes gives rise to 49 frameworks in the master library. By superimposing highly variable genetic cassettes (CDRs = complementarity determining regions) on these frameworks, the vast human antibody repertoire can be reproduced. Also included are human libraries designed with human-donor-sourced fragments encoding a light-chain variable region, a heavy-chain CDR-3, synthetic DNA encoding diversity in heavy-chain CDR-1, and synthetic DNA encoding diversity in heavy-chain CDR-2. Other libraries suitable for use will be apparent to persons skilled in the art. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

[00218] An “Fv” fragment can be produced by preferential proteolytic cleavage of an IgM, and on rare occasions IgG or IgA immunoglobulin molecule. Fv fragments are, however, more commonly derived using recombinant techniques known in the art. The

Fv fragment includes a non-covalent  $V_H::V_L$  heterodimer including an antigen-binding site which retains much of the antigen recognition and binding capabilities of the native antibody molecule. *See, e.g., Inbar et al. (1972) Proc. Nat. Acad. Sci. USA 69:2659-2662; Hochman et al. (1976) Biochem 15:2706-2710; and Ehrlich et al. (1980) Biochem 19:4091-4096.*

[00219] A single chain Fv ("sFv") polypeptide is a covalently linked  $V_H::V_L$  heterodimer which is expressed from a gene fusion including  $V_H$ - and  $V_L$ -encoding genes linked by a peptide-encoding linker. Huston *et al.* (1988) *PNAS USA*. 85(16):5879-5883. A number of methods have been described to discern chemical structures for converting the naturally aggregated--but chemically separated--light and heavy polypeptide chains from an antibody V region into an sFv molecule which will fold into a three dimensional structure substantially similar to the structure of an antigen-binding site. *See, e.g., U.S. Pat. Nos. 5,091,513 and 5,132,405, to Huston et al.; and U.S. Pat. No. 4,946,778, to Ladner et al.*

[00220] Each of the above-described molecules includes a heavy chain and a light chain CDR set, respectively interposed between a heavy chain and a light chain FR set which provide support to the CDRs and define the spatial relationship of the CDRs relative to each other. As used herein, the term "CDR set" refers to the three hypervariable regions of a heavy or light chain V region. Proceeding from the N-terminus of a heavy or light chain, these regions are denoted as "CDR1," "CDR2," and "CDR3" respectively. An antigen-binding site, therefore, includes six CDRs, comprising the CDR set from each of a heavy and a light chain V region. A polypeptide comprising a single CDR, (*e.g.*, a CDR1, CDR2 or CDR3) is referred to herein as a "molecular recognition unit." Crystallographic analysis of a number of antigen-antibody complexes has demonstrated that the amino acid residues of CDRs form extensive contact with bound antigen, wherein the most extensive antigen contact is with the heavy chain CDR3. Thus, the molecular recognition units are primarily responsible for the specificity of an antigen-binding site.

[00221] As used herein, the term "FR set" refers to the four flanking amino acid sequences which frame the CDRs of a CDR set of a heavy or light chain V region. Some FR residues may contact bound antigen; however, FRs are primarily responsible for folding the V region into the antigen-binding site, particularly the FR residues directly adjacent to the CDRs. Within FRs, certain amino residues and certain structural features are very highly conserved. In this regard, all V region sequences contain an internal disulfide loop of around 90 amino acid residues. When the V regions fold into a binding-site, the CDRs are displayed as projecting loop motifs which form an antigen-binding surface. It is generally recognized that there are conserved structural regions of



FRs which influence the folded shape of the CDR loops into certain “canonical” structures--regardless of the precise CDR amino acid sequence. Further, certain FR residues are known to participate in non-covalent interdomain contacts which stabilize the interaction of the antibody heavy and light chains.

[00222] Certain embodiments include single domain antibody (sdAbs or “nanobodies”), which refer to an antibody fragment consisting of a single monomeric variable antibody domain (*see, e.g.*, U.S. Patent Nos. 5,840,526; 5,874,541; 6,005,079, 6,765,087, 5,800,988; 5,874,541; and 6,015,695). Such sdABs typically have a molecular weight of about 12–15 kDa. In certain aspects, sdABs and other antibody molecules can be derived or isolated from the unique heavy-chain antibodies of immunized camels and llamas, often referred to as camelids. *See, e.g.*, Conrath *et al.*, *JBC*. 276:7346-7350, 2001.

[00223] A number of “humanized” antibody molecules comprising an antigen-binding site derived from a non-human immunoglobulin have been described, including chimeric antibodies having rodent V regions and their associated CDRs fused to human constant domains (Winter *et al.* (1991) *Nature* 349:293-299; Lobuglio *et al.* (1989) *Proc. Nat. Acad. Sci. USA* 86:4220-4224; Shaw *et al.* (1987) *J Immunol.* 138:4534-4538; and Brown *et al.* (1987) *Cancer Res.* 47:3577-3583), rodent CDRs grafted into a human supporting FR prior to fusion with an appropriate human antibody constant domain (Riechmann *et al.* (1988) *Nature* 332:323-327; Verhoeven *et al.* (1988) *Science* 239:1534-1536; and Jones *et al.* (1986) *Nature* 321:522-525), and rodent CDRs supported by recombinantly veneered rodent FRs (European Patent Publication No. 519,596, published Dec. 23, 1992). These “humanized” molecules are designed to minimize unwanted immunological response toward rodent antihuman antibody molecules which limits the duration and effectiveness of therapeutic applications of those moieties in human recipients. *See, e.g.*, U.S. Patent Nos. 5,530,101; 5,585,089; 5,693,762; 6,180,370; and 7,022,500.

[00224] The antibodies of the present invention can be used in any of the therapeutic, diagnostic, drug discovery, protein purification, and analytical methods and compositions described herein.

## VI. ANTIBODY ALTERNATIVES AND OTHER BINDING AGENTS

[00225] According to another aspect, the present invention further provides antibody alternatives or other binding agents, such as soluble receptors, adnectins, peptides, peptide mimetics, small molecules, aptamers, etc., that exhibit binding specificity for an AARS polypeptide or its cellular binding partner as disclosed herein, or to a portion, variant or derivative thereof, and compositions and methods of using same. Binding

agents can be used in any of the therapeutic, diagnostic, drug discovery, or protein expression/purification, and analytical methods and compositions described herein. Biologic-based binding agents such as adnectins, soluble receptors, avimers, and trinectins are particularly useful.

**[00226]** In certain embodiments, such binding agents are effective for modulating one or more of the non-canonical activities mediated by an AARS polypeptide of the invention. In some embodiments, for example, the binding agent is one that binds to an AARS polypeptide and/or its binding partner, inhibits their ability to interact with each other, and/or antagonizes the non-canonical activity of the AARS polypeptide. In certain embodiments, for example, the binding agent binds to the cellular binding partner of an AARS polypeptide, and mimics the AARS polypeptide activity, such as by increasing or agonizing the non-canonical activity mediated by the AARS polypeptide. Accordingly, such binding agents may be used to diagnose, treat, or prevent diseases, disorders or other conditions that are mediated by an AARS polypeptide of the invention, such as by antagonizing or agonizing its activity partially or fully.

**[00227]** A binding agent is said to “specifically bind” to an AARS polypeptide of the invention, or its cellular binding partner, if it reacts at a detectable level (within, for example, an ELISA assay) with the polypeptide or its cellular binding partner, and does not react detectably in a statistically significant manner with unrelated polypeptides under similar conditions. In certain instances, a binding agent does not significantly interact with a full-length version of the AARS polypeptide. In certain illustrative embodiments, a binding agent has an affinity for an AARS protein fragment or its cellular binding partner of at least about 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, or 50 nM. In certain embodiments, the affinity of the binding agent for an AARS protein fragment is stronger than its affinity for a corresponding full-length AARS polypeptide, typically by about 1.5x, 2x, 2.5x, 3x, 3.5x, 4x, 4.5x, 5x, 6x, 7x, 8x, 9x, 10x, 15x, 20x, 25x, 30x, 40x, 50x, 60x, 70x, 80x, 90x, 100x, 200x, 300x, 400x, 500x, 600x, 700x, 800x, 900x, 1000x or more (including all integers in between). In certain embodiments, a binding agent has an affinity for a corresponding full-length AARS protein of at least about 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20  $\mu$ M.

**[00228]** As noted above, “peptides” are included as binding agents. The term peptide typically refers to a polymer of amino acid residues and to variants and synthetic analogues of the same. In certain embodiments, the term “peptide” refers to relatively short polypeptides, including peptides that consist of about 2, 3, 4, 5, 6, 7, 8,

9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, or 50 amino acids, including all integers and ranges (*e.g.*, 5-10, 8-12, 10-15) in between, and interact with an AARS polypeptide, its cellular binding partner, or both. Peptides can be composed of naturally-occurring amino acids and/or non-naturally occurring amino acids, as described herein.

[00229] In addition to peptides consisting only of naturally-occurring amino acids, peptidomimetics or peptide analogs are also provided. Peptide analogs are commonly used in the pharmaceutical industry as non-peptide drugs with properties analogous to those of the template peptide. These types of non-peptide compound are termed “peptide mimetics” or “peptidomimetics” (Luthman, *et al.*, *A Textbook of Drug Design and Development*, 14:386-406, 2nd Ed., Harwood Academic Publishers (1996); Joachim Grante, *Angew. Chem. Int. Ed. Engl.*, 33:1699-1720 (1994); Fauchere, J., *Adv. Drug Res.*, 15:29 (1986); Veber and Freidinger TINS, p. 392 (1985); and Evans, *et al.*, *J. Med. Chem.* 30:229 (1987)). A peptidomimetic is a molecule that mimics the biological activity of a peptide but is no longer peptidic in chemical nature. Peptidomimetic compounds are known in the art and are described, for example, in U.S. Patent No. 6,245,886.

[00230] The present invention also includes peptoids. Peptoid derivatives of peptides represent another form of modified peptides that retain the important structural determinants for biological activity, yet eliminate the peptide bonds, thereby conferring resistance to proteolysis (Simon, *et al.*, *PNAS USA*. 89:9367-9371, 1992). Peptoids are oligomers of N-substituted glycines. A number of N-alkyl groups have been described, each corresponding to the side chain of a natural amino acid. The peptidomimetics of the present invention include compounds in which at least one amino acid, a few amino acids or all amino acid residues are replaced by the corresponding N-substituted glycines. Peptoid libraries are described, for example, in U.S. Patent No. 5,811,387.

[00231] A binding agent may also include one or more small molecules. A “small molecule” refers to an organic compound that is of synthetic or biological origin (biomolecule), but is typically not a polymer. Organic compounds refer to a large class of chemical compounds whose molecules contain carbon, typically excluding those that contain only carbonates, simple oxides of carbon, or cyanides. A “biomolecule” refers generally to an organic molecule that is produced by a living organism, including large polymeric molecules (biopolymers) such as peptides, polysaccharides, and nucleic acids as well, and small molecules such as primary secondary metabolites, lipids, phospholipids, glycolipids, sterols, glycerolipids, vitamins, and hormones. A “polymer” refers generally to a large molecule or macromolecule composed of repeating structural units, which are typically connected by covalent chemical bond.

[00232] In certain embodiments, a small molecule has a molecular weight of less than 1000-2000 Daltons, typically between about 300 and 700 Daltons, and including about 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000 or 2000 Daltons. Small molecule libraries are described elsewhere herein.

[00233] Aptamers are also included as binding agents (*see, e.g.,* Ellington *et al.*, *Nature*. 346, 818-22, 1990; and Tuerk *et al.*, *Science*. 249, 505-10, 1990). Examples of aptamers included nucleic acid aptamers (*e.g.,* DNA aptamers, RNA aptamers) and peptide aptamers. Nucleic acid aptamers refer generally to nucleic acid species that have been engineered through repeated rounds of *in vitro* selection or equivalent method, such as SELEX (systematic evolution of ligands by exponential enrichment), to bind to various molecular targets such as small molecules, proteins, nucleic acids, and even cells, tissues and organisms. *See, e.g.,* U.S. Patent Nos. 6,376,190; and 6,387,620. Hence, included are nucleic acid aptamers that bind to the AARS polypeptides described herein and/or their cellular binding partners.

[00234] Peptide aptamers typically include a variable peptide loop attached at both ends to a protein scaffold, a double structural constraint that typically increases the binding affinity of the peptide aptamer to levels comparable to that of an antibody's (*e.g.,* in the nanomolar range). In certain embodiments, the variable loop length may be composed of about 10-20 amino acids (including all integers in between), and the scaffold may include any protein that has good solubility and compacity properties. Certain exemplary embodiments may utilize the bacterial protein Thioredoxin-A as a scaffold protein, the variable loop being inserted within the reducing active site (-Cys-Gly-Pro-Cys- loop in the wild protein), with the two cysteines lateral chains being able to form a disulfide bridge. Methods for identifying peptide aptamers are described, for example, in U.S. Application No. 2003/0108532. Hence, included are peptide aptamers that bind to the AARS polypeptides described herein and/or their cellular binding partners. Peptide aptamer selection can be performed using different systems known in the art, including the yeast two-hybrid system.

[00235] Also included are ADNECTINS™, AVIMERS™, anaphones and anticalins that specifically bind to an AARS protein fragment of the invention. ADNECTINS™ refer to a class of targeted biologics derived from human fibronectin, an abundant extracellular protein that naturally binds to other proteins. *See, e.g.,* U.S. Application Nos. 2007/0082365; 2008/0139791; and 2008/0220049. ADNECTINS™ typically consists of a natural fibronectin backbone, as well as the multiple targeting domains of a specific portion of human fibronectin. The targeting domains can be engineered to

enable an ADNECTIN™ to specifically recognize a therapeutic target of interest, such as an AARS protein fragment of the invention.

[00236] AVIMERS™ refer to multimeric binding proteins or peptides engineered using *in vitro* exon shuffling and phage display. Multiple binding domains are linked, resulting in greater affinity and specificity compared to single epitope immunoglobulin domains. See, e.g., Silverman *et al.*, *Nature Biotechnology*. 23:1556-1561, 2005; U.S. Patent No. 7,166,697; and U.S. Application Nos. 2004/0175756, 2005/0048512, 2005/0053973, 2005/0089932 and 2005/0221384.

[00237] Also included are designed ankyrin repeat proteins (DARPin), which include a class of non-immunoglobulin proteins that can offer advantages over antibodies for target binding in drug discovery and drug development. Among other uses, DARPins are ideally suited for *in vivo* imaging or delivery of toxins or other therapeutic payloads because of their favorable molecular properties, including small size and high stability. The low-cost production in bacteria and the rapid generation of many target-specific DARPins make the DARPin approach useful for drug discovery. Additionally, DARPins can be easily generated in multispecific formats, offering the potential to target an effector DARPin to a specific organ or to target multiple receptors with one molecule composed of several DARPins. See, e.g., Stumpp *et al.*, *Curr Opin Drug Discov Devel.* 10:153-159, 2007; U.S. Application No. 2009/0082274; and PCT/EP2001/10454.

[00238] Certain embodiments include “monobodies,” which typically utilize the 10th fibronectin type III domain of human fibronectin (FNfn10) as a scaffold to display multiple surface loops for target binding. FNfn10 is a small (94 residues) protein with a  $\beta$ -sandwich structure similar to the immunoglobulin fold. It is highly stable without disulfide bonds or metal ions, and it can be expressed in the correctly folded form at a high level in bacteria. The FNfn10 scaffold is compatible with virtually any display technologies. See, e.g., Batori *et al.*, *Protein Eng.* 15:1015-20, 2002; and Wojcik *et al.*, *Nat Struct Mol Biol.*, 2010; and U.S. Patent No. 6,673,901.

[00239] Anticalins refer to a class of antibody mimetics, which are typically synthesized from human lipocalins, a family of binding proteins with a hypervariable loop region supported by a structurally rigid framework. See, e.g., U.S. Application No. 2006/0058510. Anticalins typically have a size of about 20 kDa. Anticalins can be characterized by a barrel structure formed by eight antiparallel  $\beta$ -strands (a stable  $\beta$ -barrel scaffold) that are pairwise connected by four peptide loops and an attached  $\alpha$ -helix. In certain aspects, conformational deviations to achieve specific binding are made in the hypervariable loop region(s). See, e.g., Skerra, *FEBS J.* 275:2677-83, 2008, herein incorporated by reference.

VII. BIOASSAYS AND ANALYTICAL ASSAYS FOR DRUG RELEASE ASSAYS AND PRODUCT SPECIFICATIONS, DIAGNOSTICS, AND REAGENTS

[00240] Also included are bioassays that relate to the AARS protein fragments and related agents as therapeutic and diagnostic reagents. Examples include bioassays and analytical assays that measure purity, biological activity, affinity, solubility, pH, endotoxin levels, among others, many of which are described herein. Also included are assays that establish dose response curves and/or provide one or more bases for comparison between different batches of agents. Batch comparisons can be based on any one or more of chemical characterization, biological characterization, and clinical characterization. For protein agents, also included are methods of evaluating the potency, stability, pharmacokinetics, and immunogenicity of a selected agent. Among other uses, these and other methods can be used for lot releasing testing of biologic or chemical agents, including the AARS protein fragments, antibodies, binding agents, polynucleotides such as antisense agents and vectors, and others described herein.

[00241] Certain embodiments include the use of bioaffinity assays. Such assays can be used to assess the binding affinity, for example, between an AARS protein fragment and a cellular binding partner, or between an AARS protein fragment and an antibody. Binding affinity can also be measured between an AARS protein fragment and an alternate binding agent such as a candidate or lead test compound (*e.g.*, small molecule modulator of an AARS), or between an AARS cellular binding partner and a candidate or lead test compound. Certain exemplary binding affinity assays may utilize ELISA assays, as described herein and known in the art. Certain assays utilize high-performance receptor binding chromatography (*see, e.g.*, Roswall *et al.*, *Biologicals*. 24:25-39, 1996). Other exemplary binding affinity assays may utilize surface plasmon resonance (SPR)-based technologies. Examples include BIACore technologies, certain of which integrate SPR technology with a microfluidics system to monitor molecular interactions in real time at concentrations ranging from pM to mM. Also included are KinExa™ assays, which provide accurate measurements of binding specificity, binding affinity, and binding kinetics/rate constants.

[00242] Certain embodiments relate to immunoassays for evaluating or optimizing the immunogenicity of protein agents. Examples include *ex vivo* human cellular assays and *in vitro* immuno-enzymatic assays to provide useful information on the immunogenic potential of a therapeutic protein. *Ex vivo* cell-response assays can be used, for example, to reproduce the cellular co-operation between antigen-presenting cells (APCs) and T-cells, and thereby measure T-cells activation after contact with a protein of interest. Certain *in vitro* enzymatic assays may utilize a collection of recombinant HLA-DR molecules that cover a significant portion of a relevant human

population, and may include automated immuno-enzymatic assays for testing the binding of peptides (stemming from the fragmentation of the therapeutic protein) with the HLA-DR molecules. Also included are methods of reducing the immunogenicity of a selected protein, such as by using these and related methods to identify and then remove or alter one or more T-cell epitopes from a protein agent.

[00243] Also included are biological release assays (*e.g.*, cell-based assays) for measuring parameters such as specific biological activities, including non-canonical biological activities, and cytotoxicity. Certain specific biological assays include, for example, cell-based assays that utilize a cellular binding partner (*e.g.*, cell-surface receptor) of a selected AARS protein fragment, which is functionally coupled to a readout, such as a fluorescent or luminescent indicator of a non-canonical biological activity, as described herein. For instance, specific embodiments include a cell that comprises a cell-surface receptor or an extracellular portion thereof that binds to an AARS protein fragment, wherein the cell comprises a detector or readout. Also included are *in vivo* biological assays to characterize the pharmacokinetics of an agent, such as an AARS polypeptide or antibody, typically utilizing engineered mice or other mammal (*see, e.g.*, Lee *et al.*, *The Journal of Pharmacology*. 281:1431-1439, 1997). Examples of cytotoxicity-based biological assays include release assays (*e.g.*, chromium or europium release assays to measure apoptosis; *see, e.g.*, von Zons *et al.*, *Clin Diagn Lab Immunol*.4:202–207, 1997), among others, which can assess the cytotoxicity of AARS protein fragments, whether for establishing dose response curves, batch testing, or other properties related to approval by various regulatory agencies, such as the Food and Drug Administration (FDA).

[00244] Such assays can be used, for example, to develop a dose response curve for a selected AARS protein fragment or other agent, and/or to compare the dose response curve of different batches of proteins or other agents. A dose-response curve is an X-Y graph that relates the magnitude of a stressor to the response of a receptor; the response may be a physiological or biochemical response, such as a non-canonical biological activity in a cell *in vitro* or in a cell or tissue *in vivo*, a therapeutically effective amount as measured *in vivo* (*e.g.*, as measured by  $EC_{50}$ ), or death, whether measured *in vitro* or *in vivo* (*e.g.*, cell death, organismal death). Death is usually indicated as an  $LD_{50}$ , a statistically-derived dose that is lethal to 50% of a modeled population, though it can be indicated by  $LC_{01}$  (lethal dose for 1% of the animal test population),  $LC_{100}$  (lethal dose for 100% of the animal test population), or  $LC_{LO}$  (lowest dose causing lethality). Almost any desired effect or endpoint can be characterized in this manner.

[00245] The measured dose of a response curve is typically plotted on the X axis and the response is plotted on the Y axis. More typically, the logarithm of the dose is

plotted on the X axis, most often generating a sigmoidal curve with the steepest portion in the middle. The No Observable Effect Level (NOEL) refers to the lowest experimental dose for which no measurable effect is observed, and the threshold dose refers to the first point along the graph that indicates a response above zero. As a general rule, stronger drugs generate steeper dose response curves. For many drugs, the desired effects are found at doses slightly greater than the threshold dose, often because lower doses are relatively ineffective and higher doses lead to undesired side effects. For *in vivo* generated dose response curves, a curve can be characterized by values such as  $\mu\text{g/kg}$ ,  $\text{mg/kg}$ , or  $\text{g/kg}$  of body-weight, if desired.

[00246] For batch comparisons, it can be useful to calculate the coefficient of variation (CV) between different dose response curves of different batches (*e.g.*, between different batches of AARS protein fragments, antibodies, or other agents), in part because the CV allows comparison between data sets with different units or different means. For instance, in certain exemplary embodiments, two or three or more different batches of AARS protein fragments or other agents have a CV between them of less than about 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% for a 4, 5, 6, 7, or 8 point dose curve. In certain embodiments, the dose response curve is measured in a cell-based assay, and its readout relates to an increase or a decrease in a selected non-canonical activity of the AARS protein fragment. In certain embodiments, the dose response curve is measured in a cell release assay or animal model (*e.g.*, mouse model), and its readout relates to cell death or animal death. Other variations will be apparent to persons skilled in the art.

#### VIII. EXPRESSION AND PURIFICATION SYSTEMS

[00247] Embodiments of the present invention include methods and related compositions for expressing and purifying the AARS protein fragments or other polypeptide-based agents of the invention. Such recombinant AARS polypeptides can be conveniently prepared using standard protocols as described for example in Sambrook, *et al.*, (1989, *supra*), in particular Sections 16 and 17; Ausubel *et al.*, (1994, *supra*), in particular Chapters 10 and 16; and Coligan *et al.*, Current Protocols in Protein Science (John Wiley & Sons, Inc. 1995-1997), in particular Chapters 1, 5 and 6. As one general example, AARS polypeptides may be prepared by a procedure including one or more of the steps of: (a) preparing a construct comprising a polynucleotide sequence that encodes a AARS polypeptide and that is operably linked to a regulatory element; (b) introducing the construct into a host cell; (c) culturing the host cell to express the AARS polypeptide; and (d) isolating the AARS polypeptide from the host cell.



[00248] AARS polynucleotides are described elsewhere herein. In order to express a desired polypeptide, a nucleotide sequence encoding the polypeptide, or a functional equivalent, may be inserted into appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding a polypeptide of interest and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described in Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (1989), and Ausubel *et al.*, Current Protocols in Molecular Biology (1989).

[00249] A variety of expression vector/host systems are known and may be utilized to contain and express polynucleotide sequences. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (*e.g.*, baculovirus); plant cell systems transformed with virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (*e.g.*, Ti or pBR322 plasmids); or animal cell systems, including mammalian cell and more specifically human cell systems.

[00250] The “control elements” or “regulatory sequences” present in an expression vector are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the PBLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSPORT1 plasmid (Gibco BRL, Gaithersburg, Md.) and the like may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are generally preferred. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding a polypeptide, vectors based on SV40 or EBV may be advantageously used with an appropriate selectable marker.

[00251] In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the expressed polypeptide. For example, when large quantities are needed, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited

to, the multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding the polypeptide of interest may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of  $\beta$ -galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke & Schuster, *J. Biol. Chem.* 264:5503-5509 (1989)); and the like. pGEX Vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

[00252] Certain embodiments may employ *E. coli*-based expression systems (*see, e.g.,* Structural Genomics Consortium *et al., Nature Methods*. 5:135-146, 2008). These and related embodiments may rely partially or totally on ligation-independent cloning (LIC) to produce a suitable expression vector. In specific embodiments, protein expression may be controlled by a T7 RNA polymerase (*e.g.,* pET vector series). These and related embodiments may utilize the expression host strain BL21(DE3), a  $\lambda$ DE3 lysogen of BL21 that supports T7-mediated expression and is deficient in lon and ompT proteases for improved target protein stability. Also included are expression host strains carrying plasmids encoding tRNAs rarely used in *E. coli*, such as Rosetta™ (DE3) and Rosetta 2 (DE3) strains. Cell lysis and sample handling may also be improved using reagents sold under the trademarks BENZONASE® nuclease and BUGBUSTER® Protein Extraction Reagent. For cell culture, auto-inducing media can improve the efficiency of many expression systems, including high-throughput expression systems. Media of this type (*e.g.,* Overnight Express™ Autoinduction System) gradually elicit protein expression through metabolic shift without the addition of artificial inducing agents such as IPTG. Particular embodiments employ hexahistidine tags (such as those sold under the trademark HIS•TAG® fusions), followed by immobilized metal affinity chromatography (IMAC) purification, or related techniques. In certain aspects, however, clinical grade proteins can be isolated from *E. coli* inclusion bodies, without or without the use of affinity tags (*see, e.g.,* Shimp *et al., Protein Expr Purif.* 50:58-67, 2006). As a further example, certain embodiments may employ a cold-shock induced *E. coli* high-yield production system, because over-expression of proteins in *Escherichia coli* at low temperature improves their solubility and stability (*see, e.g.,* Qing *et al., Nature Biotechnology*. 22:877-882, 2004).

[00253] Also included are high-density bacterial fermentation systems. For example, high cell density cultivation of *Ralstonia eutropha* allows protein production at cell densities of over 150 g/L, and the expression of recombinant proteins at titers exceeding 10 g/L.

[00254] In the yeast *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel *et al.* (supra) and Grant *et al.*, *Methods Enzymol.* 153:516-544 (1987). Also included are *Pichia pandoris* expression systems (see, e.g., Li *et al.*, *Nature Biotechnology.* 24, 210 – 215, 2006; and Hamilton *et al.*, *Science*, 301:1244, 2003). Certain embodiments include yeast systems that are engineered to selectively glycosylate proteins, including yeast that have humanized N-glycosylation pathways, among others (see, e.g., Hamilton *et al.*, *Science.* 313:1441-1443, 2006; Wildt *et al.*, *Nature Reviews Microbiol.* 3:119-28, 2005; and Gerngross *et al.*, *Nature-Biotechnology.* 22:1409 -1414, 2004; U.S. Patent Nos. 7,629,163; 7,326,681; and 7,029,872). Merely by way of example, recombinant yeast cultures can be grown in Fernbach Flasks or 15L, 50L, 100L, and 200L fermentors, among others.

[00255] In cases where plant expression vectors are used, the expression of sequences encoding polypeptides may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, *EMBO J.* 6:307-311 (1987)). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi *et al.*, *EMBO J.* 3:1671-1680 (1984); Broglie *et al.*, *Science* 224:838-843 (1984); and Winter *et al.*, *Results Probl. Cell Differ.* 17:85-105 (1991)). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, e.g., Hobbs in McGraw Hill, *Yearbook of Science and Technology*, pp. 191-196 (1992)).

[00256] An insect system may also be used to express a polypeptide of interest. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* cells. The sequences encoding the polypeptide may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the polypeptide-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which the polypeptide of interest may be expressed (Engelhard *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 91:3224-3227 (1994)). Also included

are baculovirus expression systems, including those that utilize SF9, SF21, and Tni cells (*see, e.g.,* Murphy and Piwnica-Worms, *Curr Protoc Protein Sci.* Chapter 5:Unit5.4, 2001). Insect systems can provide post-translation modifications that are similar to mammalian systems.

[00257] In mammalian host cells, a number of viral-based expression systems are generally available. For example, in cases where an adenovirus is used as an expression vector, sequences encoding a polypeptide of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the polypeptide in infected host cells (Logan & Shenk, *Proc. Natl. Acad. Sci. U.S.A.* 81:3655-3659 (1984)). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

[00258] Examples of useful mammalian host cell lines include monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells sub-cloned for growth in suspension culture, Graham *et al.*, *J. Gen Virol.* 36:59 (1977)); baby hamster kidney cells (BHK, ATCC CCL 10); mouse sertoli cells (TM4, Mather, *Biol. Reprod.* 23:243-251 (1980)); monkey kidney cells (CV1 ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor (MMT 060562, ATCC CCL51); TR1 cells (Mather *et al.*, *Annals N.Y. Acad. Sci.* 383:44-68 (1982)); MRC 5 cells; FS4 cells; and a human hepatoma line (Hep G2). Other useful mammalian host cell lines include Chinese hamster ovary (CHO) cells, including DHFR-CHO cells (Urlaub *et al.*, *PNAS USA* 77:4216 (1980)); and myeloma cell lines such as NSO and Sp2/0. For a review of certain mammalian host cell lines suitable for antibody production, *see, e.g.,* Yazaki and Wu, *Methods in Molecular Biology*, Vol. 248 (B. K.C Lo, ed., Humana Press, Totowa, N.J., 2003), pp. 255-268. Certain preferred mammalian cell expression systems include CHO and HEK293-cell based expression systems. Mammalian expression systems can utilize attached cell lines, for example, in T-flasks, roller bottles, or cell factories, or suspension cultures, for example, in 1L and 5L spinners, 5L, 14L, 40L, 100L and 200L stir tank bioreactors, or 20/50L and 100/200L WAVE bioreactors, among others known in the art.

[00259] Also included is cell-free expression of proteins. These and related embodiments typically utilize purified RNA polymerase, ribosomes, tRNA and

ribonucleotides; these reagents may be produced by extraction from cells or from a cell-based expression system.

[00260] Specific initiation signals may also be used to achieve more efficient translation of sequences encoding a polypeptide of interest. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf. *et al.*, *Results Probl. Cell Differ.* 20:125-162 (1994)).

[00261] In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, post-translational modifications such as acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a “prepro” form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as yeast, CHO, HeLa, MDCK, HEK293, and W138, in addition to bacterial cells, which have or even lack specific cellular machinery and characteristic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

[00262] For long-term, high-yield production of recombinant proteins, stable expression is generally preferred. For example, cell lines which stably express a polynucleotide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type. Transient production, such as by transient transfection or infection, can also

be employed. Exemplary mammalian expression systems that are suitable for transient production include HEK293 and CHO-based systems.

[00263] Any number of selection systems may be used to recover transformed or transduced cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler *et al.*, *Cell* 11:223-232 (1977)) and adenine phosphoribosyltransferase (Lowy *et al.*, *Cell* 22:817-823 (1990)) genes which can be employed in tk- or aprt- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 77:3567-70 (1980)); npt, which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin *et al.*, *J. Mol. Biol.* 150:1-14 (1981)); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman & Mulligan, *Proc. Natl. Acad. Sci. U.S.A.* 85:8047-51 (1988)). The use of visible markers has gained popularity with such markers as green fluorescent protein (GFP) and other fluorescent proteins (*e.g.*, RFP, YFP), anthocyanins,  $\beta$ -glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (*see, e.g.*, Rhodes *et al.*, *Methods Mol. Biol.* 55:121-131 (1995)).

[00264] Embodiments of the present invention also include high-throughput protein production systems, or micro-production systems. Certain aspects may utilize, for example, hexa-histidine fusion tags for protein expression and purification on metal chelate-modified slide surfaces or MagneHis Ni-Particles (*see, e.g.*, Kwon *et al.*, *BMC Biotechnol.* 9:72, 2009; and Lin *et al.*, *Methods Mol Biol.* 498:129-41, 2009)). Also included are high-throughput cell-free protein expression systems (*see, e.g.*, Sitaraman *et al.*, *Methods Mol Biol.* 498:229-44, 2009). These and related embodiments can be used, for example, to generate microarrays of AARS protein fragment(s), which can then be used for screening libraries to identify agents that interact with the AARS protein fragment(s).

[00265] A variety of protocols for detecting and measuring the expression of polynucleotide-encoded products, using binding agents or antibodies such as polyclonal or monoclonal antibodies specific for the product, are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), western immunoblots, radioimmunoassays (RIA), and fluorescence activated cell sorting (FACS). These and

other assays are described, among other places, in Hampton *et al.*, *Serological Methods, a Laboratory Manual* (1990) and Maddox *et al.*, *J. Exp. Med.* 158:1211-1216 (1983).

[00266] A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits. Suitable reporter molecules or labels, which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

[00267] Host cells transformed with a polynucleotide sequence of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. Certain specific embodiments utilize serum free cell expression systems. Examples include HEK293 cells and CHO cells that can grown on serum free medium (see, e.g., Rosser *et al.*, *Protein Expr. Purif.* 40:237-43, 2005; and U.S. Patent number 6,210,922).

[00268] The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides of the invention may be designed to contain signal sequences which direct secretion of the encoded polypeptide through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding a polypeptide of interest to nucleotide sequence encoding a polypeptide domain which will facilitate purification and / or detection of soluble proteins. Examples of such domains include cleavable and non-cleavable affinity purification and epitope tags such as avidin, FLAG tags, poly-histidine tags (e.g., 6xHis), cMyc tags, V5-tags, glutathione S-transferase (GST) tags, and others.

[00269] The protein produced by a recombinant cell can be purified and characterized according to a variety of techniques known in the art. Exemplary systems for performing protein purification and analyzing protein purity include fast protein liquid chromatography (FPLC) (e.g., AKTA and Bio-Rad FPLC systems), high-pressure liquid chromatography (HPLC) (e.g., Beckman and Waters HPLC). Exemplary chemistries for purification include ion exchange chromatography (e.g., Q,

S), size exclusion chromatography, salt gradients, affinity purification (*e.g.*, Ni, Co, FLAG, maltose, glutathione, protein A/G), gel filtration, reverse-phase, ceramic HYPERD® ion exchange chromatography, and hydrophobic interaction columns (HIC), among others known in the art. Also included are analytical methods such as SDS-PAGE (*e.g.*, coomassie, silver stain), immunoblot, Bradford, and ELISA, which may be utilized during any step of the production or purification process, typically to measure the purity of the protein composition.

[00270] Also included are methods of concentrating AARS protein fragments, and composition comprising concentrated soluble proteins. In different aspects such concentrated solutions of AARS polypeptides may comprise proteins at a concentration of about 5 mg/mL; or about 8 mg/mL; or about 10 mg/mL; about 15 mg/mL; or about 20 mg/mL.

[00271] In one aspect such compositions may be substantially monodisperse, meaning that the AARS polypeptide compositions exist primarily (*i.e.* at least about 90%, or greater) in one apparent molecular weight form when assessed for example, by size exclusion chromatography, dynamic light scattering, or analytical ultracentrifugation.

[00272] In another aspect, such compositions have a purity (on a protein basis) of at least about 90%, or in some aspects at least about 95% purity, or in some embodiments, at least 98% purity. Purity may be determined via any routine analytical method as known in the art.

[00273] In another aspect, such compositions have a high molecular weight aggregate content of less than about 10%, compared to the total amount of protein present, or in some embodiments such compositions have a high molecular weight aggregate content of less than about 5%, or in some aspects such compositions have a high molecular weight aggregate content of less than about 3%, or in some embodiments a high molecular weight aggregate content of less than about 1%. High molecular weight aggregate content may be determined via a variety of analytical techniques including for example, by size exclusion chromatography, dynamic light scattering, or analytical ultracentrifugation.

[00274] In certain embodiments, as noted herein, the AARS polypeptide compositions have an endotoxin content of less than about 10 EU / mg of AARS polypeptide, or less than about 5 EU / mg of AARS polypeptide, less than about 3 EU / mg of AARS polypeptide, or less than about 1 EU / mg of AARS polypeptide.

[00275] Examples of concentration approaches contemplated herein include lyophilization, which is typically employed when the solution contains few soluble components other than the protein of interest. Lyophilization is often performed after



HPLC run, and can remove most or all volatile components from the mixture. Also included are ultrafiltration techniques, which typically employ one or more selective permeable membranes to concentrate a protein solution. The membrane allows water and small molecules to pass through and retains the protein; the solution can be forced against the membrane by mechanical pump, gas pressure, or centrifugation, among other techniques.

[00276] In certain embodiments, the reagents, AARS protein fragments, or related agents (*e.g.*, antibodies) have a purity of at least about 90%, as measured according to routine techniques in the art. In certain embodiments, such as diagnostic compositions or certain therapeutic compositions, the AARS compositions of the present invention have a purity of at least about 95%. In specific embodiments, such as therapeutic or pharmaceutical compositions, the AARS compositions of the present invention have a purity of at least about 97% or 98% or 99%. In other embodiments, such as when being used as reference or research reagents, AARS protein fragments can be of lesser purity, and may have a purity of at least about 50%, 60%, 70%, or 80%. Purity can be measured overall or in relation to selected components, such as other proteins, *e.g.*, purity on a protein basis.

[00277] Purified AARS protein fragments can also be characterized according to their biological characteristics. Examples include binding affinity or binding kinetics to a selected ligand (*e.g.*, a cellular binding partner of the AARS protein fragment such as a cell-surface receptor or an extracellular domain thereof), and the presence or levels of one or more canonical or non-canonical biological activity, as described herein. Binding affinity and binding kinetics can be measured according to a variety of techniques known in the art, such as BIACORE® and related technologies that utilize surface plasmon resonance (SPR), an optical phenomenon that enables detection of unlabeled interactants in real time. SPR-based biosensors can be used in determination of active concentration, screening and characterization in terms of both affinity and kinetics. The presence or levels of one or more canonical or non-canonical biological activities can be measured according to cell-based assays, including those that utilize a cellular binding partner (*e.g.*, cell-surface receptor) of a selected AARS protein fragment, which is functionally coupled to a readout or indicator, such as a fluorescent or luminescent indicator of a non-canonical biological activity, as described herein.

[00278] In certain embodiments, as noted above, the AARS polypeptide compositions are about substantially endotoxin free, including, for example, about 95% endotoxin free, preferably about 99% endotoxin free, and more preferably about 99.99% endotoxin free. The presence of endotoxins can be detected according to routine techniques in the art, as described herein. In specific embodiments, the AARS

compositions are made from a eukaryotic cell such as a mammalian or human cell in substantially serum free media.

[00279] In certain embodiments, the AARS polypeptide compositions comprise less than about 10% wt/wt high molecular weight aggregates, or less than about 5% wt/wt high molecular weight aggregates, or less than about 2% wt/wt high molecular weight aggregates, or less than about or less than about 1% wt/wt high molecular weight aggregates.

[00280] Also included are protein-based analytical assays and methods, which can be used to assess, for example, protein purity, size, solubility, and degree of aggregation, among other characteristics. Protein purity can be assessed a number of ways. For instance, purity can be assessed based on primary structure, higher order structure, size, charge, hydrophobicity, and glycosylation. Examples of methods for assessing primary structure include N- and C-terminal sequencing and peptide-mapping (*see, e.g., Allen et al., Biologicals*. 24:255-275, 1996)). Examples of methods for assessing higher order structure include circular dichroism (*see, e.g., Kelly et al., Biochim Biophys Acta*. 1751:119-139, 2005), fluorescent spectroscopy (*see, e.g., Meagher et al., J. Biol. Chem.* 273:23283-89, 1998), FT-IR, amide hydrogen-deuterium exchange kinetics, differential scanning calorimetry, NMR spectroscopy, immunoreactivity with conformationally sensitive antibodies. Higher order structure can also be assessed as a function of a variety of parameters such as pH, temperature, or added salts. Examples of methods for assessing protein characteristics such as size include analytical ultracentrifugation and size exclusion HPLC (SEC-HPLC), and exemplary methods for measuring charge include ion-exchange chromatography and isoelectric focusing. Hydrophobicity can be assessed, for example, by reverse-phase HPLC and hydrophobic interaction chromatography HPLC. Glycosylation can affect pharmacokinetics (*e.g., clearance*), conformation or stability, receptor binding, and protein function, and can be assessed, for example, by mass spectrometry and nuclear magnetic resonance (NMR) spectroscopy.

[00281] As noted above, certain embodiments include the use of SEC-HPLC to assess protein characteristics such as purity, size (*e.g., size homogeneity*) or degree of aggregation, and/or to purify proteins, among other uses. SEC, also including gel-filtration chromatography (GFC) and gel-permeation chromatography (GPC), refers to a chromatographic method in which molecules in solution are separated in a porous material based on their size, or more specifically their hydrodynamic volume, diffusion coefficient, and/or surface properties. The process is generally used to separate biological molecules, and to determine molecular weights and molecular weight distributions of polymers. Typically, a biological or protein sample (such as a protein

extract produced according to the protein expression methods provided herein and known in the art) is loaded into a selected size-exclusion column with a defined stationary phase (the porous material), preferably a phase that does not interact with the proteins in the sample. In certain aspects, the stationary phase is composed of inert particles packed into a dense three-dimensional matrix within a glass or steel column. The mobile phase can be pure water, an aqueous buffer, an organic solvent, or a mixture thereof. The stationary-phase particles typically have small pores and/or channels which only allow molecules below a certain size to enter. Large particles are therefore excluded from these pores and channels, and their limited interaction with the stationary phase leads them to elute as a “totally-excluded” peak at the beginning of the experiment. Smaller molecules, which can fit into the pores, are removed from the flowing mobile phase, and the time they spend immobilized in the stationary-phase pores depends, in part, on how far into the pores they penetrate. Their removal from the mobile phase flow causes them to take longer to elute from the column and results in a separation between the particles based on differences in their size. A given size exclusion column has a range of molecular weights that can be separated. Overall, molecules larger than the upper limit will not be trapped by the stationary phase, molecules smaller than the lower limit will completely enter the solid phase and elute as a single band, and molecules within the range will elute at different rates, defined by their properties such as hydrodynamic volume. For examples of these methods in practice with pharmaceutical proteins, see Bruner *et al.*, *Journal of Pharmaceutical and Biomedical Analysis*. 15: 1929-1935, 1997.

[00282] Protein purity for clinical applications is also discussed, for example, by Anicetti *et al.* (*Trends in Biotechnology*. 7:342-349, 1989). More recent techniques for analyzing protein purity include, without limitation, the LabChip GXII, an automated platform for rapid analysis of proteins and nucleic acids, which provides high throughput analysis of titer, sizing, and purity analysis of proteins. In certain non-limiting embodiments, clinical grade proteins such as protein fragments and antibodies can be obtained by utilizing a combination of chromatographic materials in at least two orthogonal steps, among other methods (*see, e.g.*, *Therapeutic Proteins: Methods and Protocols*. Vol. 308, Eds., Smales and James, Humana Press Inc., 2005). Typically, protein agents (*e.g.*, AARS protein fragments, antibodies, binding agents) and other agents (*e.g.*, antisense, RNAi, small molecules) are substantially endotoxin-free, as measured according to techniques known in the art and described herein.

[00283] Protein solubility assays are also included. Such assays can be utilized, for example, to determine optimal growth and purification conditions for recombinant production, to optimize the choice of buffer(s), and to optimize the choice of AARS

protein fragments or variants thereof. Solubility or aggregation can be evaluated according to a variety of parameters, including temperature, pH, salts, and the presence or absence of other additives. Examples of solubility screening assays include, without limitation, microplate-based methods of measuring protein solubility using turbidity or other measure as an end point, high-throughput assays for analysis of the solubility of purified recombinant proteins (*see, e.g., Stenvall et al., Biochim Biophys Acta.* 1752:6-10, 2005), assays that use structural complementation of a genetic marker protein to monitor and measure protein folding and solubility *in vivo* (*see, e.g., Wigley et al., Nature Biotechnology.* 19:131-136, 2001), and electrochemical screening of recombinant protein solubility in *Escherichia coli* using scanning electrochemical microscopy (SECM) (*see, e.g., Nagamine et al., Biotechnology and Bioengineering.* 96:1008-1013, 2006), among others. AARS protein fragments with increased solubility (or reduced aggregation) can be identified or selected for according to routine techniques in the art, including simple *in vivo* assays for protein solubility (*see, e.g., Maxwell et al., Protein Sci.* 8:1908-11, 1999).

[00284] Protein solubility and aggregation can also be measured by dynamic light scattering techniques. Aggregation is a general term that encompasses several types of interactions or characteristics, including soluble/insoluble, covalent/noncovalent, reversible/irreversible, and native/denatured interactions and characteristics. For protein therapeutics, the presence of aggregates is typically considered undesirable because of the concern that aggregates may cause an immunogenic reaction (*e.g.,* small aggregates), or may cause adverse events on administration (*e.g.,* particulates). Dynamic light scattering refers to a technique that can be used to determine the size distribution profile of small particles in suspension or polymers such as proteins in solution. This technique, also referred to as photon correlation spectroscopy (PCS) or quasi-elastic light scattering (QELS), uses scattered light to measure the rate of diffusion of the protein particles. Fluctuations of the scattering intensity can be observed due to the Brownian motion of the molecules and particles in solution. This motion data can be conventionally processed to derive a size distribution for the sample, wherein the size is given by the Stokes radius or hydrodynamic radius of the protein particle. The hydrodynamic size depends on both mass and shape (conformation). Dynamic scattering can detect the presence of very small amounts of aggregated protein (<0.01% by weight), even in samples that contain a large range of masses. It can also be used to compare the stability of different formulations, including, for example, applications that rely on real-time monitoring of changes at elevated temperatures. Accordingly, certain embodiments include the use of dynamic light scattering to

analyze the solubility and/or presence of aggregates in a sample that contains an AARS protein fragment, antibody, or other agent of the invention.

*IX. DIAGNOSTIC METHODS AND COMPOSITIONS*

[00285] AARS agents such as AARS protein fragments, AARS polynucleotides, and antibodies and other binding agents described herein can be used in diagnostic assays and diagnostic compositions. Included are biochemical, histological, and cell-based methods and compositions, among others.

[00286] These and related embodiments include the detection of the AARS polynucleotide sequence(s) or corresponding AARS polypeptide sequence(s) or portions thereof of one or more newly identified AARS protein fragments, also referred to as AARS polypeptides. For instance, certain aspects include detection of the AARS polynucleotide sequence(s) or corresponding polypeptide sequence(s) or portions thereof of one or more newly identified AARS splice variants, and/or one or more splice junctions of those splice variants. In certain embodiments, the polynucleotide or corresponding polypeptide sequence(s) of at least one of the splice junctions is unique to that particular AARS splice variant.

[00287] Also included is the direct detection of AARS protein fragments, including splice variants, proteolytic fragments, and others. In certain embodiments, the presence or levels of one or more newly identified AARS protein fragments associate or correlate with one or more cellular types or cellular states. Hence, the presence or levels of an AARS polypeptide or polynucleotide can be used to distinguish between different cellular types or different cellular states. The presence or levels of AARS protein fragments or their related polynucleotides can be detected according to polynucleotide and/or polypeptide-based diagnostic techniques, as described herein and known in the art.

[00288] Certain aspects can employ the AARS protein fragments, antibody, or AARS polynucleotides as part of a companion diagnostic method, typically to assess whether a subject or population subjects will respond favorably to a specific medical treatment. For instance, a given AARS therapeutic agent (*e.g.*, protein fragment, antisense, RNAi, antibody, binding agent) could be identified as suitable for a subject or certain populations of subjects based on whether the subject(s) have one or more selected biomarkers for a given disease or condition. Examples of biomarkers include serum/tissue markers as well as markers that can be identified by medical imaging techniques. In certain embodiments, a naturally-occurring AARS protein fragment (or its corresponding polynucleotide) may itself provide a serum and/or tissue biomarker that can be utilized to measure drug outcome or assess the desirability of drug use in a

specific subject or a specific population of subjects. In certain aspects, the identification of an AARS polypeptide or polynucleotide reference sequence may include characterizing the differential expression of that sequence, whether in a selected subject, selected tissue, or otherwise, as described herein and known in the art.

**[00289]** Certain of the methods provided herein rely on the differential expression of an AARS polypeptide or polynucleotide to characterize the condition or state of a cell, tissue, or subject, and to distinguish it from another cell, tissue, or subject. Non-limiting examples include methods of detecting the presence or levels of an AARS polypeptide or polynucleotide in a biological sample to distinguish between cells or tissues of different species, cells of different tissues or organs, cellular developmental states such as neonatal and adult, cellular differentiation states, conditions such as healthy, diseased and treated, intracellular and extracellular fractions, in addition to primary cell cultures and other cell cultures, such as immortalized cell cultures.

**[00290]** Differential expression includes a statistically significant difference in one or more gene expression levels of an AARS polynucleotide or polypeptide reference sequence compared to the expression levels of the same sequence in an appropriate control. The statistically significant difference may relate to either an increase or a decrease in expression levels, as measured by RNA levels, protein levels, protein function, or any other relevant measure of gene expression such as those described herein. Also included is a comparison between an AARS polynucleotide or polypeptide of the invention and a full-length or wild-type cytosolic or mitochondrial AARS sequence, typically of the same or corresponding type. Differential expression can be detected by a variety of techniques in the art and described herein, including polynucleotide and polypeptide based techniques, such as real-time PCR, subtractive hybridization, polynucleotide and polypeptide arrays, and others.

**[00291]** A result is typically referred to as statistically significant if it is unlikely to have occurred by chance. The significance level of a test or result relates traditionally to a frequentist statistical hypothesis testing concept. In simple cases, statistical significance may be defined as the probability of making a decision to reject the null hypothesis when the null hypothesis is actually true (a decision known as a Type I error, or “false positive determination”). This decision is often made using the p-value: if the p-value is less than the significance level, then the null hypothesis is rejected. The smaller the p-value, the more significant the result. Bayes factors may also be utilized to determine statistical significance (*see, e.g., Goodman S., Ann Intern Med 130:1005-13, 1999*).

**[00292]** In more complicated, but practically important cases, the significance level of a test or result may reflect an analysis in which the probability of making a decision

to reject the null hypothesis when the null hypothesis is actually true is no more than the stated probability. This type of analysis allows for those applications in which the probability of deciding to reject may be much smaller than the significance level for some sets of assumptions encompassed within the null hypothesis.

**[00293]** In certain exemplary embodiments, statistically significant differential expression may include situations wherein the expression level of a given AARS sequence provides at least about a 1.2X, 1.3X, 1.4X, 1.5X, 1.6X, 1.7X, 1.8X, 1.9X, 2.0X, 2.2X, 2.4X, 2.6X, 2.8X, 3.0X, 4.0X, 5.0X, 6.0X, 7.0X, 8.0X, 9.0X, 10.0X, 15.0X, 20.0X, 50.0X, 100.0X, or greater difference in expression (*i.e.*, differential expression that may be higher or lower expression) in a suspected biological sample as compared to an appropriate control, including all integers and decimal points in between (*e.g.*, 1.24X, 1.25X, 2.1X, 2.5X, 60.0X, 75.0X, etc.). In certain embodiments, statistically significant differential expression may include situations wherein the expression level of a given AARS sequence provides at least about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 percent (%) or greater difference in expression (*i.e.*, differential expression that may be higher or lower) in a suspected biological sample as compared to an appropriate control, including all integers and decimal points in between.

**[00294]** As an additional example, differential expression may also be determined by performing Z-testing, *i.e.*, calculating an absolute Z score, as described herein and known in the art (*see* Example 1). Z-testing is typically utilized to identify significant differences between a sample mean and a population mean. For example, as compared to a standard normal table (*e.g.*, a control tissue), at a 95% confidence interval (*i.e.*, at the 5% significance level), a Z-score with an absolute value greater than 1.96 indicates non-randomness. For a 99% confidence interval, if the absolute Z is greater than 2.58, it means that  $p < .01$ , and the difference is even more significant—the null hypothesis can be rejected with greater confidence. In these and related embodiments, an absolute Z-score of 1.96, 2, 2.58, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more, including all decimal points in between (*e.g.*, 10.1, 10.6, 11.2, etc.), may provide a strong measure of statistical significance. In certain embodiments, an absolute Z-score of greater than 6 may provide exceptionally high statistical significance.

**[00295]** Substantial similarity relates generally to the lack of a statistically significant difference in the expression levels between the biological sample and the reference control. Examples of substantially similar expression levels may include situations wherein the expression level of a given SSCIGS provides less than about a .05X, 0.1X, 0.2X, 0.3X, 0.4X, 0.5X, 0.6X, 0.7X, 0.8X, 0.9X, 1.0X, 1.1X, 1.2X, 1.3X, or 1.4X

difference in expression (*i.e.*, differential expression that may be higher or lower expression) in a suspected biological sample as compared to a reference sample, including all decimal points in between (*e.g.*, .15X, 0.25X, 0.35X, etc.). In certain embodiments, differential expression may include situations wherein the expression level of a given AARS sequence provides less than about 0.25. 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50 percent (%) difference in expression (*i.e.*, differential expression that may be higher or lower) in a suspected biological sample as compared to a reference sample, including all decimal points in between.

[00296] In certain embodiments, such as when using an Affymetrix Microarray to measure the expression levels of an AARS polynucleotide or polypeptide reference sequence, differential expression may also be determined by the mean expression value summarized by Affymetrix Microarray Suite 5 software (Affymetrix, Santa Clara, CA), or other similar software, typically with a scaled mean expression value of 1000.

[00297] Embodiments of the present invention include methods of detecting the presence or levels of an AARS polynucleotide or polypeptide reference sequence or a portion thereof to distinguish between cells or tissues or other biological sample of a different organism or species, wherein the presence or levels of that sequence associates with a selected organism or species. General examples include methods of distinguishing between humans and any combination of bacteria, fungi, plants, and other non-human animals. Included within animals are methods of distinguishing between humans and any combination of vertebrates and invertebrates, including vertebrates such as fish, amphibians, reptiles, birds, and non-human mammals, and invertebrates such as insects, mollusks, crustaceans, and corals. Included within non-human mammals are methods of distinguishing between humans and any combination of non-human mammals from the Order Afrosoricida, Macroscelidea, Tubulidentata, Hyracoidea, Proboscidea, Sirenia, Cingulata, Pilosa, Scandentia, Dermoptera, Primates, Rodentia, Lagomorpha, Erinaceomorpha, Soricomorpha, Chiroptera, Pholidota, Cetacea, Carnivora, Perissodactyla, or Artiodactyla. Included within the Primate Order are monkeys, apes, gorillas, and chimpanzees, among others known in the art. Accordingly, the presence or levels of an AARS polynucleotide or polypeptide reference sequence or variant, as described herein, may be used to identify the source of a given biological sample, such as a cell, tissue, or organ, by distinguishing between any combination of these organisms, or by distinguishing between humans and any one or more of these organisms, such as a panel of organisms. In certain embodiments, the source of a given biological sample may also be determined by comparing the presence or levels of an AARS sequence or a portion thereof to a pre-determined value.



[00298] Embodiments of the present invention include methods of detecting the presence or levels of an AARS polynucleotide or polypeptide reference sequence or a portion thereof to distinguish between cells or other biological samples that originate from different tissues or organs. Non-limiting examples include methods of distinguishing between a cell or other biological sample that originates from any combination of skin (*e.g.*, dermis, epidermis, subcutaneous layer), hair follicles, nervous system (*e.g.*, brain, spinal cord, peripheral nerves), auditory system or balance organs (*e.g.*, inner ear, middle ear, outer ear), respiratory system (*e.g.*, nose, trachea, lungs), gastroscophogcal tissues, the gastrointestinal system (*e.g.*, mouth, esophagus, stomach, small intestines, large intestines, rectum), vascular system (*e.g.*, heart, blood vessels and arteries), liver, gallbladder, lymphatic/immune system (*e.g.*, lymph nodes, lymphoid follicles, spleen, thymus, bone marrow), uro-genital system (*e.g.*, kidneys, ureter, bladder, urethra, cervix, Fallopian tubes, ovaries, uterus, vulva, prostate, bulbourethral glands, epididymis, prostate, seminal vesicles, testicles), musculoskeletal system (*e.g.*, skeletal muscles, smooth muscles, bone, cartilage, tendons, ligaments), adipose tissue, mammaries, and the endocrine system (*e.g.*, hypothalamus, pituitary, thyroid, pancreas, adrenal glands). Hence, based on the association of an AARS polynucleotide or polypeptide sequence as described herein, these methods may be used to identify or characterize the tissue or organ from which a cell or other biological sample is derived.

[00299] Embodiments of the present invention include methods of detecting the presence or levels of an AARS polynucleotide or polypeptide reference sequence or a portion thereof to distinguish between or characterize the developmental or differentiation state of the cell. Also included are methods of differentiating between germ cells, stem cells, and somatic cells. Examples of developmental states include neonatal and adult. Examples of cellular differentiation states include all of the discreet and identifiable stages between a totipotent cell, a pluripotent cell, a multipotent progenitor stem cell and a mature, fully differentiated cell.

[00300] A totipotent cell has total potential, typically arises during sexual and asexual reproduction, and includes and spores and zygotes, though in certain instances cells can dedifferentiate and regain totipotency. A pluripotent cell includes a stem cell that has the potential to differentiate into any of the three germ layers, including the endoderm (interior stomach lining, gastrointestinal tract, the lungs), the mesoderm (muscle, bone, blood, urogenital), and the ectoderm (epidermal tissues and nervous system). Multipotent progenitor cells are typically capable of differentiating into a limited number of tissue types. Examples of multipotent cells include, without limitation, hematopoietic stem cells (adult stem cells) from the bone marrow that give

rise to immune cells such as red blood cells, white blood cells, and platelets, mesenchymal stem cells (adult stem cells) from the bone marrow that give rise to stromal cells, fat cells, and various types of bone cells, epithelial stem cells (progenitor cells) that give rise to the various types of skin cells, and muscle satellite cells (progenitor cells) that contribute to differentiated muscle tissue. Accordingly, the presence or levels of particular AARS polynucleotide or polypeptide sequence (*e.g.*, splice junction of an AARS splice variant, AARS proteolytic fragment), can be used to distinguish between or characterize the above-noted cellular differentiation states, as compared to a control or a predetermined level.

[00301] Embodiments of the present invention include methods of detecting the presence or levels of an AARS polynucleotide or polypeptide reference sequence to characterize or diagnose the condition of a cell, tissue, organ, or subject, in which that condition may be characterized as healthy, diseased, at risk for being diseased, or treated. For such diagnostic purposes, the term “diagnostic” or “diagnosed” includes identifying the presence or nature of a pathologic condition, characterizing the risk of developing such a condition, and/or measuring the change (or no change) of a pathologic condition in response to therapy. Diagnostic methods may differ in their sensitivity and specificity. In certain embodiments, the “sensitivity” of a diagnostic assay refers to the percentage of diseased cells, tissues or subjects which test positive (percent of “true positives”). Diseased cells, tissues or subjects not detected by the assay are typically referred to as “false negatives.” Cells, tissues or subjects that are not diseased and which test negative in the assay may be termed “true negatives.” In certain embodiments, the “specificity” of a diagnostic assay may be defined as one (1) minus the false positive rate, where the “false positive” rate is defined as the proportion of those samples or subjects without the disease and which test positive. While a particular diagnostic method may not provide a definitive diagnosis of a condition, it suffices if the method provides a positive indication that aids in diagnosis.

[00302] In certain instances, the presence or risk of developing a pathologic condition can be diagnosed by comparing the presence or levels of one or more selected AARS polynucleotide or polypeptide reference sequences or portions thereof that correlate with the condition, whether by increased or decreased levels, as compared to a suitable control. A “suitable control” or “appropriate control” includes a value, level, feature, characteristic, or property determined in a cell or other biological sample of a tissue or organism, *e.g.*, a control or normal cell, tissue or organism, exhibiting, for example, normal traits, such as the absence of the condition. In certain embodiments, a “suitable control” or “appropriate control” is a predefined value, level, feature,

characteristic, or property. Other suitable controls will be apparent to persons skilled in the art. Examples of diseases and conditions are described elsewhere herein.

**[00303]** Embodiments of the present invention include AARS polynucleotide or nucleic acid-based detection techniques, which offer certain advantages due to sensitivity of detection. Hence, certain embodiments relate to the use or detection of AARS polynucleotides as part of a diagnostic method or assay. The presence and/or levels of AARS polynucleotides may be measured by any method known in the art, including hybridization assays such as Northern blot, quantitative or qualitative polymerase chain reaction (PCR), quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots, or *in situ* hybridization such as fluorescent *in situ* hybridization (FISH), among others. Certain of these methods are described in greater detail below.

**[00304]** AARS polynucleotides such as DNA and RNA can be collected and/or generated from blood, biological fluids, tissues, organs, cell lines, or other relevant sample using techniques known in the art, such as those *described* in Kingston. (2002 *Current Protocols in Molecular Biology*, Greene Publ. Assoc. Inc. & John Wiley & Sons, Inc., NY, NY (*see, e.g.*, as described by Nelson *et al. Proc Natl Acad Sci U S A*, 99: 11890-11895, 2002) and elsewhere. Further, a variety of commercially available kits for constructing RNA are useful for making the RNA to be used in the present invention. RNA may be constructed from organs/tissues/cells procured from normal healthy subjects; however, this invention also contemplates construction of RNA from diseased subjects. Certain embodiments contemplate using any type of organ from any type of subject or animal. For test samples RNA may be procured from an individual (*e.g.*, any animal, including mammals) with or without visible disease and from tissue samples, biological fluids (*e.g.*, whole blood) or the like.

**[00305]** In certain embodiments, amplification or construction of cDNA sequences may be helpful to increase detection capabilities. The instant disclosure, as well as the art, provides the requisite level of detail to perform such tasks. In one exemplary embodiment, whole blood is used as the source of RNA and accordingly, RNA stabilizing reagents are optionally used, such as PAX tubes, as described, for example, in Thach *et al.*, *J. Immunol. Methods*. Dec 283(1-2):269-279, 2003 and Chai *et al.*, *J. Clin. Lab Anal.* 19(5):182-188, 2005 (both of which are incorporated by reference). Complementary DNA (cDNA) libraries can be generated using techniques known in the art, such as those described in Ausubel *et al.* (2001 *Current Protocols in Molecular Biology*, Greene Publ. Assoc. Inc. & John Wiley & Sons, Inc., NY, NY); Sambrook *et al.* (1989 *Molecular Cloning*, Second Ed., Cold Spring Harbor Laboratory, Plainview, NY); Maniatis *et al.* (1982 *Molecular Cloning*, Cold Spring Harbor Laboratory,

Plainview, NY) and elsewhere. Further, a variety of commercially available kits for constructing cDNA libraries are useful for making the cDNA libraries of the present invention. Libraries can be constructed from organs/tissues/cells procured from normal, healthy subjects.

**[00306]** Certain embodiments may employ hybridization methods for detecting AARS polynucleotide sequences. Methods for conducting polynucleotide hybridization assays have been well developed in the art. Hybridization assay procedures and conditions will vary depending on the application and are selected in accordance with the general binding methods known including those referred to in: Maniatis *et al.* Molecular Cloning: A Laboratory Manual (2nd Ed. Cold Spring Harbor, N.Y., 1989); Berger and Kimmel Methods in Enzymology, Vol. 152, Guide to Molecular Cloning Techniques (Academic Press, Inc., San Diego, Calif., 1987); Young and Davis, *PNAS*. 80: 1194 (1983). Methods and apparatus for carrying out repeated and controlled hybridization reactions have been described in U.S. Patent Nos. 5,871,928, 5,874,219, 6,045,996 and 6,386,749, 6,391,623 each of which are incorporated herein by reference

**[00307]** Certain embodiments may employ nucleic acid amplification methods for detecting AARS polynucleotide sequences. The term “amplification” or “nucleic acid amplification” refers to the production of multiple copies of a target nucleic acid that contains at least a portion of the intended specific target nucleic acid sequence. The multiple copies may be referred to as amplicons or amplification products. In certain embodiments, the amplified target contains less than the complete target gene sequence (introns and exons) or an expressed target gene sequence (spliced transcript of exons and flanking untranslated sequences). For example, specific amplicons may be produced by amplifying a portion of the target polynucleotide by using amplification primers that hybridize to, and initiate polymerization from, internal positions of the target polynucleotide. Preferably, the amplified portion contains a detectable target sequence that may be detected using any of a variety of well-known methods.

**[00308]** “Selective amplification” or “specific amplification,” as used herein, refers to the amplification of a target nucleic acid sequence according to the present invention wherein detectable amplification of the target sequence is substantially limited to amplification of target sequence contributed by a nucleic acid sample of interest that is being tested and is not contributed by target nucleic acid sequence contributed by some other sample source, *e.g.*, contamination present in reagents used during amplification reactions or in the environment in which amplification reactions are performed.

**[00309]** The term “amplification conditions” refers to conditions permitting nucleic acid amplification according to the present invention. Amplification conditions may, in some embodiments, be less stringent than “stringent hybridization conditions” as

described herein. Oligonucleotides used in the amplification reactions of the present invention hybridize to their intended targets under amplification conditions, but may or may not hybridize under stringent hybridization conditions. On the other hand, detection probes of the present invention typically hybridize under stringent hybridization conditions. Acceptable conditions to carry out nucleic acid amplifications according to the present invention can be easily ascertained by someone having ordinary skill in the art depending on the particular method of amplification employed.

**[00310]** Many well-known methods of nucleic acid amplification require thermocycling to alternately denature double-stranded nucleic acids and hybridize primers; however, other well-known methods of nucleic acid amplification are isothermal. The polymerase chain reaction (U.S. Pat. Nos. 4,683,195; 4,683,202; 4,800,159; 4,965,188), commonly referred to as PCR, uses multiple cycles of denaturation, annealing of primer pairs to opposite strands, and primer extension to exponentially increase copy numbers of the target sequence. In a variation called RT-PCR, reverse transcriptase (RT) is used to make a complementary DNA (cDNA) from mRNA, and the cDNA is then amplified by PCR to produce multiple copies of DNA.

**[00311]** As noted above, the term “PCR” refers to multiple amplification cycles that selectively amplify a target nucleic acid species. Included are quantitative PCR (qPCR), real-time PCR), reverse transcription PCR (RT-PCR) and quantitative reverse transcription PCR (qRT-PCR) is well described in the art. The term “pPCR” refers to quantitative polymerase chain reaction, and the term “qRT-PCR” refers to quantitative reverse transcription polymerase chain reaction. qPCR and qRT-PCR may be used to amplify and simultaneously quantify a targeted cDNA molecule. It enables both detection and quantification of a specific sequence in a cDNA pool, such as a selected AARS gene or transcript.

**[00312]** The term “real-time PCR” may use DNA-binding dye to bind to all double-stranded (ds) DNA in PCR, causing fluorescence of the dye. An increase in DNA product during PCR therefore leads to an increase in fluorescence intensity and is measured at each cycle, thus allowing DNA concentrations to be quantified. However, dsDNA dyes such as SYBR Green will bind to all dsDNA PCR products. Fluorescence is detected and measured in the real-time PCR thermocycler, and its geometric increase corresponding to exponential increase of the product is used to determine the threshold cycle (“Ct”) in each reaction.

**[00313]** The term “Ct Score” refers to the threshold cycle number, which is the cycle at which PCR amplification has surpassed a threshold level. If there is a higher quantity of mRNA for a particular gene in a sample, it will cross the threshold earlier than a lowly expressed gene since there is more starting RNA to amplify. Therefore, a low Ct

score indicates high gene expression in a sample and a high Ct score is indicative of low gene expression.

**[00314]** Certain embodiments may employ the ligase chain reaction (Weiss, *Science*, 254: 1292, 1991), commonly referred to as LCR, which uses two sets of complementary DNA oligonucleotides that hybridize to adjacent regions of the target nucleic acid. The DNA oligonucleotides are covalently linked by a DNA ligase in repeated cycles of thermal denaturation, hybridization and ligation to produce a detectable double-stranded ligated oligonucleotide product.

**[00315]** Another method is strand displacement amplification (Walker, G. *et al.*, 1992, *Proc. Natl. Acad. Sci. USA* 89:392-396; U.S. Pat. Nos. 5,270,184 and 5,455,166), commonly referred to as SDA, which uses cycles of annealing pairs of primer sequences to opposite strands of a target sequence, primer extension in the presence of a dNTP $\alpha$ S to produce a duplex hemiphosphorothioated primer extension product, endonuclease-mediated nicking of a hemimodified restriction endonuclease recognition site, and polymerase-mediated primer extension from the 3' end of the nick to displace an existing strand and produce a strand for the next round of primer annealing, nicking and strand displacement, resulting in geometric amplification of product. Thermophilic SDA (tSDA) uses thermophilic endonucleases and polymerases at higher temperatures in essentially the same method (European Pat. No. 0 684 315).

**[00316]** Other amplification methods include, for example: nucleic acid sequence based amplification (U.S. Pat. No. 5,130,238), commonly referred to as NASBA; one that uses an RNA replicase to amplify the probe molecule itself (Lizardi, P. *et al.*, 1988, *BioTechnol.* 6: 1197-1202), commonly referred to as Q $\beta$  replicase; a transcription based amplification method (Kwoh, D. *et al.*, 1989, *Proc. Natl. Acad. Sci. USA* 86:1173-1177); self-sustained sequence replication (Guatelli, J. *et al.*, 1990, *Proc. Natl. Acad. Sci. USA* 87: 1874-1878); and, transcription mediated amplification (U.S. Pat. Nos. 5,480,784 and 5,399,491), commonly referred to as TMA. For further discussion of known amplification methods *see* Persing, David H., 1993, "In Vitro Nucleic Acid Amplification Techniques" in *Diagnostic Medical Microbiology: Principles and Applications* (Persing *et al.*, Eds.), pp. 51-87 (American Society for Microbiology, Washington, DC).

**[00317]** Illustrative transcription-based amplification systems of the present invention include TMA, which employs an RNA polymerase to produce multiple RNA transcripts of a target region (U.S. Pat. Nos. 5,480,784 and 5,399,491). TMA uses a "promoter-primer" that hybridizes to a target nucleic acid in the presence of a reverse transcriptase and an RNA polymerase to form a double-stranded promoter from which the RNA polymerase produces RNA transcripts. These transcripts can become

templates for further rounds of TMA in the presence of a second primer capable of hybridizing to the RNA transcripts. Unlike PCR, LCR or other methods that require heat denaturation, TMA is an isothermal method that uses an RNase H activity to digest the RNA strand of an RNA:DNA hybrid, thereby making the DNA strand available for hybridization with a primer or promoter-primer. Generally, the RNase H activity associated with the reverse transcriptase provided for amplification is used.

**[00318]** In an illustrative TMA method, one amplification primer is an oligonucleotide promoter-primer that comprises a promoter sequence which becomes functional when double-stranded, located 5' of a target-binding sequence, which is capable of hybridizing to a binding site of a target RNA at a location 3' to the sequence to be amplified. A promoter-primer may be referred to as a "T7-primer" when it is specific for T7 RNA polymerase recognition. Under certain circumstances, the 3' end of a promoter-primer, or a subpopulation of such promoter-primers, may be modified to block or reduce primer extension. From an unmodified promoter-primer, reverse transcriptase creates a cDNA copy of the target RNA, while RNase H activity degrades the target RNA. A second amplification primer then binds to the cDNA. This primer may be referred to as a "non-T7 primer" to distinguish it from a "T7-primer." From this second amplification primer, reverse transcriptase creates another DNA strand, resulting in a double-stranded DNA with a functional promoter at one end. When double-stranded, the promoter sequence is capable of binding an RNA polymerase to begin transcription of the target sequence to which the promoter-primer is hybridized. An RNA polymerase uses this promoter sequence to produce multiple RNA transcripts (*i.e.*, amplicons), generally about 100 to 1,000 copies. Each newly-synthesized amplicon can anneal with the second amplification primer. Reverse transcriptase can then create a DNA copy, while the RNase H activity degrades the RNA of this RNA:DNA duplex. The promoter-primer can then bind to the newly synthesized DNA, allowing the reverse transcriptase to create a double-stranded DNA, from which the RNA polymerase produces multiple amplicons. Thus, a billion-fold isothermic amplification can be achieved using two amplification primers.

**[00319]** In certain embodiments, other techniques may be used to evaluate RNA transcripts of the transcripts from a particular cDNA library, including microarray analysis (Han, M., *et al.*, *Nat Biotechnol*, 19: 631-635, 2001; Bao, P., *et al.*, *Anal Chem*, 74: 1792-1797, 2002; Schena *et al.*, *Proc. Natl. Acad. Sci. USA* 93:10614-19, 1996; and Heller *et al.*, *Proc. Natl. Acad. Sci. USA* 94:2150-55, 1997) and SAGE (serial analysis of gene expression). Like MPSS, SAGE is digital and can generate a large number of signature sequences. (see *e.g.*, Velculescu, V. E., *et al.*, *Trends Genet*, 16: 423-425., 2000; Tuteja R. and Tuteja N. *Bioessays*. 2004 Aug; 26(8):916-22),

although orders of magnitude fewer than that are available from techniques such as MPSS.

**[00320]** In certain embodiments, the term “microarray” includes a “nucleic acid microarray” having a substrate-bound plurality of nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Nucleic acid microarrays include all the devices so called in Schena (ed.), *DNA Microarrays: A Practical Approach* (Practical Approach Series), Oxford University Press (1999); Nature Genet. 21(1) (suppl.): 1-60 (1999); Schena (ed.), *Microarray Biochip: Tools and Technology*, Eaton Publishing Company/BioTechniques Books Division (2000). Nucleic acid microarrays may include a substrate-bound plurality of nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as described, for example, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4): 1665-1670 (2000). Examples of nucleic acid microarrays may be found in U.S. Pat. Nos. 6,391,623, 6,383,754, 6,383,749, 6,380,377, 6,379,897, 6,376,191, 6,372,431, 6,351,712, 6,344,316, 6,316,193, 6,312,906, 6,309,828, 6,309,824, 6,306,643, 6,300,063, 6,287,850, 6,284,497, 6,284,465, 6,280,954, 6,262,216, 6,251,601, 6,245,518, 6,263,287, 6,251,601, 6,238,866, 6,228,575, 6,214,587, 6,203,989, 6,171,797, 6,103,474, 6,083,726, 6,054,274, 6,040,138, 6,083,726, 6,004,755, 6,001,309, 5,958,342, 5,952,180, 5,936,731, 5,843,655, 5,814,454, 5,837,196, 5,436,327, 5,412,087, and 5,405,783, the disclosures of which are incorporated by reference.

**[00321]** Additional examples include nucleic acid arrays that are commercially available from Affymetrix (Santa Clara, Calif.) under the brand name GENECHIP™. Further exemplary methods of manufacturing and using arrays are provided in, for example, US. Pat. Nos. 7,028,629; 7,011,949; 7,011,945; 6,936,419; 6,927,032; 6,924,103; 6,921,642; and 6,818,394.

**[00322]** The present invention as related to arrays and microarrays also contemplates many uses for polymers attached to solid substrates. These uses include gene expression monitoring, profiling, library screening, genotyping and diagnostics. Gene expression monitoring and profiling methods and methods useful for gene expression monitoring and profiling are shown in U.S. Pat. Nos. 5,800,992, 6,013,449, 6,020,135, 6,033,860, 6,040,138, 6,177,248 and 6,309,822. Genotyping and uses therefore are shown in U.S. Ser. Nos. 10/442,021, 10/013,598 (U.S. Application No. 2003/0036069), and U.S. Pat. Nos. 5,925,525, 6,268,141, 5,856,092, 6,267,152, 6,300,063, 6,525,185, 6,632,611, 5,858,659, 6,284,460, 6,361,947, 6,368,799, 6,673,579 and 6,333,179. Other methods of nucleic acid amplification, labeling and



analysis that may be used in combination with the methods disclosed herein are embodied in U.S. Pat. Nos. 5,871,928, 5,902,723, 6,045,996, 5,541,061, and 6,197,506.

[00323] As will be apparent to persons skilled in the art, certain embodiments may employ oligonucleotides, such as primers or probes, for amplification or detection, as described herein. Oligonucleotides of a defined sequence and chemical structure may be produced by techniques known to those of ordinary skill in the art, such as by chemical or biochemical synthesis, and by *in vitro* or *in vivo* expression from recombinant nucleic acid molecules, *e.g.*, bacterial or viral vectors. In certain embodiments, an oligonucleotide does not consist solely of wild-type chromosomal DNA or the *in vivo* transcription products thereof.

[00324] Oligonucleotides or primers may be modified in any way, as long as a given modification is compatible with the desired function of a given oligonucleotide. One of ordinary skill in the art can easily determine whether a given modification is suitable or desired for any given oligonucleotide of the present invention. Relevant AARS oligonucleotides are described in greater detail elsewhere herein.

[00325] While the design and sequence of oligonucleotides depends on their function as described herein, several variables are generally taken into account. Among the most relevant are: length, melting temperature ( $T_m$ ), specificity, complementarity with other oligonucleotides in the system, G/C content, polypyrimidine (T, C) or polypurine (A, G) stretches, and the 3'-end sequence. Controlling for these and other variables is a standard and well known aspect of oligonucleotide design, and various computer programs are readily available to screen large numbers of potential oligonucleotides for optimal ones.

[00326] Certain embodiments therefore include methods for detecting a target AARS polynucleotide in a sample, the polynucleotide comprising the sequence of a reference AARS polynucleotide, as described herein, comprising a) hybridizing the sample with a probe comprising a sequence complementary to the target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. Also included are methods for detecting a target AARS polynucleotide in a sample, the polynucleotide comprising the sequence of a reference AARS polynucleotide, as described herein, comprising a) amplifying the target polynucleotide or fragment thereof, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof. Specific embodiments relate to the

detection of AARS splice variants, such as by detecting a unique splice junction of the splice variant, whether by hybridization, amplification, or other detection method.

[00327] Embodiments of the present invention include a variety of AARS polypeptide-based detection techniques, including antibody-based detection techniques. Included in these embodiments are the use of AARS polypeptides to generate antibodies or other binders, which may then be used in diagnostic methods and compositions to detect or quantitate selected AARS polypeptides in a cell or other biological sample, typically from a subject.

[00328] Certain embodiments may employ standard methodologies and detectors such as western blotting and immunoprecipitation, enzyme-linked immunosorbent assays (ELISA), flow cytometry, and immunofluorescence assays (IFA), which utilize an imaging device. These well-known methods typically utilize one or more monoclonal or polyclonal antibodies as described herein that specifically bind to a selected AARS polypeptide of the invention, or a unique region of that AARS polypeptide, and generally do not bind significantly to other AARS polypeptides, such as a full-length AARS polypeptide. In certain embodiments, the unique region of the AARS polypeptide may represent a unique three-dimensional structure that is possessed by a newly identified protein fragment of an AARS.

[00329] Certain embodiments may employ “arrays,” such as “microarrays.” In certain embodiments, a “microarray” may also refer to a “peptide microarray” or “protein microarray” having a substrate-bound collection or plurality of polypeptides, the binding to each of the plurality of bound polypeptides being separately detectable. Alternatively, the peptide microarray may have a plurality of binders, including but not limited to monoclonal antibodies, polyclonal antibodies, phage display binders, yeast 2 hybrid binders, and aptamers, which can specifically detect the binding of the AARS polypeptides described herein. The array may be based on autoantibody detection of these AARS polypeptides, as described, for example, in Robinson *et al.*, *Nature Medicine* 8(3):295-301 (2002). Examples of peptide arrays may be found in WO 02/31463, WO 02/25288, WO 01/94946, WO 01/88162, WO 01/68671, WO 01/57259, WO 00/61806, WO 00/54046, WO 00/47774, WO 99/40434, WO 99/39210, and WO 97/42507 and U.S. Pat. Nos. 6,268,210, 5,766,960, and 5,143,854, each of which are incorporated by reference.

[00330] Certain embodiments may employ MS or other molecular weight-based methods for diagnostically detecting AARS polypeptide sequences. Mass spectrometry (MS) refers generally to an analytical technique for determining the elemental composition of a sample or molecule. MS may also be used for determining the chemical structures of molecules, such as peptides and other chemical compounds.

[00331] Generally, the MS principle consists of ionizing chemical compounds to generate charged molecules or molecule fragments, and then measuring their mass-to-charge ratios. In an illustrative MS procedure: a sample is loaded onto the MS instrument, and undergoes vaporization, the components of the sample are ionized by one of a variety of methods (*e.g.*, by impacting them with an electron beam), which results in the formation of positively charged particles, the positive ions are then accelerated by a magnetic field, computations are performed on the mass-to-charge ratio ( $m/z$ ) of the particles based on the details of motion of the ions as they transit through electromagnetic fields, and, detection of the ions, which in step prior were sorted according to  $m/z$ .

[00332] An illustrative MS instruments has three modules: an ion source, which converts gas phase sample molecules into ions (or, in the case of electrospray ionization, move ions that exist in solution into the gas phase); a mass analyzer, which sorts the ions by their masses by applying electromagnetic fields; and a detector, which measures the value of an indicator quantity and thus provides data for calculating the abundances of each ion present.

[00333] The MS technique has both qualitative and quantitative uses, including identifying unknown compounds, determining the isotopic composition of elements in a molecule, and determining the structure of a compound by observing its fragmentation. Other uses include quantifying the amount of a compound in a sample or studying the fundamentals of gas phase ion chemistry (the chemistry of ions and neutrals in a vacuum). Included are gas chromatography-mass spectrometry (GC/MS or GC-MS), liquid chromatography mass spectrometry (LC/MS or LC-MS), and ion mobility spectrometry/mass spectrometry (IMS/MS or IMMS). Accordingly, MS techniques may be used according to any of the methods provided herein to measure the presence or levels of an AARS polypeptide of the invention in a biological sample, and to compare those levels to a control sample or a pre-determined value.

[00334] Certain embodiments may employ cell-sorting or cell visualization or imaging devices/techniques to detect or quantitate the presence or levels of AARS polynucleotides or polypeptides. Examples include flow cytometry or FACS, immunofluorescence analysis (IFA), and *in situ* hybridization techniques, such as fluorescent *in situ* hybridization (FISH).

[00335] Certain embodiments may employ conventional biology methods, software and systems for diagnostic purposes. Computer software products of the invention typically include computer readable medium having computer-executable instructions for performing the logic steps of the method of the invention. Suitable computer readable medium include floppy disk, CD-ROM/DVD/DVD-ROM, hard-disk drive,

flash memory, ROM/RAM, magnetic tapes and etc. The computer executable instructions may be written in a suitable computer language or combination of several languages. Basic computational biology methods are described in, for example Setubal and Meidanis *et al.*, Introduction to Computational Biology Methods (PWS Publishing Company, Boston, 1997); Salzberg, Searles, Kasif, (Ed.), Computational Methods in Molecular Biology, (Elsevier, Amsterdam, 1998); Rashidi and Buehler, Bioinformatics Basics: Application in Biological Science and Medicine (CRC Press, London, 2000) and Ouelette and Bzevanis Bioinformatics: A Practical Guide for Analysis of Gene and Proteins (Wiley & Sons, Inc., 2nd ed., 2001). See U.S. Pat. No. 6,420,108.

[00336] Certain embodiments may employ various computer program products and software for a variety of purposes, such as probe design, management of data, analysis, and instrument operation. See, U.S. Pat. Nos. 5,593,839, 5,795,716, 5,733,729, 5,974,164, 6,066,454, 6,090,555, 6,185,561, 6,188,783, 6,223,127, 6,229,911 and 6,308,170.

[00337] The whole genome sampling assay (WGSa) is described, for example in Kennedy *et al.*, *Nat. Biotech.* 21, 1233-1237 (2003), Matsuzaki *et al.*, *Gen. Res.* 14: 414-425, (2004), and Matsuzaki, *et al.*, *Nature Methods* 1:109-111 (2004). Algorithms for use with mapping assays are described, for example, in Liu *et al.*, *Bioinformatics.* 19: 2397-2403 (2003) and Di *et al.* *Bioinformatics.* 21:1958 (2005). Additional methods related to WGSa and arrays useful for WGSa and applications of WGSa are disclosed, for example, in U.S. Patent Application Nos. 60/676,058 filed Apr. 29, 2005, 60/616,273 filed Oct. 5, 2004, 10/912,445, 11/044,831, 10/442,021, 10/650,332 and 10/463,991. Genome wide association studies using mapping assays are described in, for example, Hu *et al.*, *Cancer Res.*; 65(7):2542-6 (2005), Mitra *et al.*, *Cancer Res.*, 64(21):8116-25 (2004), Butcher *et al.*, *Hum Mol Genet.*, 14(10):1315-25 (2005), and Klein *et al.*, *Science.* 308(5720):385-9 (2005).

[00338] Additionally, certain embodiments may include methods for providing genetic information over networks such as the Internet as shown, for example, in U.S. Application Nos. 10/197,621, 10/063,559 (United States Publication Number 2002/0183936), 10/065,856, 10/065,868, 10/328,818, 10/328,872, 10/423,403, and 60/482,389.

#### X. ANTISENSE AND RNAI AGENTS

[00339] Embodiments of the present invention also include antisense oligonucleotides and RNAi agents that target the AARS polynucleotide sequences, and methods of use thereof to reduce expression of a selected AARS transcript and/or protein fragment. Certain embodiments relate to targeting one or more splice junctions

(often unique) that generate a splice variant, AARS protein fragment of instant invention. Also included are methods of antisense or RNAi inhibition that target certain splice forms, either to encourage or discourage splicing of a selected protein fragment. In certain preferred embodiments, the splice junctions that generate the AARS protein fragments are over-expressed with respect to particular tissues, and are unique to that splice variant. In these and related embodiments, such splice variants are not the only source of cytosolic AARS activity in the targeted cell type. For instance, certain splice variants to be targeted may represent about 10% to 50% of the total copy number of the AARS RNA splice variants in a given cell or tissue, and preferably about 1-10% of the total copy number of the AARS RNA splice variants in a given cell or tissue. Splice variants that are about <1% of the total copy number of the AARS RNA splice variants in a given cell or tissue may also be targeted.

[00340] In certain embodiments, the antisense or RNAi agent does not target the full-length protein, because such full-length proteins are responsible for a key step in protein synthesis, and thereby avoids lethality that often results from wild-type AARS knockouts. Certain of the methods described herein can therefore be used to avoid undesired effects such as toxicities in both chronic and acute treatments, and to selectively modulate the non-canonical activities of the AARS protein fragment. However, certain embodiments may generically target AARS sequences, including full-length AARS sequences, such as to kill or substantially derange the cell physiology of a target cell or tissue.

[00341] In certain embodiments, the AARS splice variant to be targeted possesses a non-canonical biological activity. In some embodiments, the AARS splice variant has reduced or undetectable canonical AARS activity, and the antisense or RNAi-related method more specifically modulates its non-canonical activity. In certain embodiments, the antisense or RNAi-related agents can be combined with a targeted or local delivery approach to lessen systemic undesired effects to non-targeted cells or tissues. Among others described herein, exemplary cells or tissues that could be targeted this way include cancer cells, and cells to tissues that lend themselves to localized targeting, such as tumors or epithelia via topical application.

#### **A. Antisense Agents**

[00342] The terms “antisense oligomer” or “antisense compound” or “antisense oligonucleotide” are used interchangeably and refer to a sequence of cyclic subunits, each bearing a base-pairing moiety, linked by intersubunit linkages that allow the base-pairing moieties to hybridize to a target sequence in a nucleic acid (typically an RNA) by Watson-Crick base pairing, to form a nucleic acid:oligomer heteroduplex within the

target sequence, and typically thereby prevent translation of that RNA. Also included are methods of use thereof to modulate expression of a selected AARS transcript, such as a splice variant or proteolytic fragment, and/or its corresponding polypeptide.

**[00343]** Antisense oligonucleotides may contain between about 8 and 40 subunits, typically about 8-25 subunits, and preferably about 12 to 25 subunits. In certain embodiments, oligonucleotides may have exact sequence complementarity to the target sequence or near complementarity, as defined below. In certain embodiments, the degree of complementarity between the target and antisense targeting sequence is sufficient to form a stable duplex. The region of complementarity of the antisense oligomers with the target RNA sequence may be as short as 8-11 bases, but is preferably 12-15 bases or more, *e.g.*, 12-20 bases, or 12-25 bases, including all integers in between these ranges. An antisense oligomer of about 14-15 bases is generally long enough to have a unique complementary sequence in targeting the selected AARS gene. In certain embodiments, a minimum length of complementary bases may be required to achieve the requisite binding  $T_m$ , as discussed herein.

**[00344]** In certain embodiments, antisense oligomers as long as 40 bases may be suitable, where at least a minimum number of bases, *e.g.*, 10-12 bases, are complementary to the target sequence. In general, however, facilitated or active uptake in cells is optimized at oligomer lengths less than about 30. For certain oligomers, described further below, an optimum balance of binding stability and uptake generally occurs at lengths of 18-25 bases. Included are antisense oligomers (*e.g.*, PNAs, LNAs, 2'-OMe, MOE) that consist of about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 bases, in which at least about 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 contiguous or non-contiguous bases are complementary to their AARS target sequence, or variants thereof.

**[00345]** In certain embodiments, antisense oligomers may be 100% complementary to an AARS nucleic acid target sequence, or it may include mismatches, *e.g.*, to accommodate variants, as long as a heteroduplex formed between the oligomer and AARS nucleic acid target sequence is sufficiently stable to withstand the action of cellular nucleases and other modes of degradation which may occur *in vivo*. The term "target sequence" refers to a portion of the target RNA against which the oligonucleotide is directed, that is, the sequence to which the oligonucleotide will hybridize by Watson-Crick base pairing of a complementary sequence. In certain embodiments, the target sequence may be a contiguous region of an AARS mRNA (*e.g.*, a unique splice junction of an AARS mRNA), or may be composed of non-contiguous regions of the mRNA.

[00346] Oligomer backbones which are less susceptible to cleavage by nucleases are discussed below. Mismatches, if present, are less destabilizing toward the end regions of the hybrid duplex than in the middle. The number of mismatches allowed will depend on the length of the oligomer, the percentage of G:C base pairs in the duplex, and the position of the mismatch(es) in the duplex, according to well understood principles of duplex stability. Although such an antisense oligomer is not necessarily 100% complementary to the AARS nucleic acid target sequence, it is effective to stably and specifically bind to the target sequence, such that a biological activity of the nucleic acid target, *e.g.*, expression of AARS protein(s), is modulated.

[00347] The stability of the duplex formed between an oligomer and a target sequence is a function of the binding  $T_m$  and the susceptibility of the duplex to cellular enzymatic cleavage. The  $T_m$  of an antisense oligonucleotide with respect to complementary-sequence RNA may be measured by conventional methods, such as those described by Hames *et al.*, *Nucleic Acid Hybridization*, IRL Press, 1985, pp.107-108 or as described in Miyada C.G. and Wallace R.B., 1987, Oligonucleotide hybridization techniques, *Methods Enzymol.* Vol. 154 pp. 94-107. In certain embodiments, antisense oligomer may have a binding  $T_m$ , with respect to a complementary-sequence RNA, of greater than body temperature and preferably greater than 50°C.  $T_m$ 's in the range 60-80°C or greater are preferred. According to well known principles, the  $T_m$  of an oligomer compound, with respect to a complementary-based RNA hybrid, can be increased by increasing the ratio of C:G paired bases in the duplex, and/or by increasing the length (in base pairs) of the heteroduplex. At the same time, for purposes of optimizing cellular uptake, it may be advantageous to limit the size of the antisense oligomer. For this reason, compounds that show high  $T_m$  (50°C or greater) at a length of 25 bases or less are generally preferred over those requiring greater than 25 bases for high  $T_m$  values.

[00348] Antisense oligomers can be designed to block or inhibit translation of mRNA or to inhibit natural pre-mRNA splice processing, or induce degradation of targeted mRNAs, and may be said to be "directed to" or "targeted against" a target sequence with which it hybridizes. In certain embodiments, the target sequence may include any coding or non-coding sequence of an AARS mRNA transcript, and may thus be within an exon or within an intron. In certain embodiments, the target sequence is relatively unique or exceptional among AARSs (*e.g.*, a full-length AARS) and is selective for reducing expression of a selected AARS protein fragment, such as a proteolytic fragment or splice variant. In certain embodiments, the target site includes a 3' or 5' splice site of a pre-processed mRNA, or a branch point. The target sequence for a splice site may include an mRNA sequence having its 5' end 1 to about 25 to

about 50 base pairs downstream of a splice acceptor junction or upstream of a splice donor junction in a preprocessed mRNA. In certain embodiments, a target sequence may include a splice junction of an alternatively splice AARS mRNA, such as a splice junction that does not occur in the full-length AARS, or is unique or exceptional to that transcript, in that it either does not occur or only seldom occurs in other AARS splice variants. An oligomer is more generally said to be “targeted against” a biologically relevant target, such as reference AARS polynucleotide, when it is targeted against the nucleic acid of the target in the manner described herein.

**[00349]** An oligonucleotide is typically complementary to a target sequence, such as a target DNA or RNA. The terms “complementary” and “complementarity” refer to polynucleotides (*i.e.*, a sequence of nucleotides) related by the base-pairing rules. For example, the sequence “A-G-T,” is complementary to the sequence “T-C-A.” Complementarity may be “partial,” in which only some of the nucleic acids’ bases are matched according to the base pairing rules. Or, there may be “complete” or “total” complementarity (100%) between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. While perfect complementarity is often desired, some embodiments can include one or more but preferably 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 mismatches with respect to the target sequence. Variations at any location within the oligomer are included. In certain embodiments, variations in sequence near the termini of an oligomer are generally preferable to variations in the interior, and if present are typically within about 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 nucleotides of the 5’ and/or 3’ terminus.

**[00350]** The term “targeting sequence” or in certain embodiments “antisense targeting sequence” refers to the sequence in an oligonucleotide that is complementary (meaning, in addition, substantially complementary) to the target sequence in the DNA or RNA target molecule. The entire sequence, or only a portion, of the antisense compound may be complementary to the target sequence. For example, in an oligonucleotide having 20-30 bases, about 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, or 29 may be targeting sequences that are complementary to the target region. Typically, the targeting sequence is formed of contiguous bases, but may alternatively be formed of non-contiguous sequences that when placed together, *e.g.*, from opposite ends of the oligonucleotide, constitute sequence that spans the target sequence.

**[00351]** Target and targeting sequences are described as “complementary” to one another when hybridization occurs in an antiparallel configuration. A targeting sequence may have “near” or “substantial” complementarity to the target sequence and



still function for the purpose of the present invention, that is, it may still be functionally “complementary.” In certain embodiments, an oligonucleotide may have at most one mismatch with the target sequence out of 10 nucleotides, and preferably at most one mismatch out of 20. Alternatively, an oligonucleotide may have at least about 80%, 85%, 90% sequence homology, and preferably at least 95% sequence homology, with an AARS reference polynucleotide sequence described herein, or its complement.

[00352] An oligonucleotide “specifically hybridizes” to a target polynucleotide if the oligomer hybridizes to a target (*e.g.*, an AARS reference polynucleotide or its complement) under physiological conditions, with a  $T_m$  substantially greater than 45°C, preferably at least 50°C, and typically 60°C-80°C or higher. Such hybridization preferably corresponds to stringent hybridization conditions. At a given ionic strength and pH, the  $T_m$  is the temperature at which 50% of a target sequence hybridizes to a complementary polynucleotide. Again, such hybridization may occur with “near” or “substantial” complementarity of the antisense oligomer to the target sequence, as well as with exact complementarity.

[00353] The terms specifically binds or specifically hybridizes refer generally to an oligonucleotide probe or polynucleotide sequence that not only binds to its intended target gene sequence in a sample under selected hybridization conditions, but does not bind significantly to other target sequences in the sample, and thereby discriminates between its intended target and all other targets in the target pool. A probe that specifically hybridizes to its intended target sequence may also detect concentration differences under the selected hybridization conditions, as described herein.

[00354] A “nuclease-resistant” oligomeric molecule (oligomer) refers to one whose backbone is substantially resistant to nuclease cleavage, in non-hybridized or hybridized form; by common extracellular and intracellular nucleases in the body; that is, the oligomer shows little or no nuclease cleavage under normal nuclease conditions in the body to which the oligomer is exposed.

[00355] A “heteroduplex” refers to a duplex between an oligonucleotide and the complementary portion of a target polynucleotide, such as a target DNA or RNA. A “nuclease-resistant heteroduplex” refers to a heteroduplex formed by the binding of an oligomer to its complementary target, such that the heteroduplex is substantially resistant to *in vivo* degradation by intracellular and extracellular nucleases, such as RNaseH, which are capable of cutting double-stranded RNA/RNA or RNA/DNA complexes.

[00356] A “subunit” of an oligonucleotide refers to one nucleotide (or nucleotide analog) unit. The term may refer to the nucleotide unit with or without the attached intersubunit linkage, although, when referring to a “charged subunit”, the charge

typically resides within the intersubunit linkage (*e.g.*, a phosphate or phosphorothioate linkage or a cationic linkage).

**[00357]** The cyclic subunits of an oligonucleotide may be based on ribose or another pentose sugar or, in certain embodiments, alternate or modified groups. Examples of modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Also contemplated are peptide nucleic acids (PNAs), locked nucleic acids (LNAs), 2'-O-Methyl oligonucleotides (2'-OMe), 2'-methoxyethoxy oligonucleotides (MOE), among other oligonucleotides known in the art.

**[00358]** The purine or pyrimidine base pairing moiety is typically adenine, cytosine, guanine, uracil, thymine or inosine. Also included are bases such as pyridin-4-one, pyridin-2-one, phenyl, pseudouracil, 2,4,6-trimethoxy benzene, 3-methyl uracil, dihydrouridine, naphthyl, aminophenyl, 5-alkylcytidines (*e.g.*, 5-methylcytidine), 5-alkyluridines (*e.g.*, ribothymidine), 5-halouridine (*e.g.*, 5-bromouridine) or 6-azapyrimidines or 6-alkylpyrimidines (*e.g.* 6-methyluridine), propyne, quesosine, 2-thiouridine, 4-thiouridine, wybutosine, wybutoxosine, 4-acetyluridine, 5-(carboxyhydroxymethyl)uridine, 5'-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluridine,  $\beta$ -D-galactosylquosine, 1-methyladenosine, 1-methylinosine, 2,2-dimethylguanosine, 3-methylcytidine, 2-methyladenosine, 2-methylguanosine, N6-methyladenosine, 7-methylguanosine, 5-methoxyaminomethyl-2-thiouridine, 5-methylaminomethyluridine, 5-methylcarbonylmethyluridine, 5-methoxyuridine, 5-methyl-2-thiouridine, 2-methylthio-N6-isopentenyladenosine,  $\beta$ -D-mannosylquosine, uridine-5-oxyacetic acid, 2-thiocytidine, threonine derivatives and others (Burgin *et al.*, 1996, *Biochemistry*, 35, 14090; Uhlman & Peyman, *supra*). By "modified bases" in this aspect is meant nucleotide bases other than adenine (A), guanine (G), cytosine (C), thymine (T), and uracil (U), as illustrated above; such bases can be used at any position in the antisense molecule. Persons skilled in the art will appreciate that depending on the uses of the oligomers, Ts and Us are interchangeable. For instance, with other antisense chemistries such as 2'-O-methyl antisense oligonucleotides that are more RNA-like, the T bases may be shown as U.

[00359] As noted above, certain oligonucleotides provided herein include peptide nucleic acids (PNAs). Peptide nucleic acids (PNAs) are analogs of DNA in which the backbone is structurally homomorphous with a deoxyribose backbone, consisting of N-(2-aminoethyl) glycine units to which pyrimidine or purine bases are attached. PNAs containing natural pyrimidine and purine bases hybridize to complementary oligonucleotides obeying Watson-Crick base-pairing rules, and mimic DNA in terms of base pair recognition (Egholm, Buchardt *et al.* 1993). The backbone of PNAs is formed by peptide bonds rather than phosphodiester bonds, making them well-suited for antisense applications (see structure below). The backbone is uncharged, resulting in PNA/DNA or PNA/RNA duplexes that exhibit greater than normal thermal stability. PNAs are not recognized by nucleases or proteases.

[00360] PNAs may be produced synthetically using any technique known in the art. PNA is a DNA analog in which a polyamide backbone replaces the traditional phosphate ribose ring of DNA. Despite a radical structural change to the natural structure, PNA is capable of sequence-specific binding in a helix form to DNA or RNA. Characteristics of PNA include a high binding affinity to complementary DNA or RNA, a destabilizing effect caused by single-base mismatch, resistance to nucleases and proteases, hybridization with DNA or RNA independent of salt concentration and triplex formation with homopurine DNA. Panagene™ has developed its proprietary Bts PNA monomers (Bts; benzothiazole-2-sulfonyl group) and proprietary oligomerisation process. The PNA oligomerisation using Bts PNA monomers is composed of repetitive cycles of deprotection, coupling and capping. Panagene's patents to this technology include US 6969766, US 7211668, US 7022851, US 7125994, US 7145006 and US 7179896. Representative United States patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Pat. Nos. 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Further teaching of PNA compounds can be found in Nielsen *et al.*, Science, 1991, 254, 1497.

[00361] Also included are "locked nucleic acid" subunits (LNAs). The structures of LNAs are known in the art: for example, Wengel, *et al.*, Chemical Communications (1998) 455; Tetrahedron (1998) 54, 3607, and Accounts of Chem. Research (1999) 32, 301; Obika, *et al.*, Tetrahedron Letters (1997) 38, 8735; (1998) 39, 5401, and Bioorganic Medicinal Chemistry (2008)16, 9230.

[00362] Oligonucleotides may incorporate one or more LNAs; in some cases, the compounds may be entirely composed of LNAs. Methods for the synthesis of individual LNA nucleoside subunits and their incorporation into oligonucleotides are known in the art: U.S. Patents 7,572,582; 7,569,575; 7,084,125; 7,060,809; 7,053,207; 7,034,133; 6,794,499; and 6,670,461. Typical intersubunit linkers include

phosphodiester and phosphorothioate moieties; alternatively, non-phosphorous containing linkers may be employed. A preferred embodiment is an LNA containing compound where each LNA subunit is separated by a DNA subunit (*i.e.*, a deoxyribose nucleotide). Further preferred compounds are composed of alternating LNA and DNA subunits where the intersubunit linker is phosphorothioate.

**[00363]** Certain oligonucleotides may comprise morpholino-based subunits bearing base-pairing moieties, joined by uncharged or substantially uncharged linkages. The terms “morpholino oligomer” or “PMO” (phosphoramidate- or phosphorodiamidate morpholino oligomer) refer to an oligonucleotide analog composed of morpholino subunit structures, where (i) the structures are linked together by phosphorus-containing linkages, one to three atoms long, preferably two atoms long, and preferably uncharged or cationic, joining the morpholino nitrogen of one subunit to a 5' exocyclic carbon of an adjacent subunit, and (ii) each morpholino ring bears a purine or pyrimidine or an equivalent base-pairing moiety effective to bind, by base specific hydrogen bonding, to a base in a polynucleotide.

**[00364]** Variations can be made to this linkage as long as they do not interfere with binding or activity. For example, the oxygen attached to phosphorus may be substituted with sulfur (thiophosphorodiamidate). The 5' oxygen may be substituted with amino or lower alkyl substituted amino. The pendant nitrogen attached to phosphorus may be unsubstituted, monosubstituted, or disubstituted with (optionally substituted) lower alkyl. The purine or pyrimidine base pairing moiety is typically adenine, cytosine, guanine, uracil, thymine or inosine. The synthesis, structures, and binding characteristics of morpholino oligomers are detailed in U.S. Patent Nos. 5,698,685, 5,217,866, 5,142,047, 5,034,506, 5,166,315, 5,521,063, and 5,506,337, and PCT Appn. Nos. PCT/US07/11435 (cationic linkages) and US08/012804 (improved synthesis), all of which are incorporated herein by reference.

**[00365]** The morpholino subunits may also be linked by non-phosphorus-based intersubunit linkages, as described further below, where at least one linkage is modified with a pendant cationic group as described above. Other oligonucleotide analog linkages which are uncharged in their unmodified state but which could also bear a pendant amine substituent could be used. For example, a 5'-nitrogen atom on a morpholino ring could be employed in a sulfamide linkage or a urea linkage (where phosphorus is replaced with carbon or sulfur, respectively) and modified in a manner analogous to the 5'-nitrogen atom in structure (b3) above

**[00366]** Certain embodiments include substantially uncharged morpholino oligomers, such as a substantially uncharged phosphorodiamidate-linked morpholino oligomer. A substantially uncharged, phosphorus containing backbone in an

oligonucleotide analog is one in which a majority of the subunit linkages, *e.g.*, between 50-100%, typically at least 60% to 100% or 75% or 80% of its linkages, are uncharged at physiological pH, and contain a single phosphorous atom. Examples of morpholino oligonucleotides having phosphorus-containing backbone linkages include phosphoroamidate and phosphorodiamidate-linked morpholino oligonucleotides. Certain embodiments may contain positively charged groups at preferably about 10%-50% of their backbone linkages.

[00367] Properties of the morpholino-based subunits include, for example, the ability to be linked in a oligomeric form by stable, uncharged or positively charged backbone linkages, the ability to support a nucleotide base (*e.g.*, adenine, cytosine, guanine, thymidine, uracil and hypoxanthine) such that the polymer formed can hybridize with a complementary-base target nucleic acid, including target RNA,  $T_m$  values above about 45°C in relatively short oligonucleotides (*e.g.*, 10-15 bases), the ability of the oligonucleotide to be actively or passively transported into mammalian cells, and the ability of the antisense oligonucleotide:RNA heteroduplex to resist RNase and RNaseH degradation, respectively.

[00368] In certain embodiments, a substantially uncharged oligonucleotide may be modified to include charged linkages, *e.g.*, up to about 1 per every 2-5 uncharged linkages, such as about 4-5 per every 10 uncharged linkages. In certain embodiments, optimal improvement in antisense activity may be seen when about 25% of the backbone linkages are cationic. In certain embodiments, enhancement may be seen with a small number *e.g.*, 10-20% cationic linkages, or where the number of cationic linkages are in the range 50-80%, such as about 60%. In certain embodiments the cationic backbone charges may be further enhanced by distributing the bulk of the charges close of the “center-region” backbone linkages of the antisense oligonucleotide, *e.g.*, in a 20-mer oligonucleotide with 8 cationic backbone linkages, having at least 70% of these charged linkages localized in the 10 centermost linkages.

[00369] Oligonucleotides that target one or more portions of an AARS polynucleotide reference sequence or its complement may be used in any of the therapeutic, diagnostic, or drug screening methods described herein and apparent to persons skilled in the art.

### ***B. RNA Interference Agents***

[00370] Certain embodiments relate to RNA interference (RNAi) agents that target one or more mRNA transcripts of an aminoacyl-tRNA synthetase (AARS) reference polynucleotide, including fragments and splice variants thereof. Also included are

methods of use thereof to modulate the levels of a selected AARS transcript, such as an AARS splice variant or endogenous proteolytic fragment.

[00371] The term “double-stranded” means two separate nucleic acid strands comprising a region in which at least a portion of the strands are sufficiently complementary to hydrogen bond and form a duplex structure. The term “duplex” or “duplex structure” refers to the region of a double stranded molecule wherein the two separate strands are substantially complementary, and thus hybridize to each other. “dsRNA” refers to a ribonucleic acid molecule having a duplex structure comprising two complementary and anti-parallel nucleic acid strands (*i.e.*, the sense and antisense strands). Not all nucleotides of a dsRNA must exhibit Watson-Crick base pairs; the two RNA strands may be substantially complementary. The RNA strands may have the same or a different number of nucleotides.

[00372] In certain embodiments, a dsRNA is or includes a region which is at least partially complementary to the target RNA. In certain embodiments, the dsRNA is fully complementary to the target RNA. It is not necessary that there be perfect complementarity between the dsRNA and the target, but the correspondence must be sufficient to enable the dsRNA, or a cleavage product thereof, to direct sequence specific silencing, such as by RNAi cleavage of the target RNA. Complementarity, or degree of homology with the target strand, is typically most critical in the antisense strand. While perfect complementarity, particularly in the antisense strand, is often desired some embodiments can include one or more but preferably 6, 5, 4, 3, 2, or fewer mismatches with respect to the target RNA. The mismatches are most tolerated in the terminal regions, and if present are preferably in a terminal region or regions, *e.g.*, within 6, 5, 4, or 3 nucleotides of the 5' and/or 3' terminus. The sense strand need only be substantially complementary with the antisense strand to maintain the overall double-strand character of the molecule.

[00373] As used herein, “modified dsRNA” refers to a dsRNA molecule that comprises at least one alteration that renders it more resistant to nucleases (*e.g.*, protein kinase) than an identical dsRNA molecule that recognizes the same target RNA. Modified dsRNAs may include a single-stranded nucleotide overhang and/or at least one substituted nucleotide.

[00374] As used herein, a “nucleotide overhang” refers to the unpaired nucleotide or nucleotides that protrude from the duplex structure when a 3'-end of one RNA strand extends beyond the 5'-end of the other complementary strand, or vice versa. “Blunt” or “blunt end” means that there are no unpaired nucleotides at that end of the dsRNA, *i.e.*, no nucleotide overhang. A “blunt ended” dsRNA is a dsRNA that is double stranded over its entire length, *i.e.*, no nucleotide overhang at either end of the molecule.

[00375] The term “terminal base pair,” as used herein, refers to the last nucleotide base pair on one end of the duplex region of a double-stranded molecule. For example, if a dsRNA or other molecule is blunt ended (*i.e.*, has no nucleotide overhangs), the last nucleotide base pairs at both ends of the molecule are terminal base pairs. Where a dsRNA or other molecule has a nucleotide overhang at one or both ends of the duplex structure, the last nucleotide base pair(s) immediately adjacent the nucleotide overhang(s) is the terminal base pair at that end(s) of the molecule.

[00376] In certain embodiments, the methods provided herein may utilize double-stranded ribonucleic acid (dsRNA) molecules as modulating agents, for reducing expression of an AARS transcript such as a selected fragment or splice variant. dsRNAs generally comprise two single strands. One strand of the dsRNA comprises a nucleotide sequence that is substantially identical to a portion of the target gene or target region (the “sense” strand), and the other strand (the “complementary” or “antisense” strand) comprises a sequence that is substantially complementary to a portion of the target region. The strands are sufficiently complementary to hybridize to form a duplex structure. In certain embodiments, the complementary RNA strand may be less than 30 nucleotides, less than 25 nucleotides in length, or even 19 to 24 nucleotides in length. In certain aspects, the complementary nucleotide sequence may be 20-23 nucleotides in length, or 22 nucleotides in length.

[00377] In certain embodiments, at least one of the RNA strands comprises a nucleotide overhang of 1 to 4 nucleotides in length. In other embodiments, the dsRNA may further comprise at least one chemically modified nucleotide. In certain aspects, a dsRNA comprising a single-stranded overhang of 1 to 4 nucleotides may comprise a molecule wherein the unpaired nucleotide of the single-stranded overhang that is directly adjacent to the terminal nucleotide pair contains a purine base. In other aspects, the last complementary nucleotide pairs on both ends of a dsRNA are a G-C pair, or, at least two of the last four terminal nucleotide pairs are G-C pairs.

[00378] Certain embodiments of the present invention may comprise microRNAs. Micro-RNAs represent a large group of small RNAs produced naturally in organisms, some of which regulate the expression of target genes. Micro-RNAs are formed from an approximately 70 nucleotide single-stranded hairpin precursor transcript by Dicer. (V. Ambros *et al.* Current Biology 13:807, 2003). Certain micro-RNAs may be transcribed as hairpin RNA precursors, which are then processed to their mature forms by Dicer enzyme.

[00379] Certain embodiments may also employ short-interfering RNAs (siRNA). In certain embodiments, the first strand of the double-stranded oligonucleotide contains two more nucleoside residues than the second strand. In other embodiments, the first

strand and the second strand have the same number of nucleosides; however, the first and second strands may be offset such that the two terminal nucleosides on the first and second strands are not paired with a residue on the complementary strand. In certain instances, the two nucleosides that are not paired are thymidine residues.

**[00380]** Also included are short hairpin RNAs (shRNAs) and micro RNAs (miRNAs). A double-stranded structure of an shRNA is formed by a single self-complementary RNA strand, and RNA duplex formation may be initiated either inside or outside the cell. MicroRNAs (miRNAs) are small non-coding RNAs of 20–22 nucleotides, typically excised from ~70 nucleotide foldback RNA precursor structures known as pre-miRNAs.

**[00381]** In instances when the modulating agent comprises siRNA, the agent should include a region of sufficient homology to the target region, and be of sufficient length in terms of nucleotides, such that the siRNA agent, or a fragment thereof, can mediate down regulation of the target RNA. It will be understood that the term “ribonucleotide” or “nucleotide” can, in the case of a modified RNA or nucleotide surrogate, also refer to a modified nucleotide, or surrogate replacement moiety at one or more positions. Thus, an siRNA agent is or includes a region which is at least partially complementary to the target RNA, as described herein.

**[00382]** In addition, an siRNA modulating agent may be modified or include nucleoside surrogates. Single stranded regions of an siRNA agent may be modified or include nucleoside surrogates, *e.g.*, the unpaired region or regions of a hairpin structure, *e.g.*, a region which links two complementary regions, can have modifications or nucleoside surrogates. Modification to stabilize one or more 3'- or 5'-terminus of an siRNA agent, *e.g.*, against exonucleases, or to favor the antisense siRNA agent to enter into RISC are also useful. Modifications can include C3 (or C6, C7, C12) amino linkers, thiol linkers, carboxyl linkers, non-nucleotidic spacers (C3, C6, C9, C12, abasic, triethylene glycol, hexaethylene glycol), special biotin or fluorescein reagents that come as phosphoramidites and that have another DMT-protected hydroxyl group, allowing multiple couplings during RNA synthesis.

**[00383]** siRNA agents may include, for example, molecules that are long enough to trigger the interferon response (which can be cleaved by Dicer (Bernstein *et al.* 2001. *Nature*, 409:363-366) and enter a RISC (RNAi-induced silencing complex)), in addition to molecules which are sufficiently short that they do not trigger the interferon response (which molecules can also be cleaved by Dicer and/or enter a RISC), *e.g.*, molecules which are of a size which allows entry into a RISC, *e.g.*, molecules which resemble Dicer-cleavage products. An siRNA modulating agent, or a cleavage product thereof,



can down regulate a target gene, *e.g.*, by inducing RNAi with respect to a target RNA, preferably an AARS target such as a selected splice variant.

[00384] Each strand of an siRNA agent can be equal to or less than 35, 30, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, or 15 nucleotides in length. The strand is preferably at least 19 nucleotides in length. For example, each strand can be between 21 and 25 nucleotides in length. Preferred siRNA agents have a duplex region of 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotide pairs, and one or more overhangs, preferably one or two 3' overhangs, of 2-3 nucleotides.

[00385] In addition to homology to target RNA and the ability to down regulate a target gene, an siRNA agent may have one or more of the following properties: it may, despite modifications, even to a very large number, or all of the nucleosides, have an antisense strand that can present bases (or modified bases) in the proper three dimensional framework so as to be able to form correct base pairing and form a duplex structure with a homologous target RNA which is sufficient to allow down regulation of the target, *e.g.*, by cleavage of the target RNA; it may, despite modifications, even to a very large number, or all of the nucleosides, still have "RNA-like" properties, *i.e.*, it may possess the overall structural, chemical and physical properties of an RNA molecule, even though not exclusively, or even partly, of ribonucleotide-based content. For example, an siRNA agent can contain, *e.g.*, a sense and/or an antisense strand in which all of the nucleotide sugars contain *e.g.*, 2' fluoro in place of 2' hydroxyl. This deoxyribonucleotide-containing agent can still be expected to exhibit RNA-like properties. While not wishing to be bound by theory, the electronegative fluorine prefers an axial orientation when attached to the C2' position of ribose. This spatial preference of fluorine can, in turn, force the sugars to adopt a C<sub>3'</sub>-endo pucker. This is the same puckering mode as observed in RNA molecules and gives rise to the RNA-characteristic A-family-type helix. Further, since fluorine is a good hydrogen bond acceptor, it can participate in the same hydrogen bonding interactions with water molecules that are known to stabilize RNA structures. Generally, it is preferred that a modified moiety at the 2' sugar position will be able to enter into H-bonding which is more characteristic of the OH moiety of a ribonucleotide than the H moiety of a deoxyribonucleotide.

[00386] A "single strand RNAi agent" as used herein, is an RNAi agent which is made up of a single molecule. It may include a duplexed region, formed by intra-strand pairing, *e.g.*, it may be, or include, a hairpin or pan-handle structure. Single strand RNAi modulating agents are preferably antisense with regard to the target molecule. A single strand RNAi agent should be sufficiently long that it can enter the RISC and participate in RISC mediated cleavage of a target mRNA. A single strand RNAi agent

is at least 14, and more preferably at least 15, 20, 25, 29, 35, 40, or 50 nucleotides in length. It is preferably less than 200, 100, or 60 nucleotides in length.

[00387] Hairpin RNAi modulating agents may have a duplex region equal to or at least 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotide pairs. The duplex region may preferably be equal to or less than 200, 100, or 50, in length. Certain ranges for the duplex region are 15-30, 17 to 23, 19 to 23, and 19 to 21 nucleotides pairs in length. The hairpin may have a single strand overhang or terminal unpaired region, preferably the 3', and preferably of the antisense side of the hairpin. In certain embodiments, overhangs are 2-3 nucleotides in length.

[00388] Certain modulating agents utilized according to the methods provided herein may comprise RNAi oligonucleotides such as chimeric oligonucleotides, or "chimeras," which contain two or more chemically distinct regions, each made up of at least one monomer unit, *i.e.*, a nucleotide in the case of an oligonucleotide compound. These oligonucleotides typically contain at least one region wherein the oligonucleotide is modified so as to confer upon the oligonucleotide increased resistance to nuclease degradation, increased cellular uptake, and/or increased binding affinity for the target nucleic acid. Consequently, comparable results can often be obtained with shorter oligonucleotides when chimeric oligonucleotides are used, compared to phosphorothioate oligodeoxynucleotides. Chimeric oligonucleotides may be formed as composite structures of two or more oligonucleotides, modified oligonucleotides, oligonucleotides and/or oligonucleotide mimetics as described above. Such oligonucleotides have also been referred to in the art as hybrids or gapmers. Representative United States patents that teach the preparation of such hybrid structures include, but are not limited to, U.S. Pat. Nos. 5,013,830; 5,149,797; 5,220,007; 5,256,775; 5,366,878; 5,403,711; 5,491,133; 5,565,350; 5,623,065; 5,652,355; 5,652,356; 5,700,922; and 5,955,589, each of which is herein incorporated by reference. In certain embodiments, the chimeric oligonucleotide is RNA-DNA, DNA-RNA, RNA-DNA-RNA, DNA-RNA-DNA, or RNA-DNA-RNA-DNA, wherein the oligonucleotide is between 5 and 60 nucleotides in length.

[00389] In one aspect of the invention RNAi agents relate to an oligonucleotide comprising at least one ligand tethered to an altered or non-natural nucleobase. A large number of compounds can function as the altered base. The structure of the altered base is important to the extent that the altered base should not substantially prevent binding of the oligonucleotide to its target, *e.g.*, mRNA. In certain embodiments, the altered base is difluorotolyl, nitropyrryl, nitroimidazolyl, nitroindolyl, naphthalenyl, anthracenyl, pyridinyl, quinolinyl, pyrenyl, or the divalent radical of any one of the non-natural nucleobases described herein. In certain embodiments, the non-natural

nucleobase is difluorotolyl, nitropyrrolyl, or nitroimidazolyl. In certain embodiments, the non-natural nucleobase is difluorotolyl. A wide variety of ligands are known in the art and are amenable to the present invention. For example, the ligand can be a steroid, bile acid, lipid, folic acid, pyridoxal, B12, riboflavin, biotin, aromatic compound, polycyclic compound, crown ether, intercalator, cleaver molecule, protein-binding agent, or carbohydrate. In certain embodiments, the ligand is a steroid or aromatic compound. In certain instances, the ligand is cholesteryl.

[00390] In other embodiments, the RNAi agent is an oligonucleotide tethered to a ligand for the purposes of improving cellular targeting and uptake. For example, an RNAi agent may be tethered to an antibody, or antigen binding fragment thereof. As an additional example, an RNAi agent may be tethered to a specific ligand binding molecule, such as a polypeptide or polypeptide fragment that specifically binds a particular cell-surface receptor.

[00391] In other embodiments, the modulating agent comprises a non-natural nucleobase, as described herein. In certain instances, the ribose sugar moiety that naturally occurs in nucleosides is replaced with a hexose sugar. In certain aspects, the hexose sugar is an allose, altrose, glucose, mannose, gulose, idose, galactose, talose, or a derivative thereof. In a preferred embodiment, the hexose is a D-hexose. In certain instances, the ribose sugar moiety that naturally occurs in nucleosides is replaced with a polycyclic heteroalkyl ring or cyclohexenyl group. In certain instances, the polycyclic heteroalkyl group is a bicyclic ring containing one oxygen atom in the ring. In certain instances, the polycyclic heteroalkyl group is a bicyclo[2.2.1]heptane, a bicyclo[3.2.1]octane, or a bicyclo[3.3.1]nonane. Examples of modified RNAi agents also include oligonucleotides containing modified backbones or non-natural internucleoside linkages, as described herein.

[00392] The present invention further encompasses oligonucleotides employing ribozymes. Synthetic RNA molecules and derivatives thereof that catalyze highly specific endoribonuclease activities are known as ribozymes. (*see, e.g.*, U.S. Pat. No. 5,543,508 to Haseloff *et al.*, and U.S. Pat. No. 5,545,729 to Goodchild *et al.*) The cleavage reactions are catalyzed by the RNA molecules themselves. In naturally occurring RNA molecules, the sites of self-catalyzed cleavage are located within highly conserved regions of RNA secondary structure (Buzayan *et al.*, Proc. Natl. Acad. Sci. U.S.A., 1986, 83, 8859; Forster *et al.*, Cell, 1987, 50, 9). Naturally occurring autocatalytic RNA molecules have been modified to generate ribozymes which can be targeted to a particular cellular or pathogenic RNA molecule with a high degree of specificity. Thus, ribozymes serve the same general purpose as antisense oligonucleotides (*i.e.*, modulation of expression of a specific gene) and, like

oligonucleotides, are nucleic acids possessing significant portions of single-strandedness.

[00393] In certain instances, the RNAi agents or antisense oligonucleotides for use with the methods provided herein may be modified by non-ligand group. A number of non-ligand molecules have been conjugated to oligonucleotides in order to enhance the activity, cellular distribution, cellular targeting, or cellular uptake of the oligonucleotide, and procedures for performing such conjugations are available in the scientific literature. Such non-ligand moieties have included lipid moieties, such as cholesterol (Letsinger *et al.*, *Proc. Natl. Acad. Sci. USA*, 1989, 86:6553), arginine-rich peptides, cholic acid (Manoharan *et al.*, *Bioorg. Med. Chem. Lett.*, 1994, 4:1053), a thioether, *e.g.*, hexyl-5-tritylthiol (Manoharan *et al.*, *Ann. N.Y. Acad. Sci.*, 1992, 660:306; Manoharan *et al.*, *Bioorg. Med. Chem. Lett.*, 1993, 3:2765), a thiocholesterol (Oberhauser *et al.*, *Nucl. Acids Res.*, 1992, 20:533), an aliphatic chain, *e.g.*, dodecandiol or undecyl residues (Saison-Behmoaras *et al.*, *EMBO J.*, 1991, 10:111; Kabanov *et al.*, *FEBS Lett.*, 1990, 259:327; Svinarchuk *et al.*, *Biochimie*, 1993, 75:49), a phospholipid, *e.g.*, di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate (Manoharan *et al.*, *Tetrahedron Lett.*, 1995, 36:3651; Shea *et al.*, *Nucl. Acids Res.*, 1990, 18:3777), a polyamine or a polyethylene glycol chain (Manoharan *et al.*, *Nucleosides & Nucleotides*, 1995, 14:969), or adamantane acetic acid (Manoharan *et al.*, *Tetrahedron Lett.*, 1995, 36:3651), a palmityl moiety (Mishra *et al.*, *Biochim. Biophys. Acta*, 1995, 1264:229), or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety (Crooke *et al.*, *J. Pharmacol. Exp. Ther.*, 1996, 277:923). Representative United States patents that teach the preparation of such oligonucleotide conjugates have been listed above. Typical conjugation protocols involve the synthesis of oligonucleotides bearing an aminolinker at one or more positions of the sequence. The amino group is then reacted with the molecule being conjugated using appropriate coupling or activating reagents. The conjugation reaction may be performed either with the oligonucleotide still bound to the solid support or following cleavage of the oligonucleotide in solution phase. Purification of the oligonucleotide conjugate by HPLC typically affords the pure conjugate.

[00394] Additional examples of RNAi agents may be found in U.S. Application Publication Nos. 2007/0275465, 2007/0054279, 2006/0287260, 2006/0035254, 2006/0008822, which are incorporated by reference. Also included are vector delivery systems that are capable of expressing the AARS-targeting sequences described herein. Included are vectors that express siRNA or other duplex-forming RNA interference molecules.

[00395] A vector or nucleic acid construct system can comprise a single vector or plasmid, two or more vectors or plasmids, which together contain the total DNA to be introduced into the genome of the host cell, or a transposon. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. In the present case, the vector or nucleic acid construct is preferably one which is operably functional in a mammalian cell, such as a muscle cell. The vector can also include a selection marker such as an antibiotic or drug resistance gene, or a reporter gene (*i.e.*, green fluorescent protein, luciferase), that can be used for selection or identification of suitable transformants or transfectants. Exemplary delivery systems may include viral vector systems (*i.e.*, viral-mediated transduction) including, but not limited to, retroviral (*e.g.*, lentiviral) vectors, adenoviral vectors, adeno-associated viral vectors, and herpes viral vectors, among others known in the art.

#### XI. DRUG DISCOVERY

[00396] Certain embodiments relate to the use of AARS polypeptides, antibodies, or polynucleotides in drug discovery, typically to identify agents that modulate one or more of the non-canonical activities of the reference AARS polypeptide, *e.g.*, the AARS protein fragment. For example, certain embodiments include methods of identifying one or more “cellular binding partners” of an AARS reference polypeptide, such as a cellular protein, lipid, nucleic acid or other host molecule that directly or physically interacts with the AARS polypeptide. Particular examples include for example cell-surface receptors, such as GPCRs, protein-protein interaction domains, and extracellular or intracellular domains thereof.

[00397] Also included are methods of identifying host molecules that participate in one or more non-canonical activities of the AARS polypeptide, including molecules that directly or indirectly interact with the cellular binding partner, and either regulate its role in a non-canonical activity, or are regulated by the binding partner. Such host molecules include both upstream and downstream components of the non-canonical pathway, typically related by about 1, 2, 3, 4, 5 or more identifiable steps in the pathway, relative to the cellular binding partner/AARS protein interaction.

[00398] Certain aspects include methods of identifying a compound (*e.g.*, polypeptide) or other agent that agonizes or antagonizes the non-canonical activity of an AARS reference polypeptide or active variant thereof, such as by interacting with the AARS polypeptide and/or one or more of its cellular binding partners. Also included are methods of identifying agents that modulate the expression (*e.g.*, splicing) of AARS splice variants, or modulate the activity of proteases that otherwise regulate the production of endogenous AARS protein fragments (resectins) at the protein level.

[00399] Certain embodiments therefore include methods of identifying a binding partner of an AARS reference polypeptide, comprising a) combining the AARS polypeptide with a biological sample under suitable conditions, and b) detecting specific binding of the AARS polypeptide to a binding partner, thereby identifying a binding partner that specifically binds to the AARS reference polypeptide. Also included are methods of screening for a compound that specifically binds to an AARS reference polypeptide or a binding partner of the AARS polypeptide, comprising a) combining the polypeptide or the binding partner with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide or the binding partner to the test compound, thereby identifying a compound that specifically binds to the polypeptide or its binding partner. In certain embodiments, the compound is a polypeptide or peptide. In certain embodiments, the compound is a small molecule or other (*e.g.*, non-biological) chemical compound. In certain embodiments, the compound is a peptide mimetic.

[00400] Any method suitable for detecting protein-protein interactions may be employed for identifying cellular proteins that interact with an AARS reference polypeptide, interact with one or more of its cellular binding partners, or both. Examples of traditional methods that may be employed include co-immunoprecipitation, cross-linking, and co-purification through gradients or chromatographic columns of cell lysates or proteins obtained from cell lysates, mainly to identify proteins in the lysate that interact with the AARS polypeptide.

[00401] In these and related embodiments, at least a portion of the amino acid sequence of a protein that interacts with an AARS polypeptide or its binding partner can be ascertained using techniques well known to those of skill in the art, such as via the Edman degradation technique. *See, e.g.*, Creighton *Proteins: Structures and Molecular Principles*, W. H. Freeman & Co., N.Y., pp. 34-49, 1983. The amino acid sequence obtained may be used as a guide for the generation of oligonucleotide mixtures that can be used to screen for gene sequences encoding such proteins. Screening may be accomplished, for example, by standard hybridization or PCR techniques, as described herein and known in the art. Techniques for the generation of oligonucleotide mixtures and the screening are well known. *See, e.g.*, Ausubel *et al.* *Current Protocols in Molecular Biology* Green Publishing Associates and Wiley Interscience, N.Y., 1989; and Innis *et al.*, eds. *PCR Protocols: A Guide to Methods and Applications* Academic Press, Inc., New York, 1990.

[00402] Additionally, methods may be employed in the simultaneous identification of genes that encode the binding partner or other polypeptide. These methods include, for example, probing expression libraries, in a manner similar to the well known

technique of antibody probing of lambda-gt11 libraries, using labeled AARS protein, or another polypeptide, peptide or fusion protein, *e.g.*, a variant AARS polypeptide or AARS domain fused to a marker (*e.g.*, an enzyme, fluor, luminescent protein, or dye), or an Ig-Fc domain.

[00403] One method that detects protein interactions *in vivo*, the two-hybrid system, is described in detail for illustration only and not by way of limitation. One example of this system has been described (Chien *et al.*, *PNAS USA* 88:9578 9582, 1991) and is commercially available from Clontech (Palo Alto, Calif.).

[00404] Briefly, utilizing such a system, plasmids may be constructed that encode two hybrid proteins: one plasmid consists of nucleotides encoding the DNA-binding domain of a transcription activator protein fused to an AARS reference nucleotide sequence (or, in certain embodiments, its binding partner), or a variant thereof, and the other plasmid consists of nucleotides encoding the transcription activator protein's activation domain fused to a cDNA (or collection of cDNAs) encoding an unknown protein(s) that has been recombined into the plasmid as part of a cDNA library. The DNA-binding domain fusion plasmid and the activator cDNA library may be transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (*e.g.*, HBS or lacZ) whose regulatory region contains the transcription activator's binding site. Either hybrid protein alone cannot activate transcription of the reporter gene: the DNA-binding domain hybrid cannot because it does not provide activation function and the activation domain hybrid cannot because it cannot localize to the activator's binding sites. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product.

[00405] The two-hybrid system or other such methodology may be used to screen activation domain libraries for proteins that interact with the "bait" gene product. By way of example, and not by way of limitation, an AARS reference polypeptide or variant may be used as the bait gene product. An AARS binding partner may also be used as a "bait" gene product. Total genomic or cDNA sequences are fused to the DNA encoding an activation domain. This library and a plasmid encoding a hybrid of a bait AARS gene product fused to the DNA-binding domain are co-transformed into a yeast reporter strain, and the resulting transformants are screened for those that express the reporter gene.

[00406] A cDNA library of the cell line from which proteins that interact with bait AARS gene products are to be detected can be made using methods routinely practiced in the art. For example, the cDNA fragments can be inserted into a vector such that they are translationally fused to the transcriptional activation domain of GAL4. This

library can be co-transformed along with the bait gene-GAL4 fusion plasmid into a yeast strain, which contains a lacZ gene driven by a promoter that contains GAL4 activation sequence. A cDNA encoded protein, fused to GAL4 transcriptional activation domain, that interacts with bait gene product will reconstitute an active GAL4 protein and thereby drive expression of the HIS3 gene. Colonies, which express HIS3, can be detected by their growth on Petri dishes containing semi-solid agar based media lacking histidine. The cDNA can then be purified from these strains, and used to produce and isolate the bait AARS gene-interacting protein using techniques routinely practiced in the art.

**[00407]** Also included are three-hybrid systems, which allow the detection of RNA-protein interactions in yeast. *See, e.g., Hook et al., RNA*. 11:227-233, 2005. Accordingly, these and related methods can be used to identify a cellular binding partner of an AARS polypeptide, and to identify other proteins or nucleic acids that interact with the AARS polypeptide, the cellular binding partner, or both.

**[00408]** Certain embodiments relate to the use of interactome screening approaches. Particular examples include protein domain-based screening (*see, e.g., Boxem et al., Cell*. 134:534-545, 2008; and Yu *et al., Science*. 322:10-110, 2008).

**[00409]** As noted above, once isolated, binding partners can be identified and can, in turn, be used in conjunction with standard techniques to identify proteins or other compounds with which it interacts. Certain embodiments thus relate to methods of screening for a compound that specifically binds to the binding partner of an AARS reference polypeptide, comprising a) combining the binding partner with at least one test compound under suitable conditions, and b) detecting binding of the binding partner to the test compound, thereby identifying a compound that specifically binds to the binding partner. In certain embodiments, the test compound is a polypeptide. In certain embodiments, the test compound is a chemical compound, such as a small molecule compound or peptide mimetic.

**[00410]** Certain embodiments include methods of screening for a compound that modulates the activity of an AARS reference polypeptide, comprising a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide. Certain embodiments include methods of screening for a compound that modulates the activity of a binding partner of an AARS reference polypeptide, comprising a) combining the



polypeptide with at least one test compound under conditions permissive for the activity of the binding partner, b) assessing the activity of the binding partner in the presence of the test compound, and c) comparing the activity of the binding partner in the presence of the test compound with the activity of the binding partner in the absence of the test compound, wherein a change in the activity of the binding partner in the presence of the test compound is indicative of a compound that modulates the activity of the binding partner. Typically, these and related embodiments include assessing a selected non-canonical activity that is associated with the AARS polypeptide or its binding partner. Included are *in vitro* and *in vivo* conditions, such as cell culture conditions.

**[00411]** Certain embodiments include methods of screening a compound for effectiveness as a full or partial agonist of an AARS reference polypeptide or an active fragment or variant thereof, comprising a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample, typically by measuring an increase in the non-canonical activity of the AARS polypeptide. Certain methods include a) exposing a sample comprising a binding partner of the AARS polypeptide to a compound, and b) detecting agonist activity in the sample, typically by measuring an increase in the selected non-canonical activity of the AARS polypeptide. Certain embodiments include compositions that comprise an agonist compound identified by the method and a pharmaceutically acceptable carrier or excipient.

**[00412]** Also included are methods of screening a compound for effectiveness as a full or partial antagonist of an AARS reference polypeptide, comprising a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample, typically by measuring a decrease in the non-canonical activity of the AARS polypeptide. Certain methods include a) exposing a sample comprising a binding partner of the AARS polypeptide to a compound, and b) detecting antagonist activity in the sample, typically by measuring a decrease in the selected non-canonical activity of the AARS polypeptide. Certain embodiments include compositions that comprise an antagonist compound identified by the method and a pharmaceutically acceptable carrier or excipient.

**[00413]** In certain embodiments, *in vitro* systems may be designed to identify compounds capable of interacting with or modulating an AARS reference sequence or its binding partner. Certain of the compounds identified by such systems may be useful, for example, in modulating the activity of the pathway, and in elaborating components of the pathway itself. They may also be used in screens for identifying compounds that disrupt interactions between components of the pathway; or may disrupt such interactions directly. One exemplary approach involves preparing a reaction mixture of the AARS polypeptide and a test compound under conditions and

for a time sufficient to allow the two to interact and bind, thus forming a complex that can be removed from and/or detected in the reaction mixture

[00414] *In vitro* screening assays can be conducted in a variety of ways. For example, an AARS polypeptide, a cellular binding partner, or test compound(s) can be anchored onto a solid phase. In these and related embodiments, the resulting complexes may be captured and detected on the solid phase at the end of the reaction. In one example of such a method, the AARS polypeptide and/or its binding partner are anchored onto a solid surface, and the test compound(s), which are not anchored, may be labeled, either directly or indirectly, so that their capture by the component on the solid surface can be detected. In other examples, the test compound(s) are anchored to the solid surface, and the AARS polypeptide and/or its binding partner, which are not anchored, are labeled or in some way detectable. In certain embodiments, microtiter plates may conveniently be utilized as the solid phase. The anchored component (or test compound) may be immobilized by non-covalent or covalent attachments. Non-covalent attachment may be accomplished by simply coating the solid surface with a solution of the protein and drying. Alternatively, an immobilized antibody, preferably a monoclonal antibody, specific for the protein to be immobilized may be used to anchor the protein to the solid surface. The surfaces may be prepared in advance and stored.

[00415] To conduct an exemplary assay, the non-immobilized component is typically added to the coated surface containing the anchored component. After the reaction is complete, un-reacted components are removed (*e.g.*, by washing) under conditions such that any specific complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. For instance, where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; *e.g.*, using a labeled antibody specific for the previously non-immobilized component (the antibody, in turn, may be directly labeled or indirectly labeled with a labeled anti-Ig antibody).

[00416] Alternatively, the presence or absence of binding of a test compound can be determined, for example, using surface plasmon resonance (SPR) and the change in the resonance angle as an index, wherein an AARS polypeptide or a cellular binding partner is immobilized onto the surface of a commercially available sensorchip (*e.g.*, manufactured by Biacore<sup>TM</sup>) according to a conventional method, the test compound is contacted therewith, and the sensorchip is illuminated with a light of a particular wavelength from a particular angle. The binding of a test compound can also be measured by detecting the appearance of a peak corresponding to the test compound by

a method wherein an AARS polypeptide or a cellular binding partner is immobilized onto the surface of a protein chip adaptable to a mass spectrometer, a test compound is contacted therewith, and an ionization method such as MALDI-MS, ESI-MS, FAB-MS and the like is combined with a mass spectrometer (*e.g.*, double-focusing mass spectrometer, quadrupole mass spectrometer, time-of-flight mass spectrometer, Fourier transformation mass spectrometer, ion cyclotron mass spectrometer and the like).

[00417] In certain embodiments, cell-based assays, membrane vesicle-based assays, or membrane fraction-based assays can be used to identify compounds that modulate interactions in the non-canonical pathway of the selected AARS polypeptide. To this end, cell lines that express an AARS polypeptide and/or a binding partner, or a fusion protein containing a domain or fragment of such proteins (or a combination thereof), or cell lines (*e.g.*, COS cells, CHO cells, HEK293 cells, Hela cells etc.) that have been genetically engineered to express such protein(s) or fusion protein(s) can be used. Test compound(s) that influence the non-canonical activity can be identified by monitoring a change (*e.g.*, a statistically significant change) in that activity as compared to a control or a predetermined amount.

[00418] For embodiments that relate to antisense and RNAi agents, for example, also included are methods of screening a compound for effectiveness in altering expression of an AARS reference polynucleotide, comprising a) exposing a sample comprising the AARS reference polynucleotide to a compound such as a potential antisense oligonucleotide, and b) detecting altered expression of the AARS polynucleotide. In certain non-limiting examples, these and related embodiments can be employed in cell-based assays or in cell-free translation assays, according to routine techniques in the art. Also included are the antisense and RNAi agents identified by such methods.

[00419] Antibodies to AARS protein fragments can also be used in screening assays, such as to identify an agent that specifically binds to an AARS, confirm the specificity or affinity of an agent that binds to an AARS protein fragment, or identify the site of interaction between the agent and the AARS protein fragment. Included are assays in which the antibody is used as a competitive inhibitor of the agent. For instance, an antibody that specifically binds to an AARS protein fragment with a known affinity can act as a competitive inhibitor of a selected agent, and be used to calculate the affinity of the agent for the AARS protein fragment. Also, one or more antibodies that specifically bind to known epitopes or sites of an AARS protein fragment can be used as a competitive inhibitor to confirm whether or not the agent binds at that same site. Other variations will be apparent to persons skilled in the art.

[00420] Also included are any of the above methods, or other screening methods known in the art, which are adapted for high-throughput screening (HTS). HTS

typically uses automation to run a screen of an assay against a library of candidate compounds, for instance, an assay that measures an increase or a decrease in a non-canonical activity, as described herein.

[00421] Any of the screening methods provided herein may utilize small molecule libraries or libraries generated by combinatorial chemistry. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention. Examples of methods for the synthesis of molecular libraries can be found in: (Carell *et al.*, 1994a; Carell *et al.*, 1994b; Cho *et al.*, 1993; DeWitt *et al.*, 1993; Gallop *et al.*, 1994; Zuckermann *et al.*, 1994).

[00422] Libraries of compounds may be presented in solution (Houghten *et al.*, 1992) or on beads (Lam *et al.*, 1991), on chips (Fodor *et al.*, 1993), bacteria, spores (Ladner *et al.*, U.S. Pat. No. 5,223,409, 1993), plasmids (Cull *et al.*, 1992) or on phage (Cwirla *et al.*, 1990; Devlin *et al.*, 1990; Felici *et al.*, 1991; Ladner *et al.*, U.S. Pat. No. 5,223,409, 1993; Scott and Smith, 1990). Embodiments of the present invention encompass the use of different libraries for the identification of small molecule modulators of one or more AARS protein fragments, their cellular binding partners, and/or their related non-canonical activities. Libraries useful for the purposes of the invention include, but are not limited to, (1) chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides and/or organic molecules.

[00423] Chemical libraries consist of structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening. Natural product libraries are derived from collections of microorganisms, animals, plants, or marine organisms which are used to create mixtures for screening by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of plants or marine organisms. Natural product libraries include polyketides, non-ribosomal peptides, and variants (non-naturally occurring) thereof. *See, e.g.*, Cane *et al.*, *Science* 282:63-68, 1998. Combinatorial libraries may be composed of large numbers of peptides, oligonucleotides or organic compounds as a mixture. They are relatively easy to prepare by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods.

[00424] More specifically, a combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a

given compound length (*i.e.*, the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

[00425] For a review of combinatorial chemistry and libraries created therefrom, *see, e.g.*, Huc, I. and Nguyen, R. (2001) *Comb. Chem. High Throughput Screen* 4:53-74; Lepre, C. A. (2001) *Drug Discov. Today* 6:133-140; Peng, S. X. (2000) *Biomed. Chromatogr.* 14:430-441; Bohm, H. J. and Stahl, M. (2000) *Curr. Opin. Chem. Biol.* 4:283-286; Barnes, C. and Balasubramanian, S. (2000) *Curr. Opin. Chem. Biol.* 4:346-350; Lepre, Enjalbal, C, *et al.*, (2000) *Mass Spectrom Rev.* 19:139-161; Hall, D. G., (2000) *Nat. Biotechnol.* 18:262-262; Lazo, J. S., and Wipf, P. (2000) *J. Pharmacol. Exp. Ther.* 293:705-709; Houghten, R. A., (2000) *Ann. Rev. Pharmacol. Toxicol.* 40:273-282; Kobayashi, S. (2000) *Curr. Opin. Chem. Biol.* (2000) 4:338-345; Kopylov, A. M. and Spiridonova, V. A. (2000) *Mol. Biol. (Mosk)* 34:1097-1113; Weber, L. (2000) *Curr. Opin. Chem. Biol.* 4:295-302; Dolle, R. E. (2000) *J. Comb. Chem.* 2:383-433; Floyd, C. D., *et al.*, (1999) *Prog. Med. Chem.* 36:91-168; Kundu, B., *et al.*, (1999) *Prog. Drug Res.* 53:89-156; Cabilly, S. (1999) *Mol. Biotechnol.* 12:143-148; Lowe, G. (1999) *Nat. Prod. Rep.* 16:641-651; Dolle, R. E. and Nelson, K. H. (1999) *J. Comb. Chem.* 1:235-282; Czarnick, A. W. and Keene, J. D. (1998) *Curr. Biol.* 8:R705-R707; Dolle, R. E. (1998) *Mol. Divers.* 4:233-256; Myers, P. L., (1997) *Curr. Opin. Biotechnol.* 8:701-707; and Pluckthun, A. and Cortese, R. (1997) *Biol. Chem.* 378:443.

[00426] Devices for the preparation of combinatorial libraries are commercially available (*see, e.g.*, 357 MPS, 390 MPS, Advanced Chem Tech, Louisville Ky., Symphony, Rainin, Woburn, Mass., 433A Applied Biosystems, Foster City, Calif., 9050 Plus, Millipore, Bedford, Mass.). In addition, numerous combinatorial libraries are themselves commercially available (*see, e.g.*, ComGenex, Princeton, N.J., Asinex, Moscow, Ru, Tripos, Inc., St. Louis, Mo., ChemStar, Ltd., Moscow, RU, 3D Pharmaceuticals, Exton, Pa., Martek Biosciences, Columbia, Md., etc.).

### XII. METHODS OF USE

[00427] Embodiments of the present invention include therapeutic methods of treatment. Accordingly, the AARS agents described herein, including AARS polypeptides, AARS polynucleotides, AARS polynucleotide-based vectors, AARS expressing host cells, antisense oligonucleotides, RNAi agents, as well as binding agents such as peptides, antibodies and antigen-binding fragments, peptide mimetics and other small molecules, can be used to treat a variety of non-limiting diseases or conditions associated with the non-canonical activities of a reference AARS. Examples of such non-canonical activities include modulation of extracellular signaling,

modulation of cell proliferation, modulation of cell migration, modulation of cell differentiation (*e.g.*, hematopoiesis, neurogenesis, myogenesis, osteogenesis, and adipogenesis), modulation of apoptosis or other forms of cell death, modulation of angiogenesis, modulation of cell binding, modulation of cellular metabolism, modulation of cytokine production or activity, modulation of cytokine receptor activity, modulation of cellular uptake, or secretion, immunomodulation, modulation of inflammation, modulation of metabolic processes such as glucose control, and the like.

[00428] Included are polynucleotide-based therapies, such as antisense therapies and RNAi interference therapies, which typically relate to reducing the expression of a target molecule, such as an endogenous fragment of an AARS, or a cellular binding partner of an AARS polypeptide, which otherwise contributes to its non-canonical activity. Antisense or RNAi therapies typically antagonize the non-canonical activity, such as by reducing expression of the AARS reference polypeptide. Also included are polypeptides or peptides, antibodies or antigen-binding fragment, peptide mimetics, or other small molecule-based therapies, which either agonize or antagonize the non-canonical activity of an AARS reference polypeptide, such as by interacting directly with the AARS polypeptide, its cellular binding partner(s), or both.

[00429] These and related embodiments include methods of using the AARS agents or compositions of the present invention for treating a cell, tissue or subject. The cells or tissues that may be treated or modulated by the present invention are preferably mammalian cells or tissues, or more preferably human cells or tissues. Such cells or tissues can be of a healthy state or of a diseased state.

[00430] In certain embodiments, for example, methods are provided for modulating therapeutically relevant cellular activities including, but not limited to, cellular metabolism, cell differentiation, cell proliferation, cellular uptake, cell secretion, cell death, cell mobilization, cell migration, gene transcription, mRNA translation, cell impedance, immune responses, inflammatory responses, and the like, comprising contacting a cell with an AARS agent or composition as described herein. In certain embodiments, the cell is in a subject. Accordingly, the AARS compositions may be employed in treating essentially any cell or tissue or subject that would benefit from modulation of one or more such activities.

[00431] The AARS agents and compositions may also be used in any of a number of therapeutic contexts including, for example, those relating to the treatment or prevention of neoplastic diseases, immune system diseases or conditions (*e.g.*, autoimmune diseases and inflammation), infectious diseases, metabolic diseases, neuronal/neurological diseases, muscular/cardiovascular diseases, diseases associated with aberrant hematopoiesis, diseases associated with aberrant myogenesis, diseases

associated with aberrant neurogenesis, diseases associated with aberrant adipogenesis, diseases associated with aberrant osteogenesis, diseases associated with aberrant angiogenesis, diseases associated with aberrant cell survival, diseases associated with aberrant lipid uptake, diseases associated with aging (e.g. hearing loss, peripheral or autonomic neuropathies, senile dementia, retinopathy) and others.

[00432] For example, in certain illustrative embodiments, the AARS compositions of the invention may be used to modulate angiogenesis, *e.g.*, via modulation of endothelial cell proliferation and/or signaling. Endothelial cell proliferation and/or signaling may be monitored using an appropriate cell line (*e.g.*, human microvascular endothelial lung cells (HMVEC-L) and human umbilical vein endothelial cells (HUVEC)), and using an appropriate assay (*e.g.*, endothelial cell migration assays, endothelial cell proliferation assays, tube-forming assays, matrigel plug assays, etc.), many of which are known and available in the art.

[00433] Therefore, in related embodiments, the compositions of the invention may be employed in the treatment of essentially any cell or tissue or subject that would benefit from modulation of angiogenesis. For example, in some embodiments, a cell or tissue or subject experiencing or susceptible to angiogenesis (*e.g.*, an angiogenic condition) may be contacted with a suitable composition of the invention to inhibit an angiogenic condition. In other embodiments, a cell or tissue experiencing or susceptible to insufficient angiogenesis (*e.g.*, an angiostatic condition) may be contacted with an appropriate composition of the invention in order to interfere with angiostatic activity and/or promote angiogenesis.

[00434] Also included are methods of modulating hematopoiesis and related conditions. Examples of hematopoietic processes that may be modulated by the AARS polypeptides of the invention include, without limitation, the formation of myeloid cells (*e.g.*, erythroid cells, mast cells monocytes/macrophages, myeloid dendritic cells, granulocytes such as basophils, neutrophils, and eosinophils, megakaryocytes, platelets) and lymphoid cells (*e.g.*, natural killer cells, lymphoid dendritic cells, B-cells, and T-cells). Certain specific hematopoietic processes include erythropoiesis, granulopoiesis, lymphopoiesis, megakaryopoiesis, thrombopoiesis, and others. Also included are methods of modulating the trafficking or mobilization of hematopoietic cells, including hematopoietic stem cells, progenitor cells, erythrocytes, granulocytes, lymphocytes, megakaryocytes, and thrombocytes.

[00435] The methods of modulating hematopoiesis may be practiced *in vivo*, *in vitro*, *ex vivo*, or in any combination thereof. These methods can be practiced on any biological sample, cell culture, or tissue that contains hematopoietic stem cells, hematopoietic progenitor cells, or other stem or progenitor cells that are capable of

differentiating along the hematopoietic lineage (*e.g.*, adipose tissue derived stem cells). For *in vitro* and *ex vivo* methods, stem cells and progenitor cells, whether of hematopoietic origin or otherwise, can be isolated and/or identified according to the techniques and characteristics described herein and known in the art.

[00436] The compositions of the invention may also be useful as immunomodulators for treating anti- or pro-inflammatory indications by modulating the cells that mediate, either directly or indirectly, autoimmune and/or inflammatory diseases, conditions and disorders. The utility of the compositions of the invention as immunomodulators or modulators of inflammation can be monitored using any of a number of known and available techniques in the art including, for example, migration assays (*e.g.*, using leukocytes or lymphocytes) or cell viability assays (*e.g.*, using B-cells, T-cells, monocytes or NK cells).

[00437] "Inflammation" refers generally to the biological response of tissues to harmful stimuli, such as pathogens, damaged cells (*e.g.*, wounds), and irritants. The term "inflammatory response" refers to the specific mechanisms by which inflammation is achieved and regulated, including, merely by way of illustration, immune cell activation or migration, cytokine production, vasodilation, including kinin release, fibrinolysis, and coagulation, among others described herein and known in the art.

[00438] Clinical signs of chronic inflammation are dependent upon duration of the illness, inflammatory lesions, cause and anatomical area affected. (*see, e.g.*, Kumar *et al.*, Robbins Basic Pathology-8<sup>th</sup> Ed., 2009 Elsevier, London; Miller, LM, Pathology Lecture Notes, Atlantic Veterinary College, Charlottetown, PEI, Canada). Chronic inflammation is associated with a variety of pathological conditions or diseases, including, for example, allergies, Alzheimer's disease, anemia, aortic valve stenosis, arthritis such as rheumatoid arthritis and osteoarthritis, cancer, congestive heart failure, fibromyalgia, fibrosis, heart attack, kidney failure, lupus, pancreatitis, stroke, surgical complications, inflammatory lung disease, inflammatory bowel disease, atherosclerosis, neurological disorders, diabetes, metabolic disorders, obesity, and psoriasis, among others described herein and known in the art. Hence, AARS compositions may be used to treat or manage chronic inflammation, modulate any of one or more of the individual chronic inflammatory responses, or treat any one or more diseases or conditions associated with chronic inflammation.

[00439] Criteria for assessing the signs and symptoms of inflammatory and other conditions, including for purposes of making differential diagnosis and also for monitoring treatments such as determining whether a therapeutically effective dose has been administered in the course of treatment, *e.g.*, by determining improvement according to accepted clinical criteria, will be apparent to those skilled in the art and are



exemplified by the teachings of *e.g.*, Berkow *et al.*, eds., The Merck Manual, 16<sup>th</sup> edition, Merck and Co., Rahway, N.J., 1992; Goodman *et al.*, eds., Goodman and Gilman's The Pharmacological Basis of Therapeutics, 10<sup>th</sup> edition, Pergamon Press, Inc., Elmsford, N.Y., (2001); Avery's Drug Treatment: Principles and Practice of Clinical Pharmacology and Therapeutics, 3rd edition, ADIS Press, Ltd., Williams and Wilkins, Baltimore, MD. (1987); Ebadi, Pharmacology, Little, Brown and Co., Boston, (1985); Osolci *al.*, eds., Remington's Pharmaceutical Sciences, 18<sup>th</sup> edition, Mack Publishing Co., Easton, PA (1990); Katzung, Basic and Clinical Pharmacology, Appleton and Lange, Norwalk, CT (1992).

[00440] In other embodiments, the AARS compositions of the invention may be used to modulate cellular proliferation and/or survival and, accordingly, for treating or preventing diseases, disorders or conditions characterized by abnormalities in cellular proliferation and/or survival. For example, in certain embodiments, the AARS compositions may be used to modulate apoptosis and/or to treat diseases or conditions associated with abnormal apoptosis. Apoptosis can be monitored by any of a number of available techniques known and available in the art including, for example, assays that measure fragmentation of DNA, alterations in membrane asymmetry, activation of apoptotic caspases and/or release of cytochrome C and AIF.

[00441] The progress of these and other therapies (*e.g.*, *ex vivo* therapies) can be readily monitored by conventional methods and assays and based on criteria known to the physician or other persons of skill in the art.

### XIII. PHARMACEUTICAL FORMULATIONS, ADMINISTRATION AND KITS

[00442] Embodiments of the present invention include AARS polynucleotides, AARS polypeptides, host cells expressing AARS polypeptides, binding agents, modulatory agents, or other compounds described herein, formulated in pharmaceutically-acceptable or physiologically-acceptable solutions for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy. It will also be understood that, if desired, the compositions of the invention may be administered in combination with other agents as well, such as, *e.g.*, other proteins or polypeptides or various pharmaceutically-active agents. There is virtually no limit to other components that may also be included in the compositions, provided that the additional agents do not adversely affect the modulatory or other effects desired to be achieved.

[00443] In the pharmaceutical compositions of the invention, formulation of pharmaceutically-acceptable excipients and carrier solutions is well-known to those of skill in the art, as is the development of suitable dosing and treatment regimens for

using the particular compositions described herein in a variety of treatment regimens, including *e.g.*, oral, parenteral, intravenous, intranasal, subcutaneous, and intramuscular administration and formulation.

[00444] In certain applications, the pharmaceutical or therapeutic compositions of the invention do not stimulate an immune reaction. In other embodiments, the pharmaceutical or therapeutic compositions of the invention, typically comprising one or more AARS polypeptides or polynucleotides, stimulate an immune reaction, such as by serving as an adjuvant in a vaccine or related composition, or being present in a composition together with a separate adjuvant or agent stimulates an immune response.

[00445] In certain embodiments, the AARS agents such as AARS polypeptides, AARS polynucleotides, and antibodies have a solubility that is desirable for the particular mode of administration, such intravenous administration. Examples of desirable solubilities include at least about 1 mg/ml, at least about 10 mg/ml, at least about 25 mg/ml, and at least about 50 mg/ml.

[00446] In certain applications, the pharmaceutical compositions disclosed herein may be delivered via oral administration to a subject. As such, these compositions may be formulated with an inert diluent or with an assimilable edible carrier, or they may be enclosed in hard- or soft-shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet.

[00447] In certain circumstances it will be desirable to deliver the pharmaceutical compositions disclosed herein parenterally, subcutaneously, intravenously, intramuscularly, intra-arterially, intrathecally, intraparenchymally, intraventricularly, intraurethrally, intrasternally, intracranially, intrasynovially, or even intraperitoneally as described, for example, in U.S. Pat. No. 5,543,158; U.S. Pat. No. 5,641,515 and U.S. Pat. No. 5,399,363 (each specifically incorporated herein by reference in its entirety). Suitable devices for parenteral administration include needle (including microneedle) injectors, needle-free injectors, and infusion techniques.

[00448] Solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

[00449] The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (U.S. Pat. No. 5,466,468, specifically incorporated herein by reference in its entirety). In all cases the form should be sterile

and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and should be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (e.g., glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be facilitated by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

[00450] For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, a sterile aqueous medium that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion (see, e.g., *Remington's Pharmaceutical Sciences*, 15th Edition, pp. 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject. Moreover, for human administration, preparations should meet sterility, pyrogenicity, and the general safety and purity standards as required by FDA Office of Biologics standards.

[00451] Sterile injectable solutions can be prepared by incorporating the active compounds in the required amount in the appropriate solvent with the various other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active

ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[00452] The compositions disclosed herein may be formulated in a neutral or salt form. Pharmaceutically-acceptable salts, include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms such as injectable solutions, drug-release capsules, and the like.

[00453] As used herein, “carrier” includes any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

[00454] The phrase “pharmaceutically-acceptable” refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human. The preparation of an aqueous composition that contains a protein as an active ingredient is well understood in the art. Typically, such compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared. The preparation can also be emulsified.

[00455] In certain embodiments, the pharmaceutical compositions may be delivered by intranasal sprays, inhalation, and/or other aerosol delivery vehicles. Methods for delivering genes, polynucleotides, and peptide compositions directly to the lungs via nasal aerosol sprays have been described *e.g.*, in U.S. Pat. No. 5,756,353 and U.S. Pat. No. 5,804,212 (each specifically incorporated herein by reference in its entirety). Likewise, the delivery of drugs using intranasal microparticle resins (Takenaga *et al.*, 1998) and lysophosphatidyl-glycerol compounds (U.S. Pat. No. 5,725,871, specifically incorporated herein by reference in its entirety) are also well-known in the pharmaceutical arts. Likewise, transmucosal drug delivery in the form of a

polytetrafluoroethylene support matrix is described in U.S. Pat. No. 5,780,045 (specifically incorporated herein by reference in its entirety).

[00456] The pharmaceutical compositions may be formulated to be immediate and / or sustained release. Sustained release compositions include delayed, modified, pulsed, controlled, targeted and programmed release. Thus the compositions may be formulated as a suspension or as a solid, semi-solid, or thixotropic liquid for administration as an implanted depot providing sustained release of the AARS polynucleotides, AARS polypeptides, binding agents, modulatory agents and other active agents. Examples of such formulations include without limitation, drug-coated stents and semi-solids and suspensions comprising drug-loaded poly(DL-lactic-co-glycolic)acid (PGLA), poly(DL-lactide-co-glycolide) (PLG) or poly(lactide) (PLA) lamellar vesicles or microparticles, hydrogels (Hoffman AS: *Ann. N.Y. Acad. Sci.* 944: 62-73 (2001)), poly-amino acid nanoparticles systems, sold under the trademark MEDUSA® developed by Flamel Technologies Inc., non aqueous gel systems sold under the trademark ATRIGEL® developed by Atrix, Inc., and Sucrose Acetate Isobutyrate Extended Release formulations sold under the trademark SABER® developed by Durect Corporation, and lipid-based systems developed by SkyePharma and sold under the trademark DEPOFOAM®.

[00457] Sustained release devices capable of delivering desired doses of the pharmaceutical compositions over extended periods of time are known in the art. For example, US Pat. Nos. 5,034,229; 5,557,318; 5,110,596; 5,728,396; 5,985,305; 6,113,938; 6,156,331; 6,375,978; and 6,395,292; teach osmotically-driven devices capable of delivering an active agent formulation, such as a solution or a suspension, at a desired rate over an extended period of time (i.e., a period ranging from more than one week up to one year or more). Other exemplary sustained release devices include regulator-type pumps that provide constant flow, adjustable flow, or programmable flow of beneficial agent formulations, which are available from Medtronic including the Intrathecal pumps sold under the trademark SYNCHROMED INFUSION SYSTEM®, the Johnson and Johnson systems sold under the trademark CODMAN® division pumps, and INSET® technologies pumps. Further examples of devices are described in US Pat. Nos. 6,283,949; 5,976,109; 5,836,935; and 5,511,355.

[00458] In certain embodiments, the delivery may occur by use of liposomes, nanocapsules, microparticles, microspheres, lipid particles, vesicles, and the like, for the introduction of the compositions of the present invention into suitable host cells. In particular, the compositions of the present invention may be formulated for delivery either encapsulated in a lipid particle, a liposome, a vesicle, a nanosphere, a

nanoparticle or the like. The formulation and use of such delivery vehicles can be carried out using known and conventional techniques.

[00459] In certain embodiments, the agents provided herein may be attached to a pharmaceutically acceptable solid substrate, including biocompatible and biodegradable substrates such as polymers and matrices. Examples of such solid substrates include, without limitation, polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and  $\gamma$ -ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as poly(lactic-co-glycolic acid) (PLGA) and the LUPRON DEPOT™ (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), poly-D-(-)-3-hydroxybutyric acid, collagen, metal, hydroxyapatite, bioglass, aluminate, bioceramic materials, and purified proteins.

[00460] In one particular embodiment, the solid substrate comprises biodegradable polymers sold under the trademark ATRIGEL™ (QLT, Inc., Vancouver, B.C.). The ATRIGEL® drug delivery system consists of biodegradable polymers dissolved in biocompatible carriers. Pharmaceuticals may be blended into this liquid delivery system at the time of manufacturing or, depending upon the product, may be added later by the physician at the time of use. When the liquid product is injected into the subcutaneous space through a small gauge needle or placed into accessible tissue sites through a cannula, water in the tissue fluids causes the polymer to precipitate and trap the drug in a solid implant. The drug encapsulated within the implant is then released in a controlled manner as the polymer matrix biodegrades with time.

[00461] Pharmaceutical compositions for use in the present invention may also be administered topically, (intra)dermally, or transdermally to the skin or mucosa. Typical formulations for this purpose include gels, hydrogels, lotions, solutions, creams, ointments, dusting powders, dressings, foams, films, skin patches, wafers, implants, sponges, fibers, bandages, and microemulsions. Liposomes may also be used. Typical carriers include alcohol, water, mineral oil, liquid petrolatum, white petrolatum, glycerin, polyethylene glycol, and propylene glycol. Penetration enhancers may be incorporated—see, e.g., Finnin and Morgan: *J. Pharm. Sci.* 88(10): 955-958, (1999). Other means of topical administration include delivery by electroporation, iontophoresis, phonophoresis, sonophoresis, and microneedle or needle-free injection for example using the systems sold under the trademarks POWDERJECT™, and BIOJECT™.

[00462] Methods of formulation are well known in the art and are disclosed, for example, in Remington: The Science and Practice of Pharmacy, Mack Publishing

Company, Easton, Pa., 20th edition, ISBN: 0683306472 (2000). The compositions and agents provided herein may be administered according to the methods of the present invention in any therapeutically effective dosing regimen. The dosage amount and frequency are selected to create an effective level of the agent without harmful effects. The effective amount of a compound of the present invention will depend on the route of administration, the type of warm-blooded animal being treated, and the physical characteristics of the specific warm-blooded animal under consideration. These factors and their relationship to determining this amount are well known to skilled practitioners in the medical arts. This amount and the method of administration can be tailored to achieve optimal efficacy but will depend on such factors as weight, diet, concurrent medication and other factors which those skilled in the medical arts will recognize.

[00463] In particular embodiments, the amount of a composition or agent administered will generally range from a dosage of from about 0.1 to about 100 mg/kg/day, and typically from about 0.1 to 10 mg/kg where administered orally or intravenously. In particular embodiments, a dosage is 5 mg/kg or 7.5 mg/kg. In various embodiments, the dosage is about 50-2500 mg per day, 100-2500 mg/day, 300-1800 mg/day, or 500-1800 mg/day. In one embodiment, the dosage is between about 100 to 600 mg/day. In another embodiment, the dosage is between about 300 and 1200 mg/day. In particular embodiments, the composition or agent is administered at a dosage of 100 mg/day, 240 mg/day 300 mg/day, 600 mg/day, 1000 mg/day, 1200 mg/day, or 1800 mg/day, in one or more doses per day (*i.e.*, where the combined doses achieve the desired daily dosage). In related embodiments, a dosage is 100 mg bid, 150 mg bid, 240 mg bid, 300 mg bid, 500 mg bid, or 600 mg bid. In various embodiments, the composition or agent is administered in single or repeat dosing. The initial dosage and subsequent dosages may be the same or different.

[00464] In certain embodiments, a composition or agent is administered in a single dosage of 0.1 to 10 mg/kg or 0.5 to 5 mg/kg. In other embodiments, a composition or agent is administered in a dosage of 0.1 to 50 mg/kg/day, 0.5 to 20 mg/kg/day, or 5 to 20 mg/kg/day.

[00465] In certain embodiments, a composition or agent is administered orally or intravenously, *e.g.*, by infusion over a period of time of about, *e.g.*, 10 minutes to 90 minutes. In other related embodiments, a composition or agent is administered by continuous infusion, *e.g.*, at a dosage of between about 0.1 to about 10 mg/kg/hr over a time period. While the time period can vary, in certain embodiments the time period may be between about 10 minutes to about 24 hours or between about 10 minutes to about three days.

[00466] In particular embodiments, an effective amount or therapeutically effective amount is an amount sufficient to achieve a total concentration of the composition or agent in the blood plasma of a subject with a  $C_{\max}$  of between about 0.1  $\mu\text{g/ml}$  and about 20  $\mu\text{g/ml}$  or between about 0.3  $\mu\text{g/ml}$  and about 20  $\mu\text{g/ml}$ . In certain embodiments, an oral dosage is an amount sufficient to achieve a blood plasma concentration ( $C_{\max}$ ) of between about 0.1  $\mu\text{g/ml}$  to about 5  $\mu\text{g/ml}$  or between about 0.3  $\mu\text{g/ml}$  to about 3  $\mu\text{g/ml}$ . In certain embodiments, an intravenous dosage is an amount sufficient to achieve a blood plasma concentration ( $C_{\max}$ ) of between about 1  $\mu\text{g/ml}$  to about 10  $\mu\text{g/ml}$  or between about 2  $\mu\text{g/ml}$  and about 6  $\mu\text{g/ml}$ . In a related embodiment, the total concentration of an agent in the blood plasma of the subject has a mean trough concentration of less than about 20  $\mu\text{g/ml}$  and/or a steady state concentration of less than about 20  $\mu\text{g/ml}$ . In a further embodiment, the total concentration of an agent in the blood plasma of the subject has a mean trough concentration of less than about 10  $\mu\text{g/ml}$  and/or a steady state concentration of less than about 10  $\mu\text{g/ml}$ .

[00467] In yet another embodiment, the total concentration of an agent in the blood plasma of the subject has a mean trough concentration of between about 1 ng/ml and about 10  $\mu\text{g/ml}$  and/or a steady state concentration of between about 1 ng/ml and about 10  $\mu\text{g/ml}$ . In one embodiment, the total concentration of an agent in the blood plasma of the subject has a mean trough concentration of between about 0.3  $\mu\text{g/ml}$  and about 3  $\mu\text{g/ml}$  and/or a steady state concentration of between about 0.3  $\mu\text{g/ml}$  and about 3  $\mu\text{g/ml}$ .

[00468] In particular embodiments, a composition or agent is administered in an amount sufficient to achieve in the mammal a blood plasma concentration having a mean trough concentration of between about 1 ng/ml and about 10  $\mu\text{g/ml}$  and/or a steady state concentration of between about 1 ng/ml and about 10  $\mu\text{g/ml}$ . In related embodiments, the total concentration of the agent in the blood plasma of the mammal has a mean trough concentration of between about 0.3  $\mu\text{g/ml}$  and about 3  $\mu\text{g/ml}$  and/or a steady state concentration of between about 0.3  $\mu\text{g/ml}$  and about 3  $\mu\text{g/ml}$ .

[00469] In particular embodiments of the present invention, the effective amount of a composition or agent, or the blood plasma concentration of composition or agent is achieved or maintained, *e.g.*, for at least 15 minutes, at least 30 minutes, at least 45 minutes, at least 60 minutes, at least 90 minutes, at least 2 hours, at least 3 hours, at least 4 hours, at least 8 hours, at least 12 hours, at least 24 hours, at least 48 hours, at least 3 days, at least 4 days, at least 5 days, at least 6 days, at least one week, at least 2 weeks, at least one month, at least 2 months, at least 4 months, at least 6 months, at least one year, at least 2 years, or greater than 2 years.



[00470] In certain polypeptide-based embodiments, the amount of polypeptide administered will typically be in the range of about 0.1 µg/kg to about 0.1 mg/kg to about 50 mg/kg of patient body weight. Depending on the type and severity of the disease, about 0.1 µg/kg to about 0.1 mg/kg to about 50 mg/kg body weight (*e.g.*, about 0.1-15 mg/kg/dose) of polypeptide can be an initial candidate dosage for administration to the patient, whether, for example, by one or more separate administrations, or by continuous infusion. For example, a dosing regimen may comprise administering an initial loading dose of about 4 mg/kg, followed by a weekly maintenance dose of about 2 mg/kg of the polypeptide, or about half of the loading dose. However, other dosage regimens may be useful. A typical daily dosage might range from about 0.1 µg/kg to about 1 µg/kg to 100 mg/kg or more, depending on the factors mentioned above. For repeated administrations over several days or longer, depending on the condition, the treatment is sustained until a desired suppression of disease symptoms occurs.

[00471] In particular embodiments, the effective dosage achieves the blood plasma levels or mean trough concentration of a composition or agent described herein. These may be readily determined using routine procedures.

[00472] Embodiments of the present invention, in other aspects, provide kits comprising one or more containers filled with one or more of the polypeptides, polynucleotides, antibodies, multiunit complexes, compositions thereof, *etc.*, of the invention, as described herein. The kits can include written instructions on how to use such compositions (*e.g.*, to modulate cellular signaling, angiogenesis, cancer, inflammatory conditions, diagnosis *etc.*).

[00473] The kits herein may also include a one or more additional therapeutic agents or other components suitable or desired for the indication being treated, or for the desired diagnostic application. An additional therapeutic agent may be contained in a second container, if desired. Examples of additional therapeutic agents include, but are not limited to anti-neoplastic agents, anti-inflammatory agents, antibacterial agents, antiviral agents, angiogenic agents, *etc.*

[00474] The kits herein can also include one or more syringes or other components necessary or desired to facilitate an intended mode of delivery (*e.g.*, stents, implantable depots, *etc.*).

[00475] All publications, patent applications, and issued patents cited in this specification are herein incorporated by reference as if each individual publication, patent application, or issued patent were specifically and individually indicated to be incorporated by reference.

[00476] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily

apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims. The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of noncritical parameters that could be changed or modified to yield essentially similar results.

#### *XIV. EXAMPLES*

**[00477] GENERAL METHODS** Unless indicated otherwise in the examples below, the following general methods for gene optimization, small and large scale protein expression, protein purification, transcriptional profiling and screening were used to make and characterize the AARS polypeptides described in the Examples below.

#### **GENE SYNTHESIS AND CLONING INTO EXPRESSION VECTORS**

**[00478]** Polynucleotide sequences encoding epitope tagged versions of the AARS polypeptides were codon optimized and cloned into bacterial expression vectors using the methods listed below.

**[00479]** In method (1), *E. coli* codon-optimized DNA (Welch et al., PLoS ONE 4(9): e7007 doi:10.1371/journal.pone.0007002) encoding each AARS polypeptide is synthesized by DNA 2.0 (Menlo Park, CA), and two versions of each AARS polypeptide are synthesized, containing either an N-terminal, or C-terminal combined epitope tag comprising both a six histidine tag and V5 epitope tag.

**[00480]** DNA encoding the N-terminally tagged AARS polypeptides is synthesized with a 5' extension encoding in 5' to 3' orientation, a ribosome binding site (rbs (underlined below)), NdeI restriction site, six histidine tag, and a V5 epitope tag, (AGGAGGTAAACATATGCATCATCATCATCACGGTAAGCCTATCCCTA ACCCTTTGCTCGGTCTCGATTCTACG) (**SEQ. ID. No. 1**),

**[00481]** which is fused in frame to the predicted AARS polypeptide open reading frame. In cases where the AARS polypeptide comprises a predicted native initiation methionine (ATG) residue, or the first amino acid residue of the predicted AARS polypeptide is Met, this was deleted. At the end of the predicted AARS polypeptide open reading frame, two stop codons and a XhoI site (TAATGACTCGAG) (**SEQ. ID. No. 2**) are added.

**[00482]** DNA encoding the C-terminally tagged AARS polypeptides is synthesized with a 5' extension encoding a rbs (underlined below) and NdeI restriction site that either recapitulates the predicted native start codon for the AARS polypeptide, or inserts an ATG in frame with the predicted AARS polypeptide open reading frame,

(AGGAGATAAAACATATG) (SEQ. ID. No. 3). In different embodiments, the ribosome binding site can comprise the sequences “AGGAGGTAACAT” (SEQ. ID. No. 4), “AGGAGATAAAACAT” (SEQ. ID. No. 5), or GAAGGAGATATACAT (SEQ. ID. No. 6). At the 3' end of the predicted AARS polypeptide open reading frame, a 3' extension is synthesized which encodes in 5' to 3' order, a V5 epitope tag, six histidine tag, two stop codons and a XhoI site, (GGTAAGCCTATCCCTAACCCCTCTCCTCGGTCTCGATTCTACGCACCACCATCATCACCATTAATGACTCGAG) (SEQ. ID. No. 7),

[00483] which is fused in frame to the predicted AARS polypeptide open reading frame. If the AARS polypeptide included a predicted native stop codon, this was deleted.

[00484] Synthesized DNA sequences encoding the AARS polypeptides are subcloned into pJExpress411 vector (DNA 2.0). After sequencing to confirm synthesis of the correct product, expression vectors are transformed into bacteria for protein expression as described more fully below.

[00485] In method (2), *E. coli* codon-optimized DNA (Ermolaeva MD (2001) Curr. Iss. Mol. Biol. 3 (4) 91-7) encoding each AARS polypeptide is synthesized by GENEWIZ (South Plainfield, NJ). Each polynucleotide sequence encoding the AARS polypeptide was synthesized with short 5' and 3' extensions comprising unique restriction sites for subsequent cloning.

[00486] Specifically a BamHI restriction site was inserted at the 5' end of the predicted open reading frame. In cases where the AARS polypeptide comprises a predicted native initiation methionine residue (ATG), or the first amino acid residue of the predicted AARS polypeptide is Met, this was deleted. Additionally a XhoI restriction site was inserted at the 3' end of the predicted open reading frame. In cases where the AARS polypeptide comprises a predicted native stop codon, this was deleted.

[00487] After restriction digestion, the resulting DNA sequences are subcloned into modified pET-24b vectors (EMD, Gibbstown, NJ) containing either an N-terminal(pET24b\_N-6XHis/V5), or C-terminal (pET24b\_C-V5/6XHis) combined epitope tag comprising both a six histidine and V5 epitope tag (vector modification by GENEWIZ, (South Plainfield, NJ).

[00488] After restriction digestion, and cloning, the DNA encoding the N-tagged AARS polypeptide is cloned into the N-tagged vector (pET24b\_N-6XHis/V5), which comprises a 5' DNA sequence encoding six histidines and a V5 epitope tag, (CATATGCATCATCATCATCACGGTAAGCCTATCCCTAACCCCTCTCCTCGGTCTCGATTCTACGGGATCC) (SEQ. ID. No. 8),

[00489] in frame with an initiation codon (ATG) embedded within the NdeI restriction site. This 5' extension is fused to the predicted AARS polypeptide open reading frame through a short 2 amino acid linker (GS).

[00490] At the 3' end of the predicted open reading frame, the DNA encoding the N-tagged AARS polypeptide comprises a DNA sequence encoding a 2 amino acid extension (LE) followed by two termination codons (CTCGAGTAATGA) (SEQ. ID. No. 9).

[00491] After restriction digestion, and cloning, the DNA encoding the C-tagged AARS polypeptide cloned into the C-tagged vector (pET24b\_C-V5/6XHis), comprises a 5' sequence encoding an initiation codon (ATG) embedded within the NdeI restriction site which is fused to the predicted AARS polypeptide open reading frame through a short 2 amino acid linker (GS), (CATATGGGATCC) (SEQ. ID. No. 10).

[00492] At the 3' end of the predicted open reading frame, the DNA encoding the C-tagged AARS polypeptide comprises a 3' DNA sequence encoding a short linker 2 amino acid linker (LE) followed by a V5 epitope tag followed by six histidines, and two stop codons,

CTCGAGGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGCACC  
ACCACCACCACCACTAATGA (SEQ. ID. No. 11).

[00493] In method (3) (AARS polypeptides for in vivo and cell binding experiments) AARS polypeptides were prepared as described below:

[00494] Human full length AlaRS (Invitrogen Ultimate ORF collection Clone ID IOH10952) was subcloned via recombination into pET301 (Invitrogen) using an LR clonase recombination reaction. The resulting vector was used as a template to PCR amplify DNA encoding AlaRS adding a 5' NdeI site and 3' NotI site with the following oligos:

(5' – TGGCCGCATATGGACTCTACTCTAACAGC – 3') (SEQ. ID. No. 182)

(5' – TAGCGGCCGCGTTCTTTACATCCCCGAGG – 3') (SEQ. ID. No. 183)

[00495] The amplified product was subcloned into pET21a (Novagen) in frame with the C-terminal six histidine tag using these restriction sites. The NdeI restriction site recapitulates the predicted native start codon.

[00496] At the 3' end of the predicted open reading frame, the DNA encoding C-tagged AlaRS comprises a DNA sequence encoding a 5 amino acid extension (AAALE (SEQ. ID. No. 184)) followed by a six histidine tag, and a termination codon; (5'-GCGGCCGCACTCGAGCACCACCACCACCACCTGA-3') (SEQ. ID. No. 185). DNA was sequenced to confirm correct Polynucleotide sequence (Retrogen, San Diego).

[00497] **ssDNA Preparation:** To prepare pET21a-C-6XHis-AARS ssDNA, the dsDNA vector was transformed into CJ236 bacterial cells (NEB, cat no E4141S) and plated on ampicillin (100µg/mL) and chloramphenicol (30µg/mL) containing LB-Agar plates. Plates were incubated overnight at 37 °C. A colony was used to inoculate LB medium containing ampicillin and chloramphenicol and incubated overnight at 225 rpm and 37 °C. 20mL of LB containing ampicillin and chloramphenicol was inoculated with 20µL of the overnight culture and grown for 2 hours at 225 rpm and 37 °C. The culture was infected with 5e9 pfu of M13KO7 Helper Phage (NEB, cat no N0315S). After 1 hour, kanamycin was added to the culture at a final concentration of 50 µg/mL and incubated overnight at 225 rpm and 37 °C. Bacteria was separated and discarded from culture by two centrifugations at 1900 x g. ssDNA was precipitated by incubation overnight at 4 °C with final concentrations of 4% PEG-8000 and 500mM Sodium Acetate. ssDNA was centrifuged at 12000 x g and resuspended in 2mL LB medium. ssDNA was purified from the supernatant using Qiagen QIAprep M13 kit (Qiagen, cat no 27704).

[00498] **Kunkel mutagenesis:** Kunkel mutagenesis was performed to make AlaRS1<sup>C6</sup> using the oligo below:

(5'–

CCTGTGACAGCCACAATCCTCCGGATACCCTTCATATGTATATCTCCTTCTT  
AAAG-3') (SEQ. ID. No 186).

[00499] The oligo was diluted to 170ng/µL, and 340ng of the oligo was incubated with 5U PNK kinase (Roche, cat no 10633542001) in the presence of 1X PNK kinase buffer and 0.5mM ATP

[00500] This reaction was incubated at 37 °C for 1 hour.

[00501] 100ng of ssDNA vector was incubated with 34ng of kinased oligo in annealing buffer (20mM Tris, pH7.4, 2mM MgCl<sub>2</sub>, 50mM NaCl, final concentrations) for 5 minutes on a heat block at 75 °C. Reactions were allowed to cool to room temperature while contained in the heat block.

[00502] For elongation of the plasmid, 1U of T4 DNA Polymerase (Roche, cat no 11004786001) and 1U T4 DNA Ligase (Roche, cat no 10481220001) was added to 5µL of the annealing reaction. Additionally, synthesis buffer was added to a final concentration of 1.1mM dNTPs, 2.2mM ATP, 22.2mM Tris, pH 7.4, 11.1mM MgCl<sub>2</sub>, and 4.4mM DTT). This reaction was incubated on ice for 5 minutes and then at 37 °C for 90 minutes. 5µL of the elongation reaction was transformed into 100µL DH5a cells. Transformations were plated on Ampicillin plates and incubated overnight at 37 °C.

[00503] Individual colonies were used to inoculate LB medium containing ampicillin. Cultures were grown overnight at 37 °C. DNA plasmids were prepared using Qiagen Spin Miniprep kit (Qiagen, cat no 27106) and sequence verified (Retrogen, San Diego).

[00504] After sequencing to confirm synthesis of the correct product, expression vectors are transformed into bacteria for protein expression as described more fully below.

#### **AARS POLYPEPTIDE EXPRESSION, PURIFICATION AND BIOPHYSICAL CHARACTERIZATION**

[00505] 6xHis-tagged AARS polypeptides are expressed in bacteria in a medium-throughput format and /or in larger scale flask cultures depending upon the amount of protein required. AARS polypeptides are purified using affinity and ion exchange chromatography as described below, and as specified for specific experiments.

[00506] **Bacterial cultures:** 100 ng of expression vector comprising codon optimized DNA encoding each AARS polypeptide (as described above) is transformed into BL21(DE3) (EMD chemicals, cat. no. 69450) competent *E. coli* bacteria at 42°C for 30 seconds in PCR plates. C41(DE3) (Lucigen, cat. no. 60442), HMS174(DE3) (EMD chemicals, cat. no. 69453) and Origami2(DE3) (EMD chemicals, cat. no. 71345) strains are also evaluated. The plates are placed on ice for 2 minutes and 100 µL of SOC medium is added, followed by a 1-hour incubation at 37°C. 5 mL of auto-induction medium (EMD chemicals, cat. no. 71491) supplemented with kanamycin (100 µg/mL) is added into each well of a 24-well block (Qiagen, cat. no. 19583). The transformation reactions are added to the individual wells, the block is sealed with adhesive film (VWR, cat. no 60941-078) and incubated overnight at 250 rpm in a 37°C shaker. When low temperature (25°C) conditions are used, incubation is carried out for 48 hours instead.

[00507] For larger scale expression, 200 mL of auto-induction medium supplemented with kanamycin (100 µg/mL) is added into 500-mL Erlenmeyer flasks with vent caps (Corning, cat. no. 431401). The transformation reactions are added to the individual flasks and incubated for 30 hours at 250 rpm in a 37°C shaker.

[00508] **Protein Isolation:** After the culture reached stationary phase (typical OD<sub>600</sub> of 3-6), the blocks are centrifuged at 3600 x g for 10 minutes. The medium is carefully aspirated and the blocks are frozen at -80°C or -20°C for 10 minutes. The blocks are then allowed to thaw at room temperature and 1 mL lysis buffer (100 mL Bugbuster supplemented with 200 µL lysonase (EMD chemicals, cat. no 71370) and protease inhibitors “complete mini EDTA-free” (Roche, cat. no. 11 836 170 001)) is added into

each well. The pellets are resuspended by repeat pipetting until no clump is visible and transferred to eppendorf tubes, followed by a 10-20 minute incubation on a shaker at room temperature. After centrifugation at 16,000 g for 10 minutes at 4°C, the lysates are loaded onto a TurboFilter 96 Plate included in the Ni-NTA Superflow 96 BioRobot Kit (Qiagen, cat. no. 969261) and centrifuged at 500 g for 5-10 minutes.

**[00509]** For larger scale expression, the stationary phase culture is transferred into 500-mL bottles and centrifuged at 6,000 g for 10 minutes. The medium is decanted and the pellet is stored at -80°C or -20°C before further processing. The pellet is then allowed to thaw at room temperature and 20 mL lysis buffer is added into each bottle. The pellets are resuspended by repeat pipetting until no clump is visible, followed by 20 minute incubation on a shaker at room temperature. After centrifugation at 10,000 g for 30 minutes at 4°C, the lysates are transferred to clean tubes or bottles. If trace amounts of debris are carried over during the transfer, the sample is centrifuged again or passed through a 0.45 µm cellulose acetate membrane (Corning, cat. no. 430314) for further clarification.

**[00510] Affinity Purification:** A QIAFilter 96 Plate is loaded with 200 µl Ni-NTA Superflow slurry included in the Ni-NTA Superflow 96 BioRobot Kit and the resin is equilibrated by adding 600 µl binding buffer (20 mM sodium phosphate, 500 mM sodium chloride and 10 mM imidazole, pH 7.5). A vacuum of -15 in. Hg is applied until all the buffer has passed through the resin. The clarified cell lysates from the previous step are then loaded onto the QIAFilter® 96 Plate and allowed to bind for 5 minutes. A vacuum of -3 in. Hg is applied for approximately 5 minutes until all the samples have passed through the resin. The resin is then washed with 1 mL binding buffer, followed by two washes with 1 mL binding buffer containing 0.1% Triton X-100. The resin is then washed 10 times with 1 mL binding buffer without Triton X-100. The bound 6xHis-tagged AARS polypeptides are eluted with 450 µl elution buffer (20 mM sodium phosphate, 500 mM sodium chloride and 500 mM imidazole, pH 7.5) and stored at 4°C.

**[00511]** For larger scale expression, an empty disposable column "Poly-Prep" (Bio-rad, cat. no. 731-1550) is loaded with 1 mL Ni-NTA Superflow slurry (Qiagen, cat. no. 30450) and the 0.5 mL resin is equilibrated by adding 5 mL binding buffer. The clarified cell lysate from the previous step is then loaded onto the column and allowed to pass through by gravity. The resin is first washed with 50 mL binding buffer plus 0.1% Triton X-100, then washed with 50 mL binding buffer without Triton X-100. The bound 6xHis-tagged AARS polypeptides are eluted with 2 mL elution buffer and stored at 4°C.

**[00512] *Desalting and Polishing Steps:*** For AARS polypeptides with a molecular mass of >10 kDa, the Omega 10K membrane of an AcroPrep 96 filter plate (Pall, cat. no. 5034) is rinsed with 20  $\mu$ L 1X PBS and the plate is placed onto a vacuum manifold (>10 in Hg) until all the liquid passes through. The eluates from the previous step (Ni-NTA) are dispensed into each well and the vacuum applied until all the liquid passes through. These steps are repeated until the total eluate volume (450  $\mu$ L) has been processed. AARS polypeptides are recovered by adding 180  $\mu$ L of 1X PBS pH 7.4 to each well, pipetting up and down 10 times carefully and then transferred to a clean block. This step is repeated to yield a total volume of 360  $\mu$ L per well and the block is stored at 4°C. For AARS polypeptides with a molecular mass of <10 kDa, the eluates from Ni-NTA are loaded onto an Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-3 membrane (Millipore, cat. no. UFC900308), followed by the addition of 10 mL 1X PBS and a centrifugation at 3,600 g for 10-30 minutes until the volume is less than 360  $\mu$ L. The samples are recovered and 1X PBS is added to a final volume of 360  $\mu$ L.

**[00513]** In order to remove endotoxins, an AcroPrep Advance filter plate with Mustang Q membrane (Pall, cat. no. 8171) is rinsed with 300  $\mu$ L of 1X PBS and centrifuged at 1,000 g for 5 minutes to remove the buffer. The desalted AARS polypeptides (360  $\mu$ L/well) are added to the filter plate and incubated on a shaker for 5-10 minutes. The plate is then centrifuged at 1,000 g for 5-10 minutes and the flow through fractions containing the AARS polypeptides are collected and stored at 4°C.

**[00514]** For larger scale expression, the eluates from Ni-NTA are loaded onto an Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-3 or Ultracel-10 membrane (Millipore, cat. no. UFC900308 or UFC901008) depending on the molecular weight of the AARS polypeptide and then centrifuged at 3,600 g for 10-30 minutes until the volume is reduced to 250  $\mu$ L. The samples are mixed in 10 mL 1X PBS, pH7.4 and centrifuged again at 3,600 g for 10-30 minutes until the volume is about 250  $\mu$ L. This step is repeated one more time, the supernatants are recovered and 1X PBS is added to a final volume of 1.5 mL.

**[00515]** In order to remove endotoxins, a Sartobind Q 5 strong anion exchanger membrane (Sartorius, cat. no. Q5F) is flushed with 1 mL 1X PBS and the AARS polypeptides are slowly passed through the membrane using a plastic syringe. The flow through fraction containing the AARS polypeptides is collected in a 96-deep well block that is sealed and stored at 4°C.

**[00516]** 6xHis-tagged AARS polypeptides expressed in bacteria and found in inclusion bodies are purified using affinity chromatography and a series of refolding steps, as described below.



[00517] **Bacterial cultures:** 100 ng of plasmid encoding each AARS polypeptide is transformed into BL21(DE3) (EMD chemicals, cat. no. 69450) or C41(DE3) (Lucigen, cat. no. 60442) competent *E. coli* bacteria at 42°C for 30 seconds in PCR plates. The plates are placed on ice for 2 minutes and 100 µl of SOC medium is added, followed by a 1-hour incubation at 37°C. 5 ml of auto-induction medium (EMD chemicals, cat. no. 71491) supplemented with kanamycin (100 µg/ml) is added into each well of a 24-well block (Qiagen, cat. no. 19583). The transformation reactions are added to the individual wells, the block is sealed with adhesive film (VWR, cat. no 60941-078) and incubated overnight at 250 rpm in a 37°C shaker.

[00518] For larger scale expression, 200 ml of auto-induction medium supplemented with kanamycin (100 µg/ml) is added into 500-ml Erlenmeyer flasks with vent caps (Corning, cat. no. 431401). The transformation reactions are added to the individual flasks and incubated for 30 hours at 250 rpm in a 37°C shaker.

[00519] **Isolation:** After the cultures reach stationary phase (typical OD<sub>600</sub> of 3-6), the blocks are centrifuged at 3,600 x g for 10 minutes. The medium is carefully aspirated and the blocks are frozen at -80°C or -20°C for 10 minutes. The blocks are then allowed to thaw at room temperature and 1 ml lysis buffer (100 ml Bugbuster supplemented with 200 µl lysonase (EMD chemicals, cat. no 71370) and protease inhibitor “complete mini EDTA-free” (Roche, cat. no. 11 836 170 001)) is added into each well. The pellets are resuspended by repeat pipetting until no clump is visible and transferred to eppendorf tubes, followed by a 10-20 minute incubation on a shaker at room temperature. After centrifugation at 16,000 x g for 10 minutes at 4°C, the soluble lysates are discarded and the inclusion bodies are thoroughly resuspended in denaturing binding buffer (20 mM sodium phosphate, 500 mM sodium chloride, 6 M guanidine hydrochloride, 10 mM imidazole, pH 7.5). The samples are centrifuged at 16,000 g for 10 minutes and the supernatants loaded onto a TurboFilter 96 Plate included in the Ni-NTA Superflow 96 BioRobot Kit (Qiagen, cat. no. 969261) followed by centrifugation at 500 g for 5-10 minutes. The filtrates are collected in a clean 96-well block (Greiner, cat. no. 780286).

[00520] For larger scale expression, the stationary phase culture is transferred into 500-ml bottles and centrifuged at 6,000 g for 10 minutes. The medium is decanted and the pellet is stored at -80°C or -20°C before further processing. The pellet is then allowed to thaw at room temperature and 20 ml lysis buffer is added into each bottle. The pellets are resuspended by repeat pipetting until no clump is visible, followed by 20 minute incubation on a shaker at room temperature. After centrifugation at 10,000 g for 30 minutes at 4°C, the soluble lysates are discarded and the insoluble inclusion bodies thoroughly resuspended in denaturing binding buffer.

**[00521] Affinity Purification:** A QIAFilter 96 Plate is loaded with 200  $\mu$ l Ni-NTA Superflow slurry included in the Ni-NTA Superflow 96 BioRobot Kit and the resin is equilibrated by adding 600  $\mu$ l denaturing binding buffer (see above). A vacuum of -15 in. Hg is applied until all the buffer passes through the resin. The clarified denatured samples from the previous step are then loaded onto the QIAFilter® 96 Plate and allowed to bind for 5 minutes. A vacuum of -3 in. Hg is applied for approximately 5 minutes until all the samples pass through the resin. The resin is then washed with 1 ml denaturing binding buffer, followed by five washes with 1 ml denaturing binding buffer containing 0.1% Triton 100. The resin is then washed 15 times with 1 ml denaturing binding buffer without Triton 100. The bound 6xHis-tagged AARS polypeptides are then eluted with 450  $\mu$ l denaturing elution buffer (20 mM sodium phosphate, 500 mM sodium chloride, 6 M guanidine hydrochloride and 500 mM imidazole, pH 7.5) and stored at 4°C.

**[00522]** For larger scale expression, an empty disposable column “Poly-Prep” (Bio-rad, cat. no. 731-1550) is loaded with 1 ml Ni-NTA Superflow slurry (Qiagen, cat. no. 30450) and the 0.5 ml resin is equilibrated by adding 5 ml denaturing binding buffer (see above). The denatured inclusion bodies from the previous step are then loaded onto the column and allowed to pass through by gravity. The resin is first washed with 50 ml denaturing binding buffer plus 0.1% Triton 100, then washed with 50 ml denaturing binding buffer without Triton 100. The bound 6xHis-tagged AARS polypeptides are eluted with 2 ml denaturing elution buffer and stored at 4°C.

**[00523] Refolding:** For AARS polypeptides >10 kDa, the Omega 10K membrane of an AcroPrep 96 filter plate (Pall, cat. no. 5034) is rinsed with 20  $\mu$ l 1X PBS and the plate is placed onto a vacuum manifold (>10 in. Hg) until all the liquid passes through. The eluates from the previous step (Ni-NTA) are dispensed into each well and the vacuum applied until all the liquid passes through. These steps are repeated until the total eluate volume (450  $\mu$ l) has been processed. AARS polypeptides are recovered by adding 200  $\mu$ l of refolding buffer containing 50 mM Tris, 250 mM sodium chloride, 10 mM potassium chloride, 2 mM magnesium chloride, 2 mM calcium chloride, 400 mM sucrose, 500 mM arginine, 1 mM DTT and 0.01% polysorbate 80, pH 7.4) to each well, pipetting up and down 10 times carefully, and then transferred to a clean block. This step is repeated to yield a total volume of 400  $\mu$ l per well and the block is placed on the shaker overnight at 4°C. For AARS polypeptides <10 kDa, the eluates from Ni-NTA are loaded onto an Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-3 membrane (Millipore, cat. no. UFC900308), followed by the addition of 10ml refolding buffer and a centrifugation at 3,600 g for 10-30 minutes until the volume is less than 400  $\mu$ l. The samples are recovered and extra refolding buffer is added to a final volume of 400  $\mu$ l.

The samples are transferred to a 96-well block, sealed with film and placed on a shaker overnight at 4°C.

**[00524]** For larger scale cultures, the eluates from Ni-NTA are loaded onto an Amicon Ultra-15 centrifugal filter unit with Ultracel-3 or Ultracel-10 membrane (Millipore, cat. no. UFC900308 or UFC901008 depending on the molecular weight of the AARS polypeptide) and then centrifuged at 3,600 g for 10-30 minutes until the volume is reduced to about 500 µl. For AARS polypeptides with pI > 7, the samples are diluted 20-fold in the following buffer: 50 mM sodium acetate, 10 mM sodium chloride, 0.4 mM potassium chloride, 1 mM EDTA, 400 mM sucrose, 500 mM arginine, 1 mM DTT and 0.01% polysorbate 80, pH 6.0. For AARS polypeptides with pI < 7, the samples are diluted 20-fold in the following buffer: 50 mM Tris, 250 mM sodium chloride, 10 mM potassium chloride, 2 mM magnesium chloride, 2 mM calcium chloride, 400 mM sucrose, 500 mM arginine, 1mM DTT and 0.01% polysorbate 80, pH 8.0. The samples are incubated on a shaker at 4°C overnight.

**[00525] *Desalting and Polishing Steps:*** After overnight incubation, the 96-well block is centrifuged at 3,600 g to remove any potential aggregates. The supernatants are then subjected to buffer exchange with 1X PBS (Invitrogen, cat. no. 10010). For AARS polypeptides > 10 kDa, the Omega 10K membrane of an AcroPrep 96 filter plate is rinsed with 20 µl 1X PBS and the plate is placed onto a vacuum manifold (>10 in. Hg) until all the liquid passes through. The samples in the refolding buffer are dispensed into each well and the vacuum applied until all the liquid passes through. These steps are repeated until the total sample volume (400 µl) has been processed. AARS polypeptides are recovered by adding 180 µl of 1X PBS pH 7.4 to each well, pipetting up and down 10 times carefully, and then transferred to a clean block. This step is repeated to yield a total volume of 360 µl per well and the block is stored at 4°C. For AARS polypeptides < 10 kDa, the refolded samples are loaded onto an Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-3 membrane (Millipore, cat. no. UFC900308) followed by the addition of 10 ml 1X PBS and centrifugation at 3,600 g for 10-30 minutes until the volume is less than 360 µl. The samples are recovered and 1X PBS is added to a final volume of 360 µl.

**[00526]** In order to remove endotoxins, an AcroPrep Advance filter plate with Mustang Q membrane (Pall, cat. no. 8171) is rinsed with 300 µl of 1X PBS and centrifuged at 1,000 g for 5 minutes to remove the buffer. The AARS polypeptides (360 µl/well) are added to the filter plate and incubated on a shaker for 5-10 minutes. The plate is then centrifuged at 1,000 g for 5-10 minutes and the flow through fractions containing the AARS polypeptides are collected and stored at 4°C.

[00527] For larger scale cultures, after overnight incubation, the refolded samples are centrifuged at 10,000 g for 10 minutes to remove any insoluble aggregates. The supernatant is loaded onto an Amicon Ultra-15 Centrifugal Filter Unit and centrifuged at 3,600 g until the volume is reduced to 250  $\mu$ l. The samples are mixed in 10 ml 1X PBS and centrifuged again at 3,600 g for 10-30 minutes until the volume is about 250  $\mu$ l. Note that the pH of 1X PBS is adjusted to match the pH of the refolding buffer, either pH 6.0 or pH 8.0. This step is repeated one more time, the supernatants are recovered and 1X PBS is added to a final volume of 1.5 ml.

[00528] In order to remove endotoxins, a Sartobind Q 5 strong anion exchanger membrane (Sartorius, cat. no. Q5F) is flushed with 1 ml 1X PBS and the AARS polypeptides are slowly passed through the membrane using a plastic syringe. The flow through fraction containing the AARS polypeptides is collected in a 96-deep well block that is sealed and stored at 4°C.

#### PROTEIN PRODUCTION FOR IN VIVO EXPERIMENTS

[00529] **Bacterial cultures:** 100 ng of expression vector comprising DNA encoding each AARS polypeptide (as described in Gene synthesis and cloning method (3)) is transformed into BL21(DE3)-RIPL (Agilent Technologies cat. #230280) competent *E. coli* bacteria at 42 °C for 30 seconds. 500  $\mu$ L of LB medium is added to the cells and incubated for 1 hour at 250 rpm in a 37 °C shaker, and 150  $\mu$ l of the transformation reactions are plated onto ampicillin LB agar plates and incubated overnight at 37°C.

[00530] Individual colonies are picked to start seed cultures in 30ml of LB-Amp and incubated overnight at 250 rpm in a 37 °C shaker. Seed cultures are then used to inoculate 2.5L LB-Amp in 6L Erlenmeyer flasks. After the culture reaches stationary phase (typical OD<sub>600</sub> of 0.6-0.8), the flasks are iced for 30 minutes and then induced with 1M IPTG to a final concentration on 200 $\mu$ M. Individual cultures are then incubated overnight at 250 rpm in a 30 °C shaker

[00531] **Protein Isolation:** The culture is then transferred to 500 ml Nalgene bottles (Cat#3141-0500) and centrifuged at 8,000 x g for 10 minutes at 4°C. The medium is carefully decanted and the pellets are frozen at -20°C.

[00532] The pellets are then thawed and re-suspended in 50 ml Ni-NTA pH 8.0 buffer with 50 $\mu$ l  $\beta$ -ME and one protease inhibitor tablet (Roche #11873580001). 300 mg of lysozyme (Sigma #L6878) is added and the mixture is rotated for 30 minutes at 4°C. The re-suspended pellet is then sonicated at 25, 50, and 75% for 1 minute each (10 seconds on, 5 seconds off). The sample is then spun down at 35,000 x g for 45 minutes at 4°C.

[00533] **Affinity Purification:** The supernatant is then added to 2 ml buffer equilibrated Ni-NTA agarose (Qiagen #30230) and rotated for 1 hour at 4 °C. The nickel bound protein mix is then poured through a buffer equilibrated eco column from Bio-RAD (Cat#737-4151) followed by washing with 1L Ni-NTA buffer pH 8.0 (50 mM Tris, pH 8, 300 mM NaCl, 25mM imidazole) with 0.5% Triton-X114 (Sigma #X114) and then by a 100 ml wash of endotoxin free Ni-NTA buffer pH 8.0. The purified protein is then eluted from the Ni-NTA agarose with 10 ml endotoxin free elution buffer pH 8.0 (50 mM Tris, pH 8, 300 mM NaCl, 300 mM imidazole) and dialyzed overnight in slide-a-lyzers (Pierce) against 1X PBS pH 7.4 (Invitrogen #10010) with two buffer changes an hour apart the next morning.

[00534] **Concentration and Endotoxin removal:** The dialyzed eluates from Ni-NTA are loaded onto an Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-10 membrane (Millipore, cat. # UFC901008) and then centrifuged at 3,600 x g for 10-30 minutes until the desired concentration is reached (usually 1.7 mg/ml).

[00535] In order to remove endotoxin, a Sartobind Q 15 strong anion exchanger membrane (Sartorius, cat. # Q15X) was flushed with 1 mL 1X PBS and the AARS polypeptides are slowly passed through the membrane using a plastic syringe. The flow through fraction containing the AARS polypeptides is collected and aliquoted. AARS polypeptides are then snap frozen in liquid nitrogen and stored at -80°C.

[00536] **Biophysical Characterization:** All purified AARS polypeptides are analyzed by SDS-PAGE, their concentration determined based on  $A_{280}$  and calculated extinction coefficient (ProtParam on ExPASy server). Endotoxin levels are measured by the QCL-1000 Endpoint Chromogenic LAL assay (Lonza, cat. no. 50-648U) according to the manufacturer's instructions.

[00537] **Dynamic Light Scattering:** A Wyatt Technology DynaPro 99 instrument and the temperature controller (20°C) are warmed up for 15 minutes before the experiment followed by connection of the Dynamics software to the instrument. The acquisition time is set to 10 seconds for multiple acquisitions and the laser power is set to 100%. The quartz cuvette is washed thoroughly with deionized water and methanol before the addition of the protein sample (15µL at a concentration of approximately 1 mg/mL in PBS). Air bubbles are removed by tapping the cuvette before it is inserted into the holder with the frosted side to the left. If the intensity is too high (warning message shown on the screen), the sample is further diluted with PBS until the intensity is decreased to a normal range. The data collected include hydrodynamic radius, polydispersity, predicted average molecular weight, percentage of intensity and percentage of mass.

**[00538] *Size Exclusion Chromatography:*** The protein sample is diluted to a concentration of about 5-10 mg/mL in PBS before being loaded into a 100  $\mu$ L sample loop on the General Electric AKTA FPLC. The Superdex 200 10/300 GL size exclusion column (General Electric, cat. no. 17-5175-01) is used for separation. The column is first equilibrated with 1.5 column volume (CV) of 1X PBS buffer, followed by sample injection. The column is run in 1 CV of 1x PBS buffer (isocratic flow) with absorbance at 280nm monitoring. The peak area is integrated and the percentage calculated with the Unicorn software. The elution volume is used to estimate the molecular weight based on comparison with gel filtration calibration kits (General Electric, cat. no. 28-4038-41 and 28-4038-42).

**[00539] *Protein Recovery upon Storage at High Concentration:*** 10  $\mu$ L of the AARS polypeptides concentrated to  $\geq 10$ mg/mL using an Amicon Ultra-15 filter unit (Millipore, cat. no. UFC901024 or UFC900324 depending on molecular weight) are transferred to a clean microcentrifuge tube. The sample is stored at room temperature for one week followed by centrifugation at 16,000 g for 10 minutes to pellet any precipitates. The concentration of the supernatant is determined by a Bradford protein assay and compared to the concentration measured prior to the week-long exposure to room temperature. The recovery is expressed as percentage of the starting concentration.

**[00540] *Characterization of AARS polypeptides by LC-MS:*** Purified AARS polypeptides (1 mg/mL) are diluted 1:10 into 0.1% formic acid and 0.6  $\mu$ g protein is loaded with a Dionex autosampler onto a C4 capillary column. The capillary column is prepared by cutting 150 mm of fused silica tubing (0.36 mm OD by 0.1 mm ID, Polymicro Technologies, cat. no. 2000023). The capillary is pulled at one end with a Suter Instrument Laser Fiber Puller and cut with a fused silica cutter to generate a 5  $\mu$ m tip. The capillary is packed to the length of 75 mm with C4 resin (5 $\mu$ m, 300 $\text{\AA}$ , Michrom, cat. no. PM5/64300/00) using pressure bomb. The LC-MS analysis is performed on an ThermoFisher LTQ ion trap mass spectrometer coupled to a Dionex Ultimate3000 HPLC system. The analyte is eluted from the column using a 35-minute gradient of 5-70% acetonitrile in 0.1% formic acid at a flow rate of 0.9  $\mu$ L/min. The LTQ is operated on a full MS scan mode (300-2,000 m/z) with a spray voltage of 2.5 kV.

**[00541] *Data collection and analysis:*** raw mass spectrometry data are stored in RAW files generated by XCalibur running on the LTQ XL mass spectrometer. The MS spectra of the major peaks on the chromatograph are further analyzed with ThermoFisher deconvoluting algorithm ProMass to obtain the AARS polypeptide molecular weights.

## FUNCTIONAL ANALYSIS OF AARS POLYPEPTIDES

### TRANSCRIPTIONAL PROFILING

[00542] *Background and therapeutic relevance:* In addition to traditional target identification techniques, genomic tools have recently emerged as important approaches to aid in elucidating the mechanism of action of AARS polypeptides and can provide direct insight into therapeutic relevance early in the drug discovery process. To facilitate an understanding of potential therapeutic utility, primary human cell types are cultured with AARS polypeptides and transcriptional profiling is assessed at two separate time points following incubation with AARS polypeptides.

[00543] The cell types chosen for transcriptional profiling are based on the pluripotent capabilities of the cells in question and potential to identify AARS polypeptides of direct therapeutic value. For example, Mesenchymal stem cells (MSCs) can differentiate into osteogenic, adipogenic, chondrogenic, myocardial, or neural lineages when exposed to specific stimuli, making them attractive for understanding the potential relevance of the AARS polypeptides to a broad range of cell types, and diseases.

[00544] In addition to supporting hematopoietic cells, marrow stromal cells can also be induced to differentiate into cells of different connective tissue lineage, such as bone, cartilage, and fat. The potential of Human Mesenchymal stem cells (hMSCs) to maintain multipotency and proliferate extensively in vitro provides new avenues for cell-based therapy in the restoration of damaged or diseased tissue. Recent reports also indicate that HMSCs are capable of cell fate crossing germ layer boundaries. In addition to differentiating into multi-lineages of the mesoderm, these cells can also differentiate into neurons of ectodermal origin and hepatocyte-like cells of endodermal origin. During the process of differentiation, these cells may modify expression patterns of certain lineage specific transcripts.

[00545] Accordingly the ability of specific AARS polypeptides to modulate specific patterns of genes in HMSCs in a time dependent manner demonstrates that these proteins play potentially significant roles in a broad array of differentiation pathways, as well as diseases and disorders resulting from the dysfunction, or deterioration of these processes, or the corresponding cell types. Moreover AARS polypeptides with the ability to modulate gene transcription in MSCs have significant therapeutic utility to enable the in vitro or in vivo modulation of hematopoiesis, neurogenesis, myogenesis, osteogenesis, and adipogenesis, as well as in a broad range of disorders and diseases, including for example inflammatory responses, autoimmunity, cancer, neuronal degeneration, muscular dystrophy, osteoporosis, and lipodystrophy.

[00546] Human Skeletal Muscle Cells (HskMC) can undergo differentiation to exhibit actin and myosin myofilaments, and have been used in the study of genetic muscular diseases such as Malignant Hyperthermia<sup>1</sup>. HskMC also have the potential to act as a cardiac graft, mending damage to the heart. Recently, cultured Human Skeletal Muscle cells have been used in micro gravity experiments to study the effects of low gravity environments on Human Skeletal Muscle.

[00547] Accordingly the ability of specific AARS polypeptides to modulate specific patterns of genes in HskMC in a time dependent manner demonstrates that these proteins play potentially significant roles in the processes of myogenesis, as well as diseases and disorders resulting from the dysfunction, or deterioration of these processes as well as muscle cell development or metabolism. Accordingly AARS polypeptides with the ability to modulate gene transcription in muscle cells have therapeutic utility in a broad range of diseases including for example, the treatment of metabolic disease, cachexia, various muscle wasting conditions, as well as musculoskeletal diseases.

[00548] **Methods:** The ability of AARS polypeptides to modulate gene expression is assessed using a high-throughput microfluidic real-time quantitative PCR (RT-qPCR) approach (Fluidigm Corporation). (See Petriv et al., (2010) PNAS (doi/10.1073/pnas.1009320107) in Human Marrow Stromal Cells (HMSC) and Human Skeletal Muscle Cells (HskMC). In the experiments reported here, Human HskMC (Cat # 150-05f) and HMSC (Cat # 492-05f) were purchased from Cell Applications. HMSC cells are cryopreserved at second passage and can be cultured and propagated to 10 population doublings. Here HMSC in the 6<sup>th</sup> Passage are used. Human Skeletal Muscle Cells (HskMC) are cryopreserved at second passage and can be cultured and propagated for at least 15 population doublings. In the experiments reported here HskMC at passage 6 post harvest from normal human donor are used.

[00549] In both cases, cells are plated at 50000 cells/ mL in 100 $\mu$ L volume of growth media and exposed to AARS polypeptides at a concentration of 250nM, or as otherwise indicated below, for 24 hours and 72 hours. Controls include Differentiation media with a standard cocktail to promote (1) Adipogenesis, (2) Osteogenesis, (3) Chondrogenesis and (4) Skeletal muscle myotube formation. Additional controls include untreated wells containing only growth media. Two wells were run for each Differentiation control. Controls: all media was made utilizing DMEM as the basal media. Standard literature was followed and Differentiation media was purchased from Cell Applications. Per the vendor, differentiation media contained the following additives: Skeletal muscle differentiation cocktail: FBS, insulin, glutamine, FGF, EGF; Adipogenesis cocktail: insulin, dexamethasone and IBMX; Osteogenesis cocktail:



FBS, dexamethasone, ascorbate 2 phosphate, beta-glycerophosphate; Chondrogenesis cocktail: insulin, ascorbate-2-phosphate, and TGF- $\beta$ 1.

[00550] Standard protocols for using an ABI (Applied Biosystems, Item # AM1728) TAQMAN® Gene Expression Cells-to-CT™ Kit are utilized to lyse cells and harvest genomic material. An ABI Pre-Amp Mix (Applied Biosystems, Item# 4391128) is used to initiate pre-amplification. Gene specific primers are created using a Primer 3 program and purchased from IDT technologies. Fluidigm profiling arrays (Item # BMK-M-96.96) were used for actual quantitative PCR with standard Fluidigm loading reagents and pipetting devices. **Table E1** below lists the genes profiled.

<b>Table E1</b>			
<b>List of genes assessed in transcriptional profiling</b>			
<b>Compiled UniqueList</b>	<b>refseq_nt</b>	<b>Full name</b>	<b>Synonyms</b>
ABCA1	NM_005502	ATP-binding cassette, sub-family A (ABC1), member 1	ABC-1 ABC1 CERP FLJ14958 HDLDT1 MGC164864 MGC165011 TGD
ACTB	NM_001101	actin, beta	PS1TP5BP1
ACTG1	NM_001614	actin, gamma 1	ACT ACTG DFNA20 DFNA26
ACVR2B	NM_001106	activin A receptor, type IIB	ACTRIIB ActR-IIB MGC116908
APOA1	NM_000039	apolipoprotein A-I	MGC117399
ARNT	NM_178427	aryl hydrocarbon receptor nuclear translocator	HIF-1beta HIF1B HIF1BETA TANGO bHLHe2
BAD	NM_032989	BCL2-associated agonist of cell death	BBC2 BCL2L8
BCL2	NM_000657	B-cell CLL/lymphoma 2	Bcl-2
BMP2	NM_001200	bone morphogenetic protein 2	BMP2A
BMP4	NM_130851	bone morphogenetic protein 4	BMP2B BMP2B1 MCOPS6 OFC11 ZYME
C3AR1	NM_004054	complement component 3a receptor 1	AZ3B C3AR HNFAG09
CASP3	NM_032991	caspase 3, apoptosis-related cysteine peptidase	CPP32 CPP32B SCA-1
CAV1	NM_001753	caveolin 1, caveolae protein, 22kDa	BSCL3 CGL3 MSTP085 VIP21
CDH5	NM_001795	cadherin 5, type 2 (vascular endothelium)	7B4 CD144 FLJ17376
CFLAR	NM_003879	CASP8 and FADD-like apoptosis regulator	CASH CASP8AP1 CLARP Casper FLAME FLAME-1 FLAME1 FLIP I-FLICE MRIT c-FLIP c-FLIPL c-FLIPR c-FLIPS
COMP	NM_000095	cartilage oligomeric matrix protein	EDM1 EPD1 MED MGC131819 MGC149768 PSACH THBS5
CSF1	NM_172212	colony stimulating factor 1 (macrophage)	MCSF MGC31930
CTGF	NM_001901	connective tissue growth factor	CCN2 HCS24 IGFBP8 MGC102839 NOV2
CTNNB1	NM_001904	catenin (cadherin-associated protein), beta	CTNNB DKFZp686D02253 FLJ25606 FLJ37923

**Table E1**  
**List of genes assessed in transcriptional profiling**

Compiled UniqueList	refseq_nt	Full_name	Synonyms
		1, 88kDa	
DAAM1	NM_014992	dishevelled associated activator of morphogenesis 1	FLJ41657 KIAA0666
ELN	NM_001081755	elastin	FLJ38671 FLJ43523 SVAS WBS WS
ENO1	NM_001428	enolase 1, (alpha)	ENO1L1 MPB1 NNE PPH
FABP3	NM_004102	fatty acid binding protein 3, muscle and heart (mammary-derived growth inhibitor)	FABP11 H-FABP MDGI O-FABP
FAK	NM_001199649	focal adhesion kinase	fak1
FGF4	NM_002007	fibroblast growth factor 4	HBGF-4 HST HST-1 HSTF1 K-FGF KFGF
FIGF	NM_004469	c-fos induced growth factor (vascular endothelial growth factor D)	VEGF-D VEGFD
FLT1	NM_002019	fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)	FLT VEGFR1
FOXA1	NM_004496	forkhead box A1	HNF3A MGC33105 TCF3A
GAPDH	NM_002046	glyceraldehyde-3-phosphate dehydrogenase	G3PD GAPD MGC88685
GFAP	NM_002055	glial fibrillary acidic protein	FLJ45472
SLC2A4	NM_001042	solute carrier family 2 (facilitated glucose transporter), member 4	GLUT4
HAND1	NM_004821	heart and neural crest derivatives expressed 1	Hxt Thing1 bHLHa27 eHand
HIF1A	NM_181054	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)	HIF-1alpha HIF1 HIF1-ALPHA MOP1 PASD8 bHLHe78
HK2	NM_000189	hexokinase 2	DKFZp686M1669 HKII HXK2
HMGB1	NM_002128	high-mobility group box 1	DKFZp686A04236 HMG1 HMG3 SBP-1
HNF4A	NM_178850	hepatocyte nuclear factor 4, alpha	FLJ39654 HNF4 HNF4a7 HNF4a8 HNF4a9 HNF4alpha MODY MODY1 NR2A1 NR2A21 TCF TCF14
HPRT1	NM_000194	hypoxanthine phosphoribosyltransferase 1	HGPRT HPRT
HSPB1	NM_001540	heat shock 27kDa protein 1	CMT2F DKFZp586P1322 HMN2B HS.76067 HSP27 HSP28 Hsp25 SRP27
ICAM1	NM_000201	intercellular adhesion molecule 1	BB2 CD54 P3.58
IFNG	NM_000619	interferon, gamma	IFG IFI

**Table E1**  
**List of genes assessed in transcriptional profiling**

Compiled UniqueList	refseq_nt	Full_name	Synonyms
IGF1	NM_001111285	insulin-like growth factor 1 (somatomedin C)	IGF-I IGF1A IGFI
IGF2	NM_001127598	insulin-like growth factor 2 (somatomedin A)	C11orf43 FLJ22066 FLJ44734 INSIGF pp9974
IGFBP3	NM_001013398	insulin-like growth factor binding protein 3	BP-53 IBP3
IGFBP5	NM_000599	insulin-like growth factor binding protein 5	IBP5
IKBKB	NM_001556	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta	FLJ33771 FLJ36218 FLJ38368 FLJ40509 IKK-beta IKK2 IKKB MGC131801 NFKBIKB
IL10	NM_000572	interleukin 10	CSIF IL-10 IL10A MGC126450 MGC126451 TGIF
IL1B	NM_000576	interleukin 1, beta	IL-1 IL1-BETA IL1F2
IL3	NM_000588	interleukin 3 (colony-stimulating factor, multiple)	IL-3 MCGF MGC79398 MGC79399 MULTI-CSF
IL4	NM_172348	interleukin 4	BCGF-1 BCGF1 BSF1 IL-4 MGC79402
IL5	NM_000879	interleukin 5 (colony-stimulating factor, eosinophil)	EDF IL-5 TRF
IL6R	NM_181359	interleukin 6 receptor	CD126 IL-6R-1 IL-6R-alpha IL6RA MGC104991
IL8	NM_000584	interleukin 8	CXCL8 GCP-1 GCP1 LECT LUCT LYNAP MDNCF MONAP NAF NAP-1 NAP1
ITGA5	NM_002205	integrin, alpha 5 (fibronectin receptor, alpha polypeptide)	CD49e FNRA VLA5A
KDR	NM_002253	kinase insert domain receptor (a type III receptor tyrosine kinase)	CD309 FLK1 VEGFR VEGFR2
LEP	NM_000230	leptin	FLJ94114 OB OBS
LPL	NM_000237	lipoprotein lipase	HDLCQ11 LIPD
MAPK11	NM_002751	mitogen-activated protein kinase 11	P38B P38BETA2 PRKM11 SAPK2 SAPK2B p38-2 p38Beta
MMP1	NM_002421	matrix metalloproteinase 1 (interstitial collagenase)	CLG CLGN
MMP3	NM_002422	matrix metalloproteinase 3 (stromelysin 1, progelatinase)	CHDS6 MGC126102 MGC126103 MGC126104 MMP-3 SL-1 STMY STMY1 STR1
MYH1	NM_005963	myosin, heavy chain 1, skeletal muscle, adult	MGC133384 MYHSA1 MYHa MyHC-2X/D MyHC-2x
MYH11	NM_022844	myosin, heavy chain 11, smooth muscle	AAT4 DKFZp686D10126 DKFZp686D19237 FAA4 FLJ35232 MGC126726 MGC32963 SMHC SMMHC
MYH7	NM_000257	myosin, heavy chain 7, cardiac muscle, beta	CMD1S CMH1 DKFZp451F047 MGC138376 MGC138378 MPD1 MYHCB SPMD SPMM
MYOD1	NM_002478	myogenic differentiation 1	MYF3 MYOD PUM bHLHc1
NFATC1	NM_172390	nuclear factor of	MGC138448 NF-ATC NFAT2 NFATc

**Table E1**  
**List of genes assessed in transcriptional profiling**

Compiled UniqueList	refseq_nt	Full_name	Synonyms
		activated T-cells, cytoplasmic, calcineurin-dependent 1	
NFATC2	NM_173091	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2	NFAT1 NFATP
NFKB1	NM_003998	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	DKFZp686C01211 EBP-1 KBF1 MGC54151 NF-kappa-B NF-kappaB NFKB-p105 NFKB-p50 p105 p50
NOS2	NM_000625	nitric oxide synthase 2, inducible	HEP-NOS INOS NOS NOS2A
NOTCH1	NM_017617	notch 1	TAN1 hN1
NR3C1	NM_001024094	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)	GCCR GCR GR GRL
NRP2	NM_201279	neuropilin 2	MGC126574 NP2 NPN2 PRO2714 VEGF165R2
PAX7	NM_013945	paired box 7	FLJ37460 HUP1 PAX7B RMS2
PDGFB	NM_033016	platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog)	FLJ12858 PDGF2 SIS SSV c-sis
PDK4	NM_002612	pyruvate dehydrogenase kinase, isozyme 4	FLJ40832
PLA2G1B	NM_000928	phospholipase A2, group 1B (pancreas)	MGC119834 MGC119835 PLA2 PLA2A PPLA2
PLIN1	NM_002666	lipid droplet associated protein	perilipin
PPARG	NM_138712	peroxisome proliferator-activated receptor gamma	CIMT1 GLM1 NR1C3 PPARG1 PPARG2 PPARGgamma
QARS	NM_005051	glutaminyl-tRNA synthetase	GLNRS PRO2195
RHOA	NM_001664	ras homolog gene family, member A	ARH12 ARHA RHO12 RHOH12
RUNX1	NM_001754	runt-related transcription factor 1	AML1 AML1-EVI-1 AMLCR1 CBFA2 EVI-1 PEBP2aB
RXRA	NM_002957	retinoid X receptor, alpha	FLJ00280 FLJ00318 FLJ16020 FLJ16733 MGC102720 NR2B1
SERPINE1	NM_001165413	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	PAI PAI-1 PAI1 PLANH1
SMAD2	NM_005901	SMAD family member 2	JV18 JV18-1 MADH2 MADR2 MGC22139 MGC34440 hMAD-2 hSMAD2
SMAD4	NM_005359	SMAD family member 4	DPC4 JIP MADH4
TERT	NM_198255	telomerase reverse transcriptase	EST2 TCS1 TP2 TRT hEST2
TGFB1	NM_000660	transforming growth factor, beta 1	CED DPD1 LAP TGFB TGFBeta

Table E1 List of genes assessed in transcriptional profiling			
Compiled UniqueList	refseq_nt	Full_name	Synonyms
TGFB3	NM_003239	transforming growth factor, beta 3	ARVD FLJ16571 TGF-beta3
THBS4	NM_003248	thrombospondin 4	TSP4
TNF	NM_000594	tumor necrosis factor	DIF TNF-alpha TNFA TNFSF2
TUBB	NM_178014	tubulin, beta	M40 MGC117247 MGC16435 OK/SW-cl.56 TUBB1 TUBB5
TUBB1	NM_030773	tubulin, beta 1	tubulin isoform beta (1)
TUBG1	NM_001070	tubulin, gamma 1	GCP-1 TUBG TUBGCP1
VCAM1	NM_080682	vascular cell adhesion molecule 1	CD106 DKFZp779G2333 INCAM-100 MGC99561
VEGFA	NM_003376	vascular endothelial growth factor A	MGC70609 MVCD1 VEGF VPF
VIM	NM_003380	vimentin	FLJ36605
WISP1	NM_080838	WNT1 inducible signaling pathway protein 1	CCN4 WISP1c WISP1i WISP1tc
WNT1	NM_005430	wingless-type MMTV integration site family, member 1	INT1

**[00551] *Bioinformatics Analysis:*** Data retrieved in .csv format from the Biomark machine by Fluidigm is converted to a tabular format including sample, mRNA, and replicate information along with the raw fluorescence value. PCR reactions that failed are marked as missing. Multiple experiments were combined after normalizing to total expression of mRNA species. All measured mRNA expression is filtered based on the requirement of detection in at least 2 of all of the biological replicates tested. We assessed technical, biological and set deviation mean in entire dataset.

**[00552]** For data analysis Ct values for all genes of interest are first normalized to the averaged Ct values for housekeeping genes from the corresponding sample to obtain  $\Delta\text{Ct}$  values ( $\Delta\text{Ct} = \text{Ct gene} - \text{Ct average housekeeping genes}$ ). Genes from each sample are then normalized to the same gene in untreated control to obtain  $\Delta\Delta\text{Ct}$  values ( $\Delta\Delta\text{Ct} = \Delta\text{Ct control sample} - \Delta\text{Ct treated sample}$ ).

**[00553]** To obtain fold change values up-regulated genes (i.e.  $\Delta\Delta\text{Cts}$  greater than 0) are subject to the following calculation: Fold Change =  $2^{\Delta\Delta\text{Ct}}$ . For down-regulated genes (i.e.  $\Delta\Delta\text{Cts}$  less than 0): Fold Change =  $-(2^{\Delta\Delta\text{Ct}})$ .

#### CELLULAR PROLIFERATION ASSAYS ( ASSAYS A1-A11 IN THE DATA TABLES BELOW)

**[00554] *Background and therapeutic relevance:*** The ability to modulate the rate of cellular proliferation and apoptosis of different cell types represents a fundamental

property of many therapeutic compounds, and is of direct relevance to the treatment and prevention of a broad range of diseases and disorders.

[00555] Accordingly AARS polypeptides with the ability to modulate the rate of cellular proliferation and or apoptosis have significant therapeutic utility in a broad range of diseases including, as growth factors, and differentiation factors for stem cells, and in treatment regimens to enhance or suppress the proliferation of specific cell types of interest *in vivo* or *in vitro*, including for example, haemopoietic cells, immunomodulatory cells, cancer, and for the treatment and prevention of diseases associated with aging, including for example neurodegeneration, peripheral neuropathy, and loss of muscular and soft tissue tone.

[00556] **Methods:** Effects of the AARS polypeptides on cellular proliferation is assessed using one or more of the methods listed below, and as more specifically elaborated in the methods below.

[00557] **Hoechst 33432.** Standard cell counts to assess proliferation are performed using Hoechst 33432, which is a cell-permeant nuclear counterstain that emits blue fluorescence when bound to dsDNA. It is available as a solution (Invitrogen Cat # H-3570) that is used at a final concentration of 1 µg/mL in either media or PBS. Cells are grown in 96 well plates in the presence of AARS polypeptides for a standard growth time of 48 hours, or longer depending on cell type and as described in the examples below.

[00558] **ATP-lite.** Cellular ATP levels correlate with cellular health and can be readily determined using a variety of commercially available kits. ATP-lite (Perkin-Elmer, Cat #6016947 Boston, MA 02481) which is a homogenous mixture of lysis solution and ATP-detection reagent. is pre-mixed before use and is used 1:1 v:v ratio with cultured cells. Plates are incubated for 5 minutes to promote lysis and plates are measured using a luminescent plate reader. Cells are grown in 96 well plates in the presence of AARS polypeptides for a standard growth time of 48 hours, or longer depending on cell type and as described in the examples below.

[00559] **ALAMARBLUE®** (Resazurin) is a cell viability indicator which is based on the redox state of the cells. Resazurin, the active ingredient, is a nontoxic, cell permeable compound that is blue in color and virtually nonfluorescent when present in its oxidized form. However upon entering normal viable cells, resazurin is rapidly reduced to resorufin, which produces a red fluorescence signal. Viable cells continuously convert resazurin to resorufin, thereby generating a quantitative measure of viability—and cytotoxicity. The lack of toxicity allows long-term exposure of cells to resazurin without negative impact; cells grown in the presence of resazurin were

found to produce similar numbers of viable cells as control cells, as determined by flow cytometric analysis.

**[00560] Measurements are made by adding a** solution of Resazurin / ALAMARBLUE® to cells, incubating them for 1-4 hours, and reading the fluorescence or absorbance. The amount of fluorescence or absorbance is proportional to the number of living cells and corresponds to the cells metabolic activity. Damaged and nonviable cells have lower innate metabolic activity and thus generate a proportionally lower signal than healthy cells. After incubation with ALAMARBLUE®, samples can readily be measured on fluorescence and absorbance instrumentation. For fluorescence readings: 530 nm excitation and 590 nm emission filter settings are used.

**[00561]** Cells are grown in 96 well plates in the presence of AARS polypeptides for a standard growth time of 48 hours, or longer depending on cell type and as described in the examples below.

**ACETYLATED LDL UPTAKE IN HEPG2C3A HUMAN HEPATOCYTE CELLS.  
(ASSAY B1 IN THE DATA TABLES BELOW)**

**[00562] *Background and therapeutic relevance:*** LDL is the major carrier of cholesterol in the blood, accounting for more than 60% of the cholesterol in plasma. In humans, the hepatic LDL receptor is responsible for clearing around 70 % of plasma LDL from circulation. Internalized LDL is degraded to free cholesterol and amino acids in the lysosome. The liver is the most important organ for LDL catabolism and LDL receptor activity in humans. LDL that is not internalized and remains in circulation can be transported by endothelial cells into the vessel wall, resulting in the formation of atherosclerotic plaques. Circulating LDL can also be taken up by macrophages and this can also contribute to the formation of plaques. Increasing LDL uptake into hepatic tissue is thought to be beneficial to human health and finding safe and efficacious therapeutics that may the positively regulate this process may provide new therapies for cardiovascular and metabolic diseases. To investigate whether the unique properties of AARS polypeptides can regulate uptake of acetylated LDL, a standard assay for measuring acetylated LDL uptake is employed in HepG2C3a cells.

**[00563]** Accordingly AARS polypeptides with the ability to modulate LDL uptake have significant therapeutic utility in a broad range of diseases including for example, the treatment of hypercholesteremia, hyperlipidemia, type 1 and 2 diabetes, metabolic syndrome, and vascular diseases including atherosclerosis

**[00564] *Methods:*** HEPG2C3a cells (ATCC# CRL-10741) are maintained in Eagles Minimal Essential (EMEM) medium supplemented with 10% FBS (HyClone Cat#SH30910.03), 50u/mL penicillin/50µg/mL streptomycin, (Invitrogen) in 15 mL

medium in 75 mL flasks. Cells are grown at 37°C, 5% CO<sub>2</sub>, in a humidified environment and utilized in BSL2 certified tissue culture hoods using sterile technique and appropriate personal protective equipment including goggles, gloves and lab coats. HEPG2C3a express the LDL-receptor and are competent for acetylated LDL uptake when grown on clear bottom collagen coated plates. A 100 µL volume of cells is plated on collagen coated plates (Invitrogen Cat#A11428) overnight in complete medium (above) at a cell density of 50,000 cells/mL. Cells are washed once with PBS (Invitrogen Cat# 10010) and 80 µL of serum free EMEM is added to each well. AARS polypeptides at a final concentration of 250nM per well are added in a consistent volume in sterile PBS to each well. A unique AARS polypeptide is placed in each well. Cells are serum starved and exposed to the AARS polypeptides for 16 hours. Following the 16 hour incubation, the, supernatant is collected and soluble ICAM is measured using a standard ELISA kit from RND Systems (Cat # DY643; Data Table F), and serum free media supplemented with 5µg/mL ac-LDL (Alexa Fluor 488 labeled Cat # L23380, Invitrogen) is added to each well. Following a 2 hour incubation at 37 °C 5% CO<sub>2</sub>, cells are washed twice with sterile PBS before 100 µL PBS is added to each well for quantification. Plates were analyzed for total fluorescent intensity using a bottom read on a Victor X5 fluorescent plate reader (Perkin Elmer) at an excitation wavelength centered around 485 nm, and an emission wavelength centered around 535 nm. Cells are stained with Hoechst dye and fluorescent intensity 405nm Excitation / 450nm Emission is read to confirm total cell number is consistent across the plate.

**REGULATION OF HUMAN NEUTROPHIL OXIDATIVE BURST AND ELASTASE PRODUCTION (ASSAYS C1-C3 IN THE DATA TABLES BELOW)**

**NEUTROPHIL OXIDATIVE BURST**

**[00565] *Background and therapeutic relevance:*** Phagocytosis by polymorphonuclear neutrophils and monocytes constitutes an essential arm of host defense against infections by microorganisms including bacteria and fungi. The phagocytic process can be separated into several major stages: chemotaxis (migration of phagocytes to inflammatory sites), attachment of particles to the cell surface of phagocytes, ingestion (phagocytosis) and intracellular killing by oxygen-dependent (oxidative burst) and oxygen-independent mechanisms. Reduced or missing burst activity is observed in inborne defects like the chronic granulomatous disease (CGD). CGD is a heterogeneous group of inherited disorders that usually manifests itself during the first two years of life. The disease is characterized by repeated and life-threatening infections caused by bacterial and fungal organisms. These infections typically consist



of pneumonia, lymphadenitis, or abscesses that involve lymph nodes, lungs, and liver. The NADPH oxidase is the enzyme system responsible for producing superoxide anion, which is quickly converted to hydrogen peroxide and hydroxyl radicals. Abnormalities in the constituent peptides of the NADPH oxidase enzyme system lead to the dysfunctions characteristic of CGD. Neutrophils from CGD patients fail to produce a significant oxidative burst following stimulation. Different forms of CGD are described (classical X-linked CGD and autosomal recessive patterns). The oxidative burst of granulocytes is impaired in transplantation, later stages of HIV infection, and in the elderly, making these populations more susceptible to secondary infection and exacerbations of inflammatory disease. Various immunomodulators (e.g., cytokines (GM-CSF, G-CSF, TNF) or drugs) also seem to have effects on the oxidative burst. There is the potential for proteins with the ability to up-regulate or down-regulate oxidative burst in a therapeutic fashion to be useful for a variety of different disease states.

**[00566] *Methods:*** The protein kinase C ligand phorbol 12-myristate 13-acetate (PMA) can be utilized in this assay as an agonist of the oxidative burst process. Heparinized whole blood is mixed with sterile dextran (0.6% final concentration) for 1 hour and allowed to separate into layers. The lower layer contains neutrophil, monocytes and red blood cells. An ammonium chloride lysis step is utilized to remove all RBCs and a 97% pure population of neutrophils with approximately 3% monocyte contamination remains following lysis step. Upon stimulation, granulocytes and monocytes produce reactive oxygen metabolites (superoxide anion, hydrogen peroxide, hypochlorous acid) which destroy bacteria inside the phagosome. Formation of the reactive oxidants during the oxidative burst can be monitored by the addition and oxidation of Amplex Red. The percentage of cells having produced reactive oxygen radicals are then analyzed as well as their mean fluorescence intensity using a fluorescent plate reader. The typical time course for this reaction is 10 minutes, with obvious burst being seen by 2 minutes and a drop off of signal being seen by 20 minutes. This assay can be run in agonist mode in the absence of PMA or in antagonist mode, with concomitant administration of AARS polypeptides and PMA at a concentration that is below the EC50 for this compound.

#### **REGULATION OF HUMAN NEUTROPHIL ELASTASE PRODUCTION**

**[00567] *Background and therapeutic relevance:*** Neutrophil elastase is a serine protease that has been implicated as having a specific role in the development of various human disease states. The wide range of activities make this a versatile enzyme with pleiotrophic effects in diseases of the lung and cardiovascular system. Elastase is

seen as a two edged sword and the range of activity appears to be tissue specific. Although its key physiologic role is in innate host defense, it can also participate in tissue remodeling and possesses secretagogue actions that are now recognized as important to local inflammatory signals. Neutrophil elastase activity has been implicated in the development of emphysema for several decades, however only relatively recently has a pathogenetic function been ascribed to this serine proteinase in situations where excessive extracellular matrix deposition occurs. The use of genetically manipulated animal models is starting to uncover the potential ways in which its actions might influence fibrotic lung repair. Emerging evidence suggests that the engagement of cellular pathways with more direct effects on fibrogenic mediator generation and collagen synthesis appears to underpin the actions of neutrophil elastase in promoting lung matrix accumulation. Human neutrophil elastase is also present within atherosclerotic plaques where it contributes to matrix degradation and weakening of the vessel wall associated with the complications of aneurysm formation and plaque rupture. It is joined by other extracellular proteases in these actions but the broad range of substrates and potency of this enzyme coupled with activity associated with neutrophil degranulation single this disruptive protease out as therapeutic target in atherosclerotic disease.

**[00568] Methods:** This assay uses the ENZCHEK® Elastase Assay Kit (Invitrogen Catalog # E-12056). Neutrophils are prepared from fresh human blood using a 6% dextran solution and red blood cells are lysed before plating cells in RPMI media (media should be un-supplemented with no serum, no antibiotics). A 1.0 mg/mL stock solution of the DQ elastin substrate is prepared by adding 1.0 mL of deionized water (dH<sub>2</sub>O) directly to one of the three vials containing the lyophilized substrate and mixing to dissolve. 1X Reaction Buffer is prepared by diluting 6 mL of the 10X Reaction Buffer in 54 mL dH<sub>2</sub>O. A 100 µg/mL working solution of the DQ elastin substrate is prepared by diluting the DQ elastin stock solution tenfold in 1X Reaction Buffer. Porcine pancreatic elastase stock solution is prepared by making a 100 U/mL stock solution in dH<sub>2</sub>O. To assay for elastase activity, 50 µL of 1X Reaction Buffer is pipette into each assay well containing 500,000 neutrophils/ mL in a 30 µL volume. 8µL of each AARS polypeptide is added per well, and the sample incubated for 20 minutes at 37 °C. 50 µL of 100 µg/mL DQ elastin working solution is added to each well and mixed. Samples are incubated at room temperature, protected from light, for 30 minutes. Fluorescence intensity in a fluorescence microplate reader equipped with standard fluorescein filters (ex 485/ Em 535) fluorescence may be measured over multiple time points.

**BINDING TO TOLL-LIKE RECEPTORS AND ACTIVATION OF NFkB (ASSAYS D1-D4 IN THE DATA TABLES BELOW)**

**[00569] *Background and therapeutic relevance:*** Macrophages are major players in the innate immune system and express a large repertoire of different classes of pattern recognition receptors (PRRs), including the family of Toll-like receptors (TLRs) which are powerful regulators and controllers of the immune response.

**[00570]** Stimulation of TLRs by microbial pathogens and endogenous ligands initiates signaling cascades that induce the secretion of pro-inflammatory cytokines and effector cytokines that direct downstream adaptive immune responses. Endogenous ligands, as well as microbial components, are recognized by and can activate TLRs, raising the possibility that these receptors may be critical targets for the development of new therapies for multiple diseases.

Accordingly AARS polypeptides that modulate TLR receptor activity, have therapeutic utility in a broad range of diseases and disorders including for example, inflammatory diseases and disorders, autoimmune diseases, tissue transplantation / organ rejection, cancer prevention or treatment, the modulation of haematopoiesis and infection.

***Measurement of TLR activation in RAW-BLUE cells***

**[00571]** Mouse macrophages sold under the trademark RAW-BLUE™ cells (Invivogen, Catalog code: raw-sp) express all TLRs except TLR5 and include a secreted embryonic alkaline phosphatase (SEAP) gene which is inducible by NF-kB and AP-1 transcription factors. Upon TLR stimulation, RAW-BLUE™ cells activate NF-kB and/or AP-1 leading to the secretion of SEAP which is measurable when using SEAP detection medium.

**[00572] *Methods:*** RAW-BLUE™ cells are washed twice with PBS, trypsinized and resuspended in fresh Growth Medium (Growth Medium: DMEM, 4.5 g/l glucose, 10% heat-inactivated fetal bovine serum (30 minutes at 56°C), 100 mg/mL ZEOCIN™, 2 mM L-glutamine). Cells are plated at a concentration of 50,000 cells/well in a 96 well plate in a total volume of 100 µL, and AARS polypeptides, controls, or AARS polypeptides (+LPS) are added to each well at the concentrations shown in the experiments outlined below. Cells are incubated at 37°C in a 5% CO<sub>2</sub> incubator for 18 hours. On experimental day 2, SEAP detection medium (QUANTI-BLUE™) (Invivogen Catalog code: rep-qb1) is prepared following the instructions and 120 µL is added per well to a clear flat-bottom 96-well plate, and cell supernatant is added (20 µL). Samples are incubated at 37°C for about 30 minutes to up to 2 hours. SEAP levels are determined using a spectrophotometer and reading absorbance at 650 nM.

[00573] To detect AARS polypeptides that specifically block TLR activation this assay can be modified to identify potential TLR antagonists. In this case AARS polypeptides are added to the cells at a final concentration of about 250nM per well, (or as otherwise specified in the Examples below) 1 hour prior to adding 50 ng/mL LPS. Cells are incubated and SEAP detected as described above. PBS control wells with no LPS or AARS polypeptide alone added are used to find the basal level of TLR stimulation at the time of the measurement. Control wells are pretreated with PBS and known TLR agonists and antagonists. The ratio of the background subtracted [PBS plus LPS signal] to [AARS polypeptide plus LPS signal] is used to determine percent antagonism.

#### ***Human TLR screening in Hek293 cells***

[00574] Human HEK293 cells are genetically modified and sold under the trademark HEK-Blue™ TLR cells (Invivogen). The TLR2 and TLR4 versions of this cell type selectively express all TLR2 or TLR4 and include a secreted embryonic alkaline phosphatase (SEAP) reporter gene under the control of an IFN-beta minimal promoter which is fused to five NF-kB and AP-1 transcription factors binding sites. With the use of specific TLR 2 or 4 agonists (respectively), Hek-Blue™ TLR2 and Hek-Blue™ TLR4 cells activate NF-kB and/or AP-1 leading to the secretion of SEAP which is measurable when using SEAP detection reagent. The Hek-Blue™ TLR2 cells are co-transfected with the LPS co-receptor protein CD14 to enhance TLR2 responsiveness and improve signal quality. The parent cell expresses endogenous levels of TLR1, 3, 5, 6 and also NOD1.

[00575] **Methods:** HEK-Blue™-TLR2 or HEK-Blue™-TLR4 cells are washed twice with PBS, trypsinized and resuspended in fresh Growth Medium (Growth Medium: DMEM, 4.5 g/L glucose, 10% heat-inactivated fetal bovine serum (30 minutes at 56°C), 100 mg/mL ZEOCIN™, 2 mM L-glutamine). Cells are plated at a concentration of 50,000 cells/well in a 96 well plate in a total volume of 100 µL, and AARS polypeptides, controls, or AARS polypeptides (+LPS) are added to each well at the concentrations shown in the experiments outlined below. Cells are incubated at 37°C in a 5% CO<sub>2</sub> incubator for 18 hours. On experimental day 2, SEAP detection medium (QUANTI-BLUE™) (Invivogen Catalog code: rep-qb1) is prepared following the instructions and 120 µL is added per well to a clear flat-bottom 96-well plate, and cell supernatant is added (20 µL). Samples are incubated at 37°C for about 30 minutes to up to 2 hours. SEAP levels are determined using a spectrophotometer and reading absorbance at 650 nM. Control wells are pretreated with PBS and known TLR agonists such as UltraPure LPS (TLR-4) or PAM3CSK4 (TLR-2). The ratio of the background

subtracted [PBS plus LPS signal] to [AARS polypeptide plus LPS signal] is used to determine percent agonism.

#### **CYTOKINE RELEASE (ASSAYS E1-E16 IN THE DATA TABLES BELOW)**

**[00576] *Background and therapeutic relevance:*** Cytokines are a diverse set of small cell signaling protein molecules that are used extensively for intercellular communication, and play significant roles in normal body homeostasis, including immunomodulation and regulation. Accordingly AARS polypeptides that modulate the release, or biological activities of cytokines, have therapeutic utility in a broad range of diseases and disorders including for example, inflammatory diseases and disorders, autoimmune diseases, tissue transplantation / organ rejection, cancer prevention or treatment, the modulation of haematopoiesis and infection.

##### ***Cytokine release from cells in culture***

**[00577] *Methods:*** Test cells are seeded into a 24-well plate at density of about 1 million cells/well in 1 mL of growth media. Cells are treated with either AARS polypeptide (at the concentrations shown in the examples below) or an equal volume of PBS and incubated overnight at 37° with 5% CO<sub>2</sub>. Following cell treatment, samples are centrifuged at 4 °C in a swinging bucket centrifuge at 2,000 x g for 5 minutes. Media is carefully removed so as to not disturb the cell pellet and transferred to a new tube. Samples are assayed immediately or snap frozen in liquid nitrogen for subsequent analysis. Cytokine release (including the cytokines MIF, IL-8, IL-10, Serpin E1, GM-CSF, GRO, IL-1 alpha, IL-1beta, IL-1ra, IL-6, MCP-1, MIP-1, RANTES and TNF-alpha) is determined using commercially available kits (R&D Systems, Inc, MN, USA) or via a contract research organization (MD Biosciences (St. Paul, MN)).

##### ***Cytokine Release from Human Whole Blood***

**[00578] *Methods:*** Human whole blood is obtained from normal human donors and collected with heparin in standard collection tubes. Blood is used on the same day as it is collected to ensure adequate cell health. Blood is mixed gently and plated in an 100 µL volume into 96 well polycarbonate V bottom plates. AARS polypeptides are added and slowly mixed into blood 2X using a multichannel pipet set on 50 µL. Filter tips are used for all experimentation and full PPE is worn. All experimentation occurs in a dedicated biosafety hood that is suitable for experimentation with human blood. Blood is incubated overnight at 37° C with 5% CO<sub>2</sub>. Following cell treatment, samples are centrifuged in a swinging bucket centrifuge at 2,000 x g for 5 minutes. Supernatant is collected for cytokine ELISAs ELISA are performed as described previously.

### *Cytokine Release from PBMCs*

[00579] **Methods:** To isolate peripheral blood mononuclear cells freshly isolated human whole blood is gently layered over Sigma HISTOPAQUE®-1077 at a ratio of 1:1 in 50 mL conical tubes at room temperature. Layered samples are centrifuged at 400 x g in a swinging bucket clinical centrifuge for 30 minutes at room temperature with no brake. The white cellular layer at the interface between the plasma and density gradient is then removed by pipet. These peripheral blood mononuclear cells are washed twice with RPMI-1640 (Invitrogen #22400-105) by dilution and centrifugation for 10 minutes at 250 x g. The washed PBMC were resuspended in RPMI-1640 + 10% FBS and plated at  $1 \times 10^6$  cells/mL.

### *Cytokine release from Human Synoviocytes*

[00580] **Background and therapeutic relevance:** A large number of studies have demonstrated that IL-6 and IL-8 are overproduced in several diseases, and thus may play a fundamental role in the pathogenesis of inflammatory disease. IL-6 activates endothelial cell production, leading to the release of IL-8 and monocyte chemoattractant protein, expression of adhesion molecules, and recruitment of leukocytes to inflammatory sites. These cytokines are expressed in cell types associated with inflammatory disease, including cells involved in the pathogenesis of systemic juvenile arthritis, systemic lupus erythematosus, Crohn's disease, and rheumatoid arthritis. One of the most important systemic actions of cytokine production is the induction of the acute phase response. Acute phase proteins are produced primarily by the liver and include proteins that promote the immune response through activation of complement, induction of proinflammatory cytokines, and stimulation of neutrophil chemotaxis. Alternatively, the acute phase response can be helpful, and acute-phase proteins, such as proteinase antagonists, opsonins, and procoagulants, help limit tissue destruction by resolving inflammation. In particular, IL-6 can stimulate synoviocyte proliferation and osteoclast activation, leading to synovial pannus formation and repair. IL-6 acts with IL-1 to increase production of matrix metalloproteinases, which may contribute to joint and cartilage destruction. However, IL-6 may also have protective effects in the joint, as suggested by the finding that this cytokine induces the expression of the tissue inhibitor of metalloproteinase and stimulates proteoglycan synthesis when injected into the joints of mice with antigen-induced arthritis. Human Fibroblast-Like Synoviocytes-Rheumatoid Arthritis (HFLS-RA) are isolated from synovial tissues obtained from patients with Rheumatoid Arthritis (RA). They are cryopreserved at second passage and can be cultured and propagated at least 5 population doublings. HFLS are long known

for their role in joint destruction by producing cytokines and metalloproteinases that contribute to cartilage degradation.

[00581] Accordingly AARS polypeptides with the ability to modulate the growth, differentiation, or cytokine release profile of fibroblast-like synoviocytes-rheumatoid arthritis (HFLS-RA) have therapeutic utility in a broad range of diseases including for example, the treatment of inflammatory diseases and disorders including systemic juvenile arthritis, systemic lupus erythematosus, Crohn's disease, and rheumatoid arthritis.

[00582] **Methods:** HFLS-RA, adult cells (Cell Applications Cat # 408RA-05a) are maintained in Synoviocyte Growth Medium (Cell Applications Cat #415-50) in 15 mL medium in 125 mL flasks for 1 passage before use. Cells are maintained at 37°C, 5% CO<sub>2</sub>, in a humidified environment and utilized in BSL2 certified tissue culture hoods using sterile technique and appropriate personal protective equipment including goggles, gloves and lab coats. An 80 µL volume of cells is plated overnight in growth medium at a cell density of about 50,000 cells/mL. AARS polypeptides at a final concentration of 250 nM per well (or as otherwise indicated in the examples below) are added in sterile PBS to each well following overnight adherence. Control wells contain untreated cells and are incubated with an equivalent volume of PBS. Cells are exposed to proteins or PBS in basal media (Cell Applications Cat #310-470) for 24 hours. Supernatant is removed and IL-8, IL-6 and TNFα ELISA assays are run according to manufacturer's instructions (RND Systems, Cat # DY206 and DY-208, DY-210 Duo-set kits). Proliferation is assessed with Resazurin as described previously by adding fresh media containing Resazurin to plates following supernatant removal and incubating for three hours at 37 °C. Plates are read on a fluorescent plate reader and viability / proliferation is expressed as a function of resorufin associated fluorescence of AARS polypeptide treated wells divided by resorufin associated fluorescence of PBS only treated wells.

#### ***Human Astrocyte Proliferation and inflammatory cytokine production***

[00583] **Background and therapeutic relevance:** Human astrocytes (HA) are derived from human cerebral cortex. They are cryopreserved at second passage and can be cultured and propagated 10 population doublings. HA are the most abundant cells in the central nervous system and they perform many functions such as provision of mechanical support and nutrients to neurons, and removal of wastes from neurons. In addition to playing a critical support role for optimal neuronal functioning, they also provide biochemical support of endothelial cells which form the blood-brain barrier. Recent studies have shown that astrocytes are capable of regulating neurogenesis by

instructing the stem cells to adopt a neuronal fate and controlling the function of single synapses, participate actively in the transfer and storage of information in the brain. Recognition of the importance of astrocytes in nervous system functioning is increasing. HA can serve as useful in vitro model for exploring the diversity of astrocytes functions. Astrocytes have been shown to proliferate in response to IL6 and TNFalpha. In addition, these cells are capable of making their own IL6 and TNFalpha. Thus AARS polypeptides which modulate the proliferation and cytokine production in HA have therapeutic utility in a variety of neurological diseases including neuro-inflammation, neurodegeneration, tumorigenesis of the brain, and brain ischemia and repair.

**[00584] Methods:** Human Astrocytes (HA) from Cell Applications (Cat # 882K-05f) are maintained in Cell Applications HA Cell Growth Medium (Cat # 821-500) according to manufacturer's instructions. Cells are maintained at 37°C, 5% CO<sub>2</sub>, in a humidified environment and utilized in BSL2 certified tissue culture hoods using sterile technique and appropriate personal protective equipment including goggles, gloves and lab coats. An 80 µL volume of cells is plated on collagen coated plates overnight in complete medium (above) at a cell density of 50,000 cells/mL. Cells are washed once with PBS and 80 µL of serum free growth media is added to each well. AARS polypeptides at a final concentration of 250 nM per well (or as otherwise described in the examples below) are added in a consistent volume in sterile PBS to each well. Cells are exposed to AARS polypeptides for 48 hours and spent media is removed for cytokine assessment (as described previously). Cells are exposed to proteins or PBS in basal media (Cell Applications Cat #310-470) for 48 hours. Supernatant is removed and IL-8 and IL-6 ELISA assays are run according to manufacturer's instructions (RND Systems, Cat # DY206 and DY-208, DY-210 Duo-set kits). Proliferation is assessed with Resazurin as described previously by adding fresh media containing Resazurin to plates following supernatant removal and incubating for three hours at 37 °C. Plates are read on a fluorescent plate reader and viability / proliferation is expressed as a function of resorufin associated fluorescence of AARS polypeptide treated wells divided by resorufin associated fluorescence of PBS only treated wells.

***Human lung microvascular endothelial cell (HLMVEC) proliferation and inflammatory cytokine production.***

**[00585] Background and therapeutic relevance:** The pulmonary vasculature is of great physiological/pathological significance. It is now recognized to be a tissue composed of metabolically active, functionally responsive cells, that interact with circulating substrates and formed elements in ways that regulate the composition of



systemic arterial blood, affect target organ functions, and contribute to thrombosis, hemostasis and immune reactions, as well as tumor metastasis. Human lung microvascular endothelial cells (HLMVEC) exhibit elevated expression of chemoattractant cytokines and cell adhesion molecules that provide critical cues for directed migration of leucocytes into the lung during acute lung injury. This primary cell type can be useful tool for studying various aspects of pathology and biology of the pulmonary microvasculature in vitro. Alteration in the structure and function of the microvasculature in response to inflammatory stimuli is believed to be a key factor in organ damage and under appropriate conditions, may provide a stimulus for repair. A significant cause of these vascular alterations is the induction of an inflammatory reaction involving leukocyte infiltration. A variety of studies focused on granulocyte adhesion to the endothelium have revealed that leukocyte recruitment and emigration involves a well-orchestrated adhesion cascade. The adhesion cascade begins when the granulocyte attaches to the endothelium and begins to roll in the direction of fluid flow at a low velocity. As the granulocyte rolls, it becomes activated, subsequently firmly adheres to the endothelium, and migrates across the endothelium into the extravascular space. These adhesion events are mediated, in part, by molecular interactions that occur between CAMs on the surface of the granulocytes and cognate glycoproteins present on the endothelium. A variety of studies have revealed that the endothelial cell adhesion molecule E-selectin can interact with SLe<sup>x</sup>-type glycan presenting granulocyte ligands to mediate the attachment and rolling steps of the adhesion cascade. The downstream steps of the cascade involve the interaction of endothelial-expressed intercellular adhesion molecule with granulocyte-expressed CD18 integrins.

[00586] Thus AARS polypeptides which modulate proliferation and / or cytokine production of human lung microvascular endothelial cells have therapeutic utility in a variety of vascular and pulmonary diseases including inflammatory and obstructive lung diseases including for example, pulmonary hypertension, chronic obstructive pulmonary disease, idiopathic pulmonary fibrosis, and asthma.

[00587] **Methods:** HLMVEC (Cell Applications, Catalog # 540-05) are maintained in Cell Applications Microvascular Endothelial Cell Growth Medium (Cat # 111-500), For appropriate growth, an Attachment Factor Solution containing collagen (Cell Applications, Catalog # 123-100), is used to coat plates and flasks before plating cells. Cells are maintained at 37°C, 5% CO<sub>2</sub>, in a humidified environment and utilized in BSL2 certified tissue culture hoods using sterile technique and appropriate personal protective equipment including goggles, gloves and lab coats. A 80µL volume of cells is plated on collagen coated plates overnight in complete medium (above) at a cell density of 50,000 cells/mL. Cells are washed once with PBS and 80 µL of serum free

growth media is added to each well. AARS polypeptides at a final concentration of 250 nM per well (or as otherwise described in the examples below) are added in a consistent volume in sterile PBS to each well. Cells are exposed to AARS polypeptides for 48 hours and spent media is removed for ELISA for cell adhesion molecules and cytokine assessment (as described previously). Cell adhesion molecules including soluble VCAM and/or ICAM are measured using a standard ELISA kit from RND Systems (Cat # DY643 and DY720 respectively). Proliferation is assessed with Resazurin as described previously by adding fresh media containing Resazurin to plates following supernatant removal and incubating for three hours at 37 °C. Plates are read on a fluorescent plate reader and viability / proliferation is expressed as a function of resorufin associated fluorescence of AARS polypeptide treated wells divided by resorufin associated fluorescence of PBS only treated wells.

#### **CELL ADHESION ((ASSAYS F1-F7 IN THE DATA TABLES BELOW))**

**[00588] *Background and therapeutic relevance:*** Cell Adhesion Molecules (CAMs) are proteins located on the cell surface which are involved with the binding with other cells or with the extracellular matrix (ECM) in the process called cell adhesion. These proteins are typically transmembrane receptors and are composed of three domains: an intracellular domain that interacts with the cytoskeleton, a transmembrane domain, and an extracellular domain that interacts either with other CAMs of the same kind (homophilic binding) or with other CAMs or the extracellular matrix (heterophilic binding). Most of the CAMs belong to four protein families: Ig (immunoglobulin) superfamily (IgSF CAMs), the integrins, the cadherins, and the selectins. The immunoglobulin superfamily (IgSF) cell adhesion molecules are calcium-independent transmembrane glycoproteins, including: neural cell adhesion molecules (NCAMs), intercellular cell adhesion molecules (ICAMs), vascular cell adhesion molecule (VCAM), platelet-endothelial cell adhesion molecule (PECAM-1), endothelial cell-selective adhesion molecule (ESAM), junctional adhesion molecule (JAMs), nectins, and other cell adhesion molecules.

**[00589]** Cell adhesion molecules are cell surface glycoproteins that are critical for leukocyte adhesion to the sinusoidal endothelium and transmigration and cytotoxicity in a variety of inflammatory liver diseases. ICAM-1 plays an important role in inflammation, and the increased expression of ICAM-1 on endothelial cells is reflected in the activation of endothelial cells. ICAM-1 is of particular importance since it mediates firm endothelial adhesion and facilitates leukocyte transmigration. Studies have shown that there is an upregulation of ICAM-1 on both sinusoidal cells and

hepatocytes in inflammatory liver conditions such as hepatitis B viral infection, autoimmune liver disorders, alcoholic hepatitis, and liver allograft rejection.

[00590] Thus AARS polypeptides which modulate cell adhesion molecule production and cell adhesion to endothelial cells have therapeutic utility in a variety of inflammatory diseases including for example, cardiovascular diseases, atherosclerosis, autoimmunity and pulmonary hypertension.

[00591] **Methods:** Human umbilical vein cells (ATCC, Cat # CRL-2873 ) (HUVEC) are seeded at a concentration of about  $1.2 \times 10^5$  cells / well in 12 well plates coated with human fibronectin attachment solution in the suggested ATCC media and supplements and grown according to manufacturer's instructions. Cells are stimulated with AARS polypeptides at the indicated concentrations, or PBS alone, and incubated overnight in growth media. Human acute monocytic leukemia (THP-1 (TIB-202)), cells are resuspended into 0.1% BSA/ RPMI serum free medium with calcein AM (6  $\mu$ L/mL; Invitrogen Cat # C1430) and incubated for 30 minutes. Labeled cells are collected and resuspended in RPMI medium containing 10 % FBS, and the density adjusted to  $2 \times 10^6$  cells/mL.

[00592] 100 $\mu$ L ( $2 \times 10^5$ ) labeled THP-1 cells are placed into each well of the HUVEC monolayer in 1 mL of growth media and incubated for 15 minutes. The wells are washed twice with PBS to remove unbound cells, and then the cells are read by fluorescent plate reader with an Excitation wavelength of 488 nm and an Emission wavelength of 530 nm.

#### CELLULAR DIFFERENTIATION (ASSAYS G1-G4 IN THE DATA TABLES BELOW)

##### *Adipocyte differentiation and proliferation in primary human pre-adipocyte cells.*

[00593] **Background and therapeutic relevance:** Both obesity and lipodystrophy are commonly associated with pathologies including diabetes and cardiovascular diseases. It is now recognized that adipose tissue is an endocrine organ that secretes a wide variety of factors, and dysregulated secretion affects adipogenesis as well as whole-body glucose/insulin homeostasis. Excess adipose tissue leading to obesity has become a severe public health threat. Adipose tissue development can be affected by genetic background, hormonal balance, diet, and physical activity. Adipose tissue mass can increase when fat cells are increased in size due to higher triacylglycerol accumulation. In addition, an increase in fat cell number, arising from differentiation of precursor cells into adipocytes, can also occur even in adults as observed in severe human obesity and in rodents fed a high-carbohydrate or high-fat diet. Adipocytes

specifically are thought to arise from mesenchymal cells that undergo the commitment and differentiation process, adipogenesis. Pre-adipocyte cell lines can undergo adipocyte differentiation upon treatment with adipogenic agents comprised of synthetic glucocorticoid, dexamethasone (DEX), isobutylmethylxanthine (IBMX), and insulin, have been valuable in these studies. Peroxisome proliferator-activated receptor  $\gamma$  (PPAR $\gamma$ ) and CCAAT enhancer-binding protein (C/EBP) family of transcription factors have been firmly established to play critical roles in adipocyte differentiation. Early during adipocyte differentiation, C/EBP $\beta$  and C/EBP $\delta$  are induced by DEX and IBMX, respectively, which together then induce PPAR $\gamma$  and C/EBP $\alpha$  to activate various adipocyte markers that are required for adipocyte function. Other transcription factors have also been reported to either positive or negatively regulate adipogenesis and various growth factors and hormones can affect adipocyte differentiation by regulating expression of adipogenic transcription factors. In fact, in addition to being the main site for energy storage in mammals by storing triacylglycerol and releasing fatty acids in times of need, adipose tissue secretes a wide array of molecules that are involved in diverse physiological processes including immune response, vascular function, and energy homeostasis. Cytokines such as TNF- $\alpha$  and IL-6 are secreted from adipocytes. Some of these factors may also affect growth and development of adipose tissue by autocrine/paracrine action.

[00594] Thus AARS polypeptides which have the ability to modulate the differentiation and / or proliferation of normal human pre-adipocytes have therapeutic utility in a broad range of diseases including for example, the treatment and prevention of metabolic disease, cardiovascular diseases, obesity and lipodystrophies, as well as the long term complications of diabetes.

[00595] **Methods:** HPAd (human pre-adipocytes) (Cell Application Cat # 803sD) are maintained according to vendor instructions. For culturing, cells are thawed quickly, and transferred immediately into 15mL of Adipocyte Growth Medium (Cell Application Cat # 811M-250) and plated into a standard sterile tissue culture treated flask. Media is replaced with fresh Adipocyte Growth Medium every other day until cell is >60% confluent. Cells are grown at 37°C, 5% CO<sub>2</sub> in a humidified environment and utilized in BSL2 certified tissue culture hoods using sterile technique and appropriate personal protective equipment including goggles, gloves and lab coats. Cells are plated in clear bottom black walled 96 well tissue culture treated assay plates for differentiation at a concentration of about 50,000 cells/mL. AARS polypeptides at a final concentration of 250nM per well (or as otherwise indicated in the Examples below) are added to each assay well. All cells are maintained in growth media for 2 days with the exception of the positive controls which are stimulated with adipogenic

differentiation media (Cell Applications Cat #811D-250). Cells are exposed to AARS polypeptides for 48 hours. Cell adhesion molecules including soluble VCAM and/or ICAM are measured using a standard ELISA kit from RND Systems (Cat # DY643 and DY720 respectively; Data Table F). Proliferation is assessed with Resazurin as described previously by adding fresh media containing Resazurin to plates following supernatant removal and incubating for three hours at 37 °C (Data Table A). Plates are read on a fluorescent plate reader and viability / proliferation is expressed as a function of resorufin associated fluorescence of AARS polypeptide treated wells divided by resorufin associated fluorescence of PBS only treated wells. Fresh media is added and differentiation is maintained for 16 days post initial media exchange, with fresh media exchanged every other day to maintain cell health. On day 15, cells are placed in serum free media. On day 16, differentiation to mature adipocytes is assessed with Nile Red (Invitrogen, concentration of 3  $\mu$ M final) staining and quantified with a fluorescent plate reader with the appropriate wavelengths. To perform this assay cells are fixed with 10% paraformaldehyde, washed in PBS and permeabilized in PBS containing 0.5% BSA and 0.1% Triton X-100. Cell proliferation is assessed with an intensity measurement on a fluorescent reader with Hoechst dye 33432 at a concentration of 1 $\mu$ g/mL final, as described previously. Adipogenesis is expressed as intensity of Nile Red signal. Hoechst dye signal is used to assess cellular number.

***Human skeletal muscle cell differentiation and proliferation.***

[00596] **Background and therapeutic relevance:** The development of skeletal muscle is a multistep process that involves the determination of pluripotential mesodermal cells to give rise to myoblasts, withdrawal of the myoblasts from the cell cycle and differentiation into muscle cells, and finally growth and maturation of skeletal muscle fibers. Skeletal muscle differentiation involves myoblast alignment, elongation, and fusion into multinucleate myotubes, together with the induction of regulatory and structural muscle-specific genes. At the molecular level, myogenic commitment and muscle-specific gene expression involve the skeletal muscle-specific helix-loop-helix (bHLH) MyoD family of proteins, which includes MyoD, myogenin, myf-5, and MRF4, and the myocyte enhancer-binding factor 2 (MEF2). The DNA binding activity of MyoD family proteins is attenuated by Id, which forms complexes with E2a gene products in proliferating cells and is down-regulated when they are induced to differentiate. The decision to differentiate into myotubes is influenced negatively by several factors. Treatment of myoblasts with fetal bovine serum, basic fibroblast growth factor 2, or transforming growth factor  $\beta$ 1 is known to inhibit differentiation of myoblasts. Myogenesis is also regulated negatively by oncogenes such as c-myc, c-jun,

c-fos, H-ras, and E1a. There is very little information regarding the signaling that is triggered in the myoblast upon serum withdrawal which leads to the induction of the MyoD family gene expression and to muscle differentiation. Myogenic differentiation appears to depend on the activation of integrins present on the plasma membrane of myoblasts suggesting the operation of an “outside-in” biochemical pathway in which integrin is the upstream molecular species. Interactions of insulin-like growth factor (IGF)-I and -II with their receptors are also positive regulators of skeletal muscle differentiation.

[00597] Accordingly AARS polypeptides with the ability to modulate muscle development have therapeutic utility in a broad range of diseases including for example, the treatment of metabolic disease, cachexia, various muscle wasting conditions, as well as musculoskeletal disease where muscle atrophy plays a key role in the pathogenesis and symptomology. Human Skeletal Muscle Cells (HSkMC) can undergo differentiation to exhibit actin and myosin myofilaments. HSkMC have been used in the study of genetic muscular diseases such as Malignant Hyperthermia. HSkMC also have the potential to act as a cardiac graft, mending damage to the heart, and thus AARS polypeptides with the ability to modulate muscle development also have utility as in vitro and in vivo regulators of myogenesis.

[00598] **Methods:** To assess the potential role of AARS polypeptides in this process, a standard assay of skeletal muscle cell differentiation was employed. For this assay, Human Adult Skeletal Muscle Cells (HSkMC, Cell Application Cat # 150-05f) are isolated from healthy human donors from limbal skeletal muscle. Cells are maintained in HSkMC Growth Medium (Cell Applications, Cat # 151-500). These cells can be cultured and propagated for at least 15 population doublings. For differentiation, cells are maintained in growth media for one passage and then plated at 50,000 cells per mL media in to 96 well clear bottom black walled TC treated plates treated with collagen at 100  $\mu$ L per well. Cells are allowed to adhere overnight. AARS polypeptides in PBS, or PBS alone, is added to each well at a final concentration of 250nM protein (or as otherwise indicated in the examples below). Control wells received the same volume of Differentiation Media (Cell Applications Cat # 151D-250) at this time. Cells are incubated with protein or differentiation media for 48 hours. At 48 hours, cell culture supernatant is collected from all wells and differentiation media is added at a volume of 150  $\mu$ L to the entire plate with the exception of control wells which are maintained in growth media only. Supernatant is utilized to assess cytokine production including IL6 and IL8 as described previously (Data Table E). Proliferation is assessed with Resazurin as described previously by adding fresh media containing Resazurin to plates following supernatant removal and incubating for three hours at 37 °C (Data Table A).

Cells are monitored under the microscope and media is exchanged for fresh Differentiation media every 2 days. On Day 10, media is removed and cells are fixed with 10% paraformaldehyde for 30 minutes. Cells are permeabilized with 0.1% Triton X-100 in PBS for 15 minutes and cells are stained with TR-Labeled phalloidin and Hoechst 33432 (as described previously) to define actin and nuclei respectively. Nuclear intensity is used to determine cell proliferation in each well and phalloidin intensity is used to determine total actin content. Cells are also stained with alpha actin skeletal muscle antibody (GenTex Cat # GTX101362). Digital photos using a fluorescent microscope as well as visual inspections and scoring are made of all wells.

***Human bone marrow mesenchymal stem differentiation and proliferation.***

**[00599] Background and therapeutic relevance:** Mesenchymal stem cells (MSCs) are multipotent stem cells that can differentiate into a variety of cell types, including osteoblasts, chondrocytes, myocytes, adipocytes, beta-pancreatic islets cells, and potentially, neuronal cells. Many different events contribute to the commitment of the MSC to other lineages including the coordination of a complex network of transcription factors, cofactors and signaling intermediates from numerous pathways. MSCs are of intense therapeutic interest because they represent a population of cells with the potential treat a wide range of acute and degenerative diseases.

**[00600]** Moreover AARS polypeptides with the ability to modulate the differentiation of MSCs into different developmental pathways have significant therapeutic utility to enable the in vitro or in vivo modulation of hematopoiesis, neurogenesis, myogenesis, osteogenesis, and adipogenesis, as well as in a broad range of disorders and diseases, including for example inflammatory responses, autoimmunity, cancer, neuronal degeneration, muscular dystrophy, osteoporosis, and lipodystrophy. Human MSCs are immuno-privileged, and represent an advantageous cell type for allogenic transplantation, reducing the risks of rejection and complications of transplantation. Recently, there have also been significant advances in the use of autologous mesenchymal stem cells to regenerate human tissues, including cartilage and meniscus, tendons, and bone fractures. Many studies have also investigated the use of MSCs for gene therapy, including transplantation of MSCs transfected with vascular endothelial growth factor for the improvement of heart function after MI in rats, MSCs as vehicles for interferon- $\beta$  delivery into tumors in mice and gene therapy with MSCs expressing BMPs to promote bone formation. Accordingly due to the intense interest as MSCs as direct and modified therapeutics, as well as the potential of AARS polypeptides to act as therapeutic agents to regulate the differentiation of MSCs in vivo,

AARS polypeptides were tested as potential inducers of MSC proliferation and differentiation.

**[00601] Methods:** hMSC (human marrow stromal cells) (Cell Application Cat # 492-05f) are maintained according to vendor instructions. For culturing, cells are thawed quickly, and transferred immediately into 15mL of Marrow Stromal cell Growth Medium (Cell Application Cat # 419-500) and plated into a standard sterile tissue culture treated flask. Media is replaced with fresh Marrow Stromal cell Growth Medium every other day until cells are >60% confluent. Cells are grown at 37°C, 5% CO<sub>2</sub>, in a humidified environment and utilized in BSL2 certified tissue culture hoods using sterile technique and appropriate personal protective equipment including goggles, gloves and lab coats. Cells are plated in clear bottom black walled 96 well tissue culture treated assay plates for differentiation at a concentration of 50,000 cells/mL. tRNA synthetase derived proteins at a final concentration of 250 nM per well (or as otherwise specified in the Examples below) are added to each assay well. All cells are maintained in growth media for 2 days with the exception of the positive controls, which was stimulated with osteogenic or chondrogenic differentiation media (StemPro, Invitrogen, Cat # A10072-01 and A10071-01 respectively). Cells are exposed to AARS polypeptides for 48 hours. Soluble VCAM is measured using a standard ELISA kit from RND Systems (Cat # DY643; Data Table F). Proliferation is assessed with Resazurin as described previously by adding fresh media containing Resazurin to plates following supernatant removal and incubating for three hours at 37 °C (Data Table A). Plates are read on a fluorescent plate reader and viability / proliferation is expressed as a function of resorufin associated fluorescence of AARS polypeptide treated wells divided by resorufin associated fluorescence of PBS only treated wells. Following an assessment of cell viability, resazurin is removed with two media exchanges and 0.5X differentiation media is added to all wells. Differentiation is monitored by visual inspections of all wells for 10 days post media exchange, with fresh media exchanged every other day to maintain cell health. Differentiation was assessed with alkaline phosphatase staining using ELF-97 stain (Invitrogen Cat# E6601) at day 10 post first differentiation exchange. (Yang et al, Nature Protocols (6) 187–213 (2011) doi:10.1038/nprot.2010.189).

***Human pulmonary artery smooth muscle cell (hPASMC) proliferation and differentiation.***

**[00602] Background and therapeutic relevance:** Pulmonary artery smooth muscle cells (PASMCs) in normal human adult lung blood vessels are mostly quiescent, non-migratory and are largely committed to executing their contractile function in the lung.



However, PASMCs are not terminally differentiated and possess the ability to modulate their phenotype and exit their quiescent state in response to changing local environmental cues. This differentiation state may occur in development, tissue injury, and vessel remodeling in response to changes in tissue demand. Pulmonary hypertension (PH) is associated with a variety of underlying conditions including an increase in peripheral pulmonary vascular resistance as a result of increased vascular tone and PASMC contractility and vascular remodeling. Vascular remodeling involves PASMC growth, synthesis of matrix material, and alterations in cell-cell and cell-matrix interactions in the walls of small pulmonary arteries (PAs), which lead to increased thickness of the smooth muscle component of the vessel wall and abnormal muscularization of the normally nonmuscularized, distal PAs. This process contributes to reduced lumen diameter and increased peripheral resistance. Although the precise role of the PASMCs in the initial cause of the disease is controversial, the changes that occur play a key role in the clinical consequences of the disease. A crucial step in studying cellular differentiation is identifying a set of cell-specific or cell-selective genes that contribute to the differentiated function(s) of the cell. A variety of smooth muscle cell (SMC) genes have been identified that serve as useful markers of the relative state of differentiation or maturation of the vascular SMCs, such as SM alpha-actin, SM MHC, h1-calponin, SM22-alpha, desmin, metavinculin, smoothelin and others. The most widely used marker is SM alpha-actin, partially because of the commercial availability of a number of very high-affinity and highly selective antibodies for this protein. Whether changes in PASMCs result from their inherent characteristics or from dysregulation of molecular events that govern PASMC growth remains an open question. However determining the regulatory cues and managing potential dis-regulation provides significant therapeutic insight to managing a variety of vascular and pulmonary diseases including pulmonary hypertension, vascular diseases.

[00603] Thus AARS polypeptides which have the ability to modulate the differentiation and / or proliferation of normal human PASMCs derived from adult humans have therapeutic utility in a variety of vascular and pulmonary diseases including inflammatory and obstructive lung diseases including for example, pulmonary hypertension, chronic obstructive pulmonary disease, idiopathic pulmonary fibrosis, and asthma.

[00604] **Methods:** HPASMC (Cell Applications Cat # 352-05a) are maintained in HPASMC growth media (Cell Applications Cat # 352-05a) in 15 mL medium in 125 mL flasks for 1 passage before use. Cells are maintained at 37°C, 5% CO<sub>2</sub>, in a humidified environment and utilized in BSL2 certified tissue culture hoods using sterile technique and appropriate personal protective equipment including goggles, gloves and

lab coats. An 80 $\mu$ L volume of cells is plated on collagen coated overnight in growth medium at a cell density of 50,000 cells/mL. AARS polypeptides were added in sterile PBS to each well at a final concentration of 250 nM (or as otherwise specified in the Examples below). Control wells held only an equivalent volume of PBS. Positive control samples were incubated with vendor supplied HPASMC differentiation media (Cell Applications Cat # 311D-250). Cells are exposed to AARS polypeptides or PBS in basal media (Cell Applications Cat # 310-470) for 48 hours followed by a media exchange to differentiation media for the entire plate. Supernatant is collected and utilized to assess cytokine production including IL6 and IL8 as described previously (Data Table E). Proliferation is assessed with Resazurin as described previously by adding fresh media containing Resazurin to plates following supernatant removal and incubating for three hours at 37 °C (Data Table A). Cells are monitored for 10 days with a media exchange every other day. Differentiation is assessed after fixation as described above, and permeabilization with 0.1% Triton X-100, by quantifying staining to smooth muscle actin- $\alpha$  staining using an anti-SMA- $\alpha$  antibody (GeneTex Cat #GTX101362) and an Alexa 405 conjugated secondary antibody. Proliferation is assessed with Hoechst staining after cell fixation in 10% formaldehyde for 30 minutes. Hoechst dye is read using a bottom reading fluorescent plate reader with an excitation wavelength (Ex) of 405 nm, and an emission wavelength (Em) of 450 nm. Total actin staining is assessed via the use of an Alexa-488 labeled phalloidin stain (Invitrogen Cat# A12379).

**ANALYSIS OF THE BINDING OF AARS POLYPEPTIDES TO CELLS (ASSAYS H1-H10 IN THE DATA TABLES BELOW)**

**[00605] *Background and therapeutic relevance:*** The binding of AARS polypeptides to specific cell types demonstrates that the cell type in question expresses specific receptors for the AARS polypeptide in question. Depending upon the cell type in question, cell binding implies a potential role for the AARS polypeptide in regulating the activity or behavior of the cell, or similar types of cell, in vivo. Specific examples of such regulatory roles include for example, the binding and modulation of B-cells and T-cells (immunomodulation / chemotaxis / autoimmunity / inflammation); HepG2 cells (control of metabolism, cholesterol uptake or metabolism); THP-1, jurkat, Raji cells (immunomodulation / chemotaxis / autoimmunity / inflammation), platelets (thrombopoiesis), 3T3L1 adipocytes (lipogenesis / metabolism), and C2C12 mouse myoblasts (myogenesis, osteogenesis).

***Binding to blood cells***

[00606] **Methods:** Blood is collected in EDTA tubes from healthy donors. 2mL whole blood is placed into 5mL Falcon FACS tube. 2mL of staining buffer (PBS + 2% FBS) is added, vortexed 3-5 seconds, centrifuged for 5 minutes at 300 x g. The supernatant aspirated, the wash repeated, and the pellet resuspended in 2 mL of staining buffer.

[00607] 100µl of washed blood is transferred to clean 5mL FACS sample tubes. His6- or V5-His6-tagged AARS polypeptides are added to tubes at the concentrations indicated in the specific experiments outlined below and incubated on ice for 45 minutes. After incubation, antibodies for the different cell type surface markers (BD Pharmigen Cat Nos. 560910, 555398, 555415, 340953, 560361), and FITC labeled anti-V5 tag antibody (V5-FITC, Invitrogen Cat # R96325) or FITC labeled anti-His6 antibody (AbCam Cat #ab1206) are added to tubes, incubated in the dark on ice 30 minutes. After incubation 2mL of BD FACS Lysing Solution (cat #349202) was added to tubes. Samples are vortexed, and placed on ice for 15 minutes. Samples are washed with 1 x 2mL PBS and resuspended in 2mL of 2% formaldehyde in PBS prior to FACS analysis. AARS polypeptides that bind greater than 25% of a cellular population, where antibody alone has no significant signal, is deemed a hit.

[00608] **Platelet binding assays:** 50µL of washed blood is transferred to clean 5mL FACS sample tubes, His6- or V5-His6-tagged AARS polypeptides are added to tubes at the concentrations indicated in the specific experiments outlined below and tubes are placed on ice for 45 minutes. 20 µL CD61 pan platelet antibody (BD Pharmigen, Cat # 555754) and 0.5 µL anti- V5-FITC labeled antibody (Invitrogen, R96325) or FITC labeled anti-His6 antibody (AbCam Cat #ab1206) are added to each tube. Tubes are placed on ice and protected from light for 30 minutes. Samples are brought up to a total volume in 2mL of 1% formaldehyde in PBS and analyzed by flow cytometry within 24 hours. AARS polypeptides that bind greater than 25% of a cellular population, where antibody alone has no significant signal, is deemed a hit.

[00609] **Binding to cells in culture:** Approximately  $1 \times 10^6$  cells in 100 µL complete RPMI medium are placed into 5mL FACS tubes. His6- or V5-His6-tagged AARS polypeptides are added to tubes at the concentrations indicated in the specific experiments outlined below and tubes are placed on ice for 45 minutes. Cell samples are washed twice with 1mL staining buffer (PBS + 2% FBS), and then 0.5µL of anti-V5-FITC antibody (Invitrogen R96325) or FITC labeled anti-His6 antibody (AbCam Cat #ab1206) in staining buffer with 200µg/mL human IgG, is added and the samples incubated on ice, protected from light, for 30 minutes. Samples are washed twice with 1mL staining buffer, and then brought up to a total volume in 2mL of 1% formaldehyde

in PBS and analyzed by flow cytometry within 24 hours. AARS polypeptides that bind greater than 25% of a cellular population, where antibody alone has no significant signal, is deemed a hit.

#### **ANIMAL STUDIES: MODULATION OF HAEMATOPOIESIS AND CIRCULATING CYTOKINES**

**[00610] *Background and therapeutic relevance:*** Hematopoiesis (alternatively haemopoiesis or hemopoiesis) is the formation of blood cellular components. All cellular blood components are derived from haematopoietic stem cells (HSCs) which reside in the medulla of the bone (bone marrow) and have the unique ability to give rise to all of the different mature blood cell types. HSCs are self renewing: when they proliferate, at least some of their daughter cells remain as HSCs, so the pool of stem cells does not become depleted. The other daughters of HSCs (myeloid and lymphoid progenitor cells), however can each commit to any of the alternative differentiation pathways that lead to the production of one or more specific types of blood cells, but cannot themselves self-renew. A change in the blood components in response to exposure to an AARS polypeptide therefore suggests that the AARS polypeptide is capable of modulating hematopoiesis, and regulating the development of haematopoietic stem cells.

**[00611]** All blood cells can be divided into three lineages; Erythroid cells, lymphocytes and myelocytes.

**[00612]** Erythroid cells are the oxygen carrying red blood cells. Both reticulocytes and erythrocytes are functional and are released into the blood. Accordingly a reticulocyte count estimates the rate of erythropoiesis, and a change in red blood cell count suggests that an AARS polypeptide modulates erythropoiesis.

**[00613]** Lymphocytes are the cornerstone of the adaptive immune system. They are derived from common lymphoid progenitors. The lymphoid lineage is primarily composed of T-cells and B-cells (types of white blood cells). Accordingly a change in white blood cell count or composition in response to exposure to an AARS polypeptide suggests that that the AARS polypeptide modulates lymphopoiesis.

**[00614]** Myelocytes, which include granulocytes, megakaryocytes and macrophages, and are derived from common myeloid progenitors, are involved in a variety of roles, including innate immunity, adaptive immunity, and blood clotting. Accordingly a change in myeloid cell count or composition in response to exposure to an AARS polypeptide suggests that that the AARS polypeptide modulates myelopoiesis. The same rationale can be used to establish whether the AARS polypeptides modulate granulopoiesis, by measuring changes in granulocyte number in response to exposure to the AARS polypeptides. A role for the AARS polypeptide in modulating

megakaryocytopoiesis may be inferred by a change in megakaryocyte or platelet composition or number in the blood.

[00615] Cytokine release in either wild type mice, or in various animal model systems of inflammation, provides an initial assessment of the potential ability of the AARS polypeptides to modulate inflammatory responses. The role of AARS polypeptides in modulating acute chronic inflammatory processes for example, can be readily assessed using a mouse model of diet induced obesity (DIO). The DIO model centers upon placing rodents on a high fat diet for several months leading to increased obesity, insulin resistance and immune system dysfunction. A particular consequence of this immune system dysregulation results in increased production of proinflammatory cytokines in DIO animals leading to a condition of chronic systemic inflammation. There is a growing body of evidence suggesting that low grade inflammation contributes to the development and maintenance of obesity and a diabetic phenotype that is similarly observed in the human condition termed metabolic syndrome. As such, the ability of AARS polypeptides to modulate the immune system and restore homeostatic balance towards a resolution of this chronic inflammatory state would be particularly beneficial in numerous diseases and disorders including but not limited to the treatment and prevention of the symptoms and side effects of metabolic disease, diabetes, cardiovascular diseases, atherosclerosis, obesity, as well as various autoimmune diseases and disorders, including for example, multiple sclerosis, vascular and allergic disorders.

[00616] **Methods:** Male wild type control (C57BL/6) or diet induced obesity mice (C57BL/6NHsd) are purchased from Harlan (Indianapolis, IN) and housed individually. DIO mice are fed a high fat diet (Cat. #TD.06414-60% kcal from fat) and control mice are fed a normal diet (Cat. #2018S-18% kcal from fat). DIO mice are placed on the high fat diet starting at 6 weeks of age for a total of 10 weeks. Both DIO and control mice are allowed to feed and drink *ad libitum*. At 16 weeks of age, mice are sorted and randomized into groups of 5 animals based on weight. On day 2, mice are weighed and tail vein bled (100 $\mu$ L) for pre-treatment complete blood count (CBC) analysis. On day 1, mice are weighed and intravenously injected via the tail vein with vehicle (PBS) or individual AARS polypeptides at 10mg/kg. Four hours post-injection, mice are facial vein bled (150-200 $\mu$ L) for subsequent cytokine analysis. On days 2, 3, & 4, mice are intravenously dosed as on day 1. On day 5, mice are weighed, terminated and blood are collected by heart puncture for Complete Blood Count (CBC analysis) (plasma-EDTA) and cytokine examination (serum).

[00617] **CBC and cytokine analysis:** Complete blood counts are analyzed from blood draws preceding injections (day -2) and 24 hours after the final injection (day 5).

CBC values are assessed for total white blood cell counts and overall red blood cell morphology. White blood cells are further characterized by total and fractional percentage of neutrophils, lymphocytes, monocytes, eosinophils, & basophils. Red blood cell breakdown included measurements of hemoglobin (dL), hematocrit (%), mean corpuscular volume (fL), mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration (%), and total platelet count ( $10^3/\mu\text{L}$ ). CBC analysis is performed by Antech Diagnostics (Fishers, IN).

**[00618]** Circulating cytokine levels are examined at 4 hours post-injection (day 1) and 24 hours after the final injection (day 5). Serum is isolated, snap frozen and sent to Rules Based Medicine (Austin, TX) for multi-analyte profiling. Serum samples are analyzed using the RodentMap panel encompassing 59 unique biomarkers including Apo A-1, CD40, CD40-L, CRP, ET-1, eotaxin, EGF, Factor VII, fibrinogen, FGF-9, FGF-basic, GST- $\alpha$ , GCP-2, GM-CSF, KC/GRO $\alpha$ , haptoglobin, IgA, IFN $\gamma$ , IP-10, IL-1 $\alpha$ , IL-1 $\beta$ , IL-10, IL-11, IL-12p70, IL-17A, IL-18, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, LIF, lymphotactin, M-CSF-1, MIP-1 $\alpha$ , MIP-1 $\beta$ , MIP-1 $\gamma$ , MIP-2, MIP-3 $\beta$ , MDC, MMP-9, MCP-1, MCP-3, MCP-5, MPO, myoglobin, SAP, SGOT, SCF, RANTES, TPO, tissue factor, TIMP-1, TNF- $\alpha$ , VCAM-1, VEGF-A, and vWF. A change in cytokine levels was counted as a hit if the cytokine increased by at least 2-fold or decreased by at least 50% compared to vehicle controls.

#### EXAMPLE 1

##### IDENTIFICATION OF PROTEOLYTIC FRAGMENTS AND PRODUCTS OF ALTERNATIVE SPLICING FROM AARS USING PROTEIN TOPOGRAPHY AND MIGRATION ANALYSIS PLATFORM

**[00619]** To identify AARS fragments from cell lines, conditioned media and tissues, samples are prepared in the following ways:

**[00620]** *Mouse macrophage (RAW 264.7), cytosol and conditioned media:* Cells are treated with serum free DMEM media at a density of  $15 \times 10^6$  cells / flasks. After 48 hours conditioned media and cell pellets are collected and processed. 200  $\mu\text{g}$  of protein from secreted and cytosolic proteomic fractions are separated by SDS-PAGE and gel slices are prepared for analysis by mass spectrometry.

**[00621]** *Mouse pancreas tissue:* The pancreas from three mice are chopped, dounce homogenized, and sonicated in PBS with protease inhibitors. Cytosolic proteome is isolated by centrifugation and 200  $\mu\text{g}$  of protein is separated by SDS-PAGE and gel slices are prepared for analysis by mass spectrometry.

**[00622]** *Mouse liver tissue:* Three mouse livers are chopped, dounced homogenized, and sonicated in PBS with protease inhibitors. Cytosolic proteome is isolated by

centrifugation and 200  $\mu$ g of protein is separated by SDS-PAGE and gel slices are prepared for analysis by mass spectrometry.

**[00623]** In-gel digests are analyzed by LTQ XL ion trap mass spectrometer (ThermoFisher) equipped with ultimate 3000  $\mu$ LC system (Dionex). The samples are first loaded on PepTrap (michrom) for 10 min with 5% Acetonitrile in 0.1% formic acid using Dionex autosampler. Then the samples are analyzed with a 100 $\mu$ m (inner diameter) fused silica capillary column containing 10 cm of C18 resin (michrom). Peptides are eluted from the column into mass spectrometer with a flow rate of 0.45 $\mu$ l/min using a linear gradient of 5-33.5% acetonitrile in 0.1% formic acid within 110 min.

**[00624]** LTQ is operated in data-dependent scanning mode such that one full MS scan is followed by seven MS/MS scans of the seven most abundant ions. Dynamic exclusion is enabled with repeat count equals to 1, repeat duration equals to 20 seconds, exclusion list size is 300 and exclusion duration is 60 seconds.

**[00625]** After LC-MS/MS analysis, the raw data is searched with BioWorks3.3.1(SEQUEST) using a concatenated target/decoy variant of the mouse IPI database. The SEQUEST data are filtered and sorted with DTASelect. **Tables 1, 4 and 7** show sequences identified in this way.

## EXAMPLE 2

### IDENTIFICATION OF SPLICE VARIANTS USING DEEP SEQUENCING

**[00626]** Splice variants of the aminoacyl tRNA synthetase are identified using high throughput sequencing of cDNA libraries enriched for aminoacyl tRNA synthetase transcripts. The cDNA templates are prepared from total RNA extracts of tissues such as human adult and fetal brains and enriched for aminoacyl tRNA synthetase transcripts by using primer sequences specific for all annotated exons of all annotated human aminoacyl tRNA synthetases and their associated proteins.

**[00627]** Human Total RNAs are obtained from Clontech. For cell line and mouse tissue samples, total RNAs are extracted using RNA Extract II Kit (MN). Genomic DNA is digested in the total RNA samples by DNAase I. To obtain mature messenger RNAs (mRNAs), the RNA samples are enriched twice by binding polyA<sup>+</sup> RNA and digestion of RNA without 5'-cap by 5'-phosphate dependent exonuclease. Complementary DNA (cDNA) is synthesized from mature RNAs using primers that anneal to exon sequences of aminoacyl tRNA synthetase genes. A transcriptome enriched for aminoacyl tRNA synthetase genes is amplified by multiplex PCR using the aminoacyl tRNA synthetase-exon specific cDNA and different combinations of aminoacyl tRNA synthetase-exon primers. The double-stranded aminoacyl tRNA

synthetase-enriched transcriptome PCR products are enzymatically repaired at both ends before adding A-overhangs to the 3' ends of the repaired fragments. Sequencing adaptors and index sequences are then added to the aminoacyl tRNA synthetase-enriched transcriptome PCR products to generate cDNA libraries for deep sequencing with Illumina's Multiplex Sequencing Kit. In brief, the aminoacyl tRNA synthetase-enriched transcriptome PCR products with 3'-A overhangs are ligated to the InPE adaptor oligonucleotides provided in the kits. Index sequences are added to the PCR products with InPE adaptors. To obtain enough DNA fragments for deep sequencing, the PCR products with index sequences are further amplified by PCR. Aminoacyl tRNA synthetase-enriched cDNA libraries with different indexes are pooled and sequenced using an Illumina DNA sequencing machine to get 50 base pair end reads. Sequencing reads are mapped to human or mouse genome for identification of alternative splicing events. "Splicemap" software (available for public download at <http://www-stat.stanford.edu/~kinfai/SpliceMap/>) is used to identify splice junctions.

[00628] Deep sequencing of these cDNAs are performed to generate about 1 million sequencing reads of about 50 nucleotides in length. The sequences specific for exons of the aminoacyl tRNA synthetases are queried against annotated exon junctions and new exon junctions are identified as alternative splicing events.

[00629] The columns in **Tables 2, 5, and 8** labeled "5' exon" and "3'exon" indicate, when present, which exons are fused together in the cDNA sequence. Tables 2, 5, and 8 show sequences that were identified for alternative splice events, transcripts containing such splice events, and the polypeptides expressed by those transcripts. Alternative splice variants identified by deep sequencing are identified in **Tables 2, 5, and 8** as those ones in which there are numbers greater than zero in the columns labeled as "Sequencing reads" in the human adult or fetal brain.

### EXAMPLE 3

#### IDENTIFICATION OF AARS POLYPEPTIDES USING BIOINFORMATICS

[00630] AARS protein fragments (resectin or appendacrine peptides) are identified using bioinformatics. Amino acid sequences of the full length human aminoacyl tRNA synthetase are aligned with the full length amino acid sequence of its ortholog from the bacterium *Escherichia coli* using a program such as FASTA (available at the website [http://fasta.bioch.virginia.edu/fasta\\_www2/fasta\\_www.cgi](http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi)) or the BLASTP program from the NCBI (available at the website [http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PROGRAM=blastp&BLAST\\_PROGRAMS=blastp&PAGE\\_TYPE=BlastSearch&SHOW\\_DEFAULTS=on&LINK\\_LOC=blasthome](http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome)). Resectin sequences from the human proteins are identified as sequences covering



regions where there are gaps in the bacterial sequence in the alignment, or regions with low homology between the two species. The peptide, and corresponding DNA sequences in **Tables 3, 6, and 9** include examples identified in this way.

#### **EXAMPLE 4**

##### **DIFFERENTIAL EXPRESSION OF AARS POLYPEPTIDES IDENTIFIED BY MASS SPECTROMETRY**

[00631] The PROTOMAP technique is used as described in Example 1 to compare the differential expression of Alanyl tRNA synthetases in different tissues/cell types (refer to **Tables 1, 4, and 7**) for sequences and comparisons): Aminoacyl-tRNA synthetase resectin expression is compared between mouse liver tissue and mouse pancreas tissue. Aminoacyl-tRNA synthetase resectin expression is compared between cytosol of RAW264.7 and conditioned media from RAW264.7 cells harvested after 48 hours of serum starvation.

#### **EXAMPLE 5**

##### **DIFFERENTIAL EXPRESSION OF AARS POLYPEPTIDES IDENTIFIED BY DEEP SEQUENCING**

[00632] To test for differential expression of splice events, the deep sequencing is done for cDNAs prepared from different tissues. The number of sequencing hits in the deep sequencing samples reflects relative abundance of a sequence in the samples. The columns labeled “Hs Adult” and Hs Fetal” in **Tables 2B, 5B and 8B** indicate how many times the sequence is identified in the random sequencing of the enriched cDNA from human adult brain and from human fetal brain respectively.

[00633] Expression of specific alternative splice events for aminoacyl tRNA synthetases is unexpected and indicates biological importance. The variation in relative number of reads seen in the deep sequencing of adult and fetal human brain transcriptome samples indicates that alternative splice events of aminoacyl tRNA synthetases are differentially regulated and not just artifacts due to sample handling. Additional evidence for this unexpected biological relevance is that alternative splice variants are not detected in the deep sequencing analysis of transcriptomes from adult or fetal brain for all aminoacyl tRNA synthetases (*e.g.*, cysteinyl-tRNA synthetase and threonyl-tRNA synthetase).

### EXAMPLE 6

#### ANTIBODY SCREENING

[00634] To facilitate the discovery of antibodies displaying preferential binding to specific aminoacyl tRNA synthetase fragments (*e.g.*,  $\geq 10$ -fold higher affinity when compared to the parental full length enzyme), a human antibody phage display library is screened by AbD Serotec (a division of MorphoSys™, Martinsried/Planegg, Germany) using affinity enrichment techniques (panning). Antibodies enriched after multiple rounds of screening with the aminoacyl tRNA synthetase fragments are subsequently characterized by ELISA for reactivity to the fragments, and to the parental, full length enzyme. Clones demonstrating preferential binding (*e.g.*,  $\geq 10$ -fold higher affinity) to the aminoacyl tRNA synthetase fragments are further characterized.

[00635] If the necessary specificity is not achieved at the end of this process, subtraction strategies, such as pre-adsorption steps with the full length enzyme and/or counter-screening, are used to eliminate cross reacting antibodies and drive the selection process towards the unique epitope(s) on the aminoacyl tRNA synthetase fragments.

### EXAMPLE 7

#### IDENTIFICATION OF SPLICE VARIANTS USING SYSTEMATIC PCR

[00636] cDNA templates for PCR reactions are reverse transcribed from total RNA extracts of tissues or cells (*e.g.*, human brain, IMR-32 and HEK293T). PCR reactions are performed using aminoacyl tRNA synthetase specific primers, pairing a forward primer (FP1) designed to anneal to the 5' untranslated region or exons in the 5' half of the gene with a reverse primer (RP1) designed to anneal to exons in the 3' half of the gene or the 3'UTR. Amplified DNA products are analyzed by agarose gel electrophoresis to identify PCR products that are a different size than the fragment amplified from the canonical transcripts. These different PCR products are excised and purified from the gel and ligated into a standard cloning vector for DNA sequence analysis. Alternative splicing variants are identified as different sequences from the canonical transcripts. Splice variants identified by this systematic PCR approach are shown in **Tables 2, 5 and 8** as those variants with zero sequencing reads in both the adult and fetal brains.

## EXAMPLE 8

## CODON OPTIMIZATION OF SELECTED AARS POLYNUCLEOTIDES

[00637] Representative AARS polypeptides (summarized in **Table E2**) are selected for further biochemical, biophysical and functional characterization based on one or more of the following criteria, i) the identification of AARS polypeptide proteolytic fragments, ii) the identification of AARS polypeptide splice variants, iii) the identification of AARS polypeptides by bioinformatic analysis, iv) evidence of differential expression of specific AARS polypeptides, v) the domain structure of the AARS protein, vi) the size of the AARS polypeptide, and vii) the minimization of similar duplicative sequences.

<b>Table E2</b> <b>Summary of AARS Polypeptides Selected for Codon Optimization and Bacterial Expression</b>					
<b>AARS Polypeptide Name</b>	<b>SEQ. ID Nos. for Epitope Tagged AARS polypeptides</b>	<b>SEQ. ID. Nos. for AARS Polynucleotides</b>	<b>Residues of AARS protein</b>	<b>Location of epitope tag</b>	<b>Cloning / synthesis method used</b>
AlaRS1 <sup>N1</sup>	SEQ.ID. NO.61	SEQ.ID. NO.73	1-401	N-terminal	1
AlaRS1 <sup>N1</sup>	SEQ.ID. NO.62	SEQ.ID. NO.73	1-401	C-terminal	1
AlaRS1 <sup>N2</sup>	SEQ.ID. NO.63	SEQ.ID. NO.74	1-286	N-terminal	2
AlaRS1 <sup>N2</sup>	SEQ.ID. NO.64	SEQ.ID. NO.74	1-286	C-terminal	2
AlaRS1 <sup>N3</sup>	SEQ.ID. NO.65	SEQ.ID. NO.75	1-488	N-terminal	1
AlaRS1 <sup>N3</sup>	SEQ.ID. NO.66	SEQ.ID. NO.75	1-488	C-terminal	1
AlaRS1 <sup>N4</sup>	SEQ.ID. NO.67	SEQ.ID. NO.76	1-224 + 51 aa	N-terminal	2
AlaRS1 <sup>N4</sup>	SEQ.ID. NO.68	SEQ.ID. NO.76	1-224 + 51 aa	C-terminal	2
AlaRS1 <sup>N6</sup>	SEQ.ID. NO.69	SEQ.ID. NO.77	1-497 + 24 aa	N-terminal	1
AlaRS1 <sup>N6</sup>	SEQ.ID. NO.70	SEQ.ID. NO.77	1-497 + 24 aa	C-terminal	1
AlaRS1 <sup>N8</sup>	SEQ.ID. NO.71	SEQ.ID. NO.78	1-321 + 6aa	N-terminal	2
AlaRS1 <sup>N8</sup>	SEQ.ID. NO.72	SEQ.ID. NO.78	1-321 + 6aa	C-terminal	2
AlaRS1 <sup>C1</sup>	SEQ.ID. NO. 146	SEQ.ID. NO. 167	476-968	N-terminal	1
AlaRS1 <sup>C1</sup>	SEQ.ID. NO. 147	SEQ.ID. NO. 167	476-968	C-terminal	1
AlaRS1 <sup>C2</sup>	SEQ.ID. NO. 148	SEQ.ID. NO. 168	758-968	N-terminal	2
AlaRS1 <sup>C2</sup>	SEQ.ID. NO. 149	SEQ.ID. NO. 168	758-968	C-terminal	2
AlaRS1 <sup>C3</sup>	SEQ.ID. NO. 150	SEQ.ID. NO. 169	588-968	N-terminal	1
AlaRS1 <sup>C3</sup>	SEQ.ID. NO. 151	SEQ.ID. NO. 169	588-968	C-terminal	1
AlaRS1 <sup>C4</sup>	SEQ.ID. NO. 152	SEQ.ID. NO. 170	659-968	N-terminal	1
AlaRS1 <sup>C4</sup>	SEQ.ID. NO. 153	SEQ.ID. NO. 170	659-968	C-terminal	1
AlaRS1 <sup>C4</sup>	SEQ.ID. NO. 164	SEQ.ID. NO. 176	659	N-terminal	2
AlaRS1 <sup>C4</sup>	SEQ.ID. NO. 165	SEQ.ID. NO. 176	659	C-terminal	2
AlaRS1 <sup>C5</sup>	SEQ.ID. NO. 154	SEQ.ID. NO. 171	691-968	N-terminal	1
AlaRS1 <sup>C5</sup>	SEQ.ID. NO. 155	SEQ.ID. NO. 171	691-968	C-terminal	1
AlaRS1 <sup>C6</sup>	SEQ.ID. NO. 156	SEQ.ID. NO. 172	747-968	N-terminal	2
AlaRS1 <sup>C6</sup>	SEQ.ID. NO. 157	SEQ.ID. NO. 172	747-968	C-terminal	2

<b>Table E2</b> <b>Summary of AARS Polypeptides Selected for Codon Optimization and Bacterial Expression</b>					
<b>AARS Polypeptide Name</b>	<b>SEQ. ID Nos. for Epitope Tagged AARS polypeptides</b>	<b>SEQ. ID. Nos. for AARS Polynucleotides</b>	<b>Residues of AARS protein</b>	<b>Location of epitope tag</b>	<b>Cloning / synthesis method used</b>
AlaRS1 <sup>C6</sup>	SEQ.ID. NO. 166	SEQ.ID. NO. 143	747-968	C-terminal	3
AlaRS1 <sup>C7</sup>	SEQ.ID. NO. 158	SEQ.ID. NO. 173	747-968	N-terminal	1
AlaRS1 <sup>C7</sup>	SEQ.ID. NO. 159	SEQ.ID. NO. 173	747-968	C-terminal	1
AlaRS1 <sup>C11</sup>	SEQ.ID. NO. 160	SEQ.ID. NO. 174	16 aa + 558-968	N-terminal	2
AlaRS1 <sup>C11</sup>	SEQ.ID. NO. 161	SEQ.ID. NO. 174	16 aa + 558-968	C-terminal	2
AlaRS1 <sup>C12</sup>	SEQ.ID. NO. 162	SEQ.ID. NO. 175	2 aa+727-968	N-terminal	1
AlaRS1 <sup>C12</sup>	SEQ.ID. NO. 163	SEQ.ID. NO. 175	2 aa+727-968	C-terminal	1
AlaRS1I1	SEQ.ID. NO. 179	SEQ.ID. NO. 181	476-749	N-terminal	2
AlaRS1I1	SEQ.ID. NO. 180	SEQ.ID. NO. 181	476-749	C-terminal	2

**[00638]** Polynucleotides encoding the selected AARS polypeptides listed in **Table E2**, along with the appropriate N or C-terminal epitope tag, are synthesized and cloned as described in the General Materials and Methods section using the gene synthesis methodology listed in Table E2.

#### EXAMPLE 9

##### SMALL SCALE BACTERIAL EXPRESSION AND PURIFICATION

**[00639]** The AARS polypeptides listed in **Table E2** are expressed in E.coli. as described in the General Materials and Methods section. The relative expression of soluble and inclusion body localized AARS polypeptides is summarized in **Table E3** below.

<b>Table E3</b> <b>Summary of AARS Polypeptide Bacterial Expression Characteristics</b>			
<b>AARS Polypeptide</b>	<b>Location of Epitope Tag</b>	<b>Amount of Protein Recovered from Soluble Fraction</b>	<b>Amount of Protein Recovered from Inclusion Bodies</b>
AlaRS1 <sup>N1</sup>	N-terminal	+	ND
AlaRS1 <sup>N1</sup>	C-terminal	++	ND
AlaRS1 <sup>N2</sup>	N-terminal	+	ND
AlaRS1 <sup>N2</sup>	C-terminal	+	ND
AlaRS1 <sup>N3</sup>	N-terminal	++	ND
AlaRS1 <sup>N3</sup>	C-terminal	++	ND
AlaRS1 <sup>N4</sup>	N-terminal	+	ND
AlaRS1 <sup>N4</sup>	C-terminal	+	ND

Table E3 Summary of AARS Polypeptide Bacterial Expression Characteristics			
AARS Polypeptide	Location of Epitope Tag	Amount of Protein Recovered from Soluble Fraction	Amount of Protein Recovered from Inclusion Bodies
AlaRS1 <sup>N6</sup>	N-terminal	++	ND
AlaRS1 <sup>N6</sup>	C-terminal	+	ND
AlaRS1 <sup>N8</sup>	N-terminal	+	ND
AlaRS1 <sup>N8</sup>	C-terminal	+	ND
AlaRS1 <sup>C1</sup>	N-terminal	+	ND
AlaRS1 <sup>C1</sup>	C-terminal	+	ND
AlaRS1 <sup>C2</sup>	N-terminal	+++	ND
AlaRS1 <sup>C2</sup>	C-terminal	+++	ND
AlaRS1 <sup>C3</sup>	N-terminal	+	ND
AlaRS1 <sup>C3</sup>	C-terminal	+	ND
AlaRS1 <sup>C4</sup>	N-terminal	+	ND
AlaRS1 <sup>C4</sup>	C-terminal	+	ND
AlaRS1 <sup>C5</sup>	N-terminal	+	ND
AlaRS1 <sup>C5</sup>	C-terminal	+	ND
AlaRS1 <sup>C6</sup>	N-terminal	+	ND
AlaRS1 <sup>C6</sup>	C-terminal	++	ND
AlaRS1 <sup>C7</sup>	N-terminal	+	ND
AlaRS1 <sup>C7</sup>	C-terminal	+	ND
AlaRS1 <sup>C11</sup>	N-terminal	+	ND
AlaRS1 <sup>C11</sup>	C-terminal	+	ND
AlaRS1 <sup>C12</sup>	N-terminal	+	ND
AlaRS1 <sup>C12</sup>	C-terminal	+	ND
AlaRS1 <sup>I1</sup>	N-terminal	+	++
AlaRS1 <sup>I1</sup>	C-terminal	+	+
Key			
“+” represents 0-1 mg/L AARS polypeptide expression “++” represents 1-5 mg/L AARS polypeptide expression; “+++” represents 5-10 mg/L AARS polypeptide expression; “++++” represents 10-15 mg/L AARS polypeptide expression; “+++++” represents ≥15 mg/L AARS polypeptide expression; ND: Not Determined			

[00640] Surprisingly, the protein expression data demonstrates the existence of several protein domains that exhibit high level expression of soluble protein when expressed in E.coli.

[00641] Specifically the data demonstrates that the AARS polypeptides AlaRS1<sup>N1</sup>, (amino acids 1-401), AlaRS1<sup>N3</sup>, (amino acids 1-488) and AlaRS1<sup>N6</sup> (amino acids 1-

497+24 amino acids) define the boundaries of a novel protein domain that is highly expressed in E.coli. This new AARS polypeptide domain comprises a portion of the full length Ala tRNA synthetase approximately spanning amino acids 1 to about 500.

[00642] Additionally the data demonstrates that the AARS polypeptides AlaRS1<sup>C2</sup> (amino acids 758-968) and AlaRS1<sup>C6</sup> (amino acids 747-968) define the boundaries of a second novel protein domain that is highly expressed in E.coli. This second new AARS polypeptide domain comprises a portion of the full length Ala tRNA synthetase approximately spanning amino acids from about 745 to 968.

### EXAMPLE 10

#### LARGE SCALE PRODUCTION OF AARS POLYPEPTIDES

[00643] Representative AARS polypeptides are prepared in larger amounts to enable further functional and biophysical characterization. The AARS polypeptides listed in **Table E4** are expressed in E.coli. in large scale culture as described in the General Materials and Methods section. The yields, and specified biophysical characteristics, for each expressed soluble protein are summarized below in **Table E4**.

<b>Table E4</b> <b>Summary of representative AARS Polypeptide yields and biophysical characterization</b>								
AARS Poly-peptide	Location of Epitope Tag	Yield [mg/L] <sup>(1)</sup>	Purity [%]	Endotoxin [EU/mg]	Molecular Weight	Working stock concentration [mg/mL]	Stability [percent recovery after 1 week] <sup>(2)</sup>	Aggregation [DLS]
AlaRS1 <sup>N3</sup>	C-terminal	1.5	80	6.0	C: 56,912 D: 56,907	13.8	98%	+
AlaRS1 <sup>I1</sup>	N-terminal	0.3 <sup>(3)</sup>	80	NA	ND	1.83	ND	ND
AlaRS1 <sup>C6</sup>	C-terminal	9.5	99	0.4	C: 26,771 D: 26,642 <sup>(4)</sup>	19.04	60%	++++
<b>Notes</b> (1): Yield determined by measuring protein recovery after last purification step (2): Determined as percent recovery of non aggregated material after 1 week at 25°C (3): Measured after final Amicon concentration step (4): Likely to represent MW without N-terminal methionine C: Calculated D: Determined								
<b>Key:</b> “+” represents less than 1 % high molecular protein aggregates “++” represents less than 2% high molecular protein aggregates “+++” represents less than 5% high molecular protein aggregates “++++” represents less than 10 % high molecular protein aggregates ND: Not Determined								

[00644] The results from these studies establish that representative AARS proteins from these two families of AARS proteins, (AlaRS1<sup>N3</sup> and AlaRS1<sup>C6</sup>) exhibit favorable protein expression yields, as well as good solubility, stability and aggregation characteristics, even without extensive optimization.

#### EXAMPLE 11

##### TRANSCRIPTIONAL PROFILING OF REPRESENTATIVE AARS POLYPEPTIDES

[00645] To test for the ability of the AARS polypeptides to modulate gene expression, selected AARS polypeptides were incubated with Mesenchymal stems cells or human skeletal muscle cells for the times shown in **Table E5**.

<b>Table E5</b>						
<b>Transcriptional profiling of representative AARS Polypeptides in Mesenchymal Stem Cells (MSC) or Human Skeletal Muscle Cells (HskMC)</b>						
<b>Test Sample Description</b>		<b>Cell type and Exposure Time</b>				
<b>AARS Polypeptides</b>	<b>Location of Epitope Tag</b>	<b>Concentration nM</b>	<b>MSC 24 hours</b>	<b>MSC 72 hours</b>	<b>HskMC 24 hours</b>	<b>HskMC 72 hours</b>
AlaRS1 <sup>N1</sup>	C-terminal	57	19	5	11	4
AlaRS1 <sup>N3</sup>	N-terminal	41	1	0	0	8
AlaRS1 <sup>N3</sup>	C-terminal	250	3	3	9	7
AlaRS1 <sup>N6</sup>	C-terminal	250	0	3	4	5
AlaRS1 <sup>C2</sup>	N-terminal	80	1	6	7	2
AlaRS1 <sup>C2</sup>	C-terminal	250	0	2	9	10
AlaRS1 <sup>C6</sup>	C-terminal	250	1	3	3	1
<b>Controls</b>						
Average across all AARS polypeptides screened			3	5	6	7
Osteogenesis cocktail			17	20	11	16
Chondrogenesis cocktail			17	19	14	19
Adipogenesis cocktail			19	15	16	18
SKMC Pos Ctrl			11	8	5	4
Untreated			0	0	1	1

[00646] In **Table E5**, the numbers in each column represent the number of genes which were modulated, either positively or negatively by at least 4 fold compared to the control samples, as described in the general methods section. The data shows that specific forms of the AARS polypeptides tested have the surprising ability to regulate the transcription, and hence potentially modulate the developmental fate or differentiation status when added to either Mesenchymal Stem Cells (MSC) and / or Human Skeletal Muscle Cells (HskMC). Shaded cells with bolded numbers in the table represent examples where the AARS polypeptide exhibits a significant impact on the regulation of gene transcription in the cell lines and times indicated in the table.

[00647] It is concluded that AlaRS1<sup>C2</sup>, AlaRS1<sup>N1</sup> and AlaRS1<sup>N3</sup> appear to be major regulators of mesenchymal Stem Cell and / or Human Skeletal Muscle Cell gene expression.

## EXAMPLE 12

### FUNCTIONAL PROFILING OF REPRESENTATIVE AARS POLYPEPTIDES

[00648] To test for the ability of the AARS polypeptides to modulate a range of phenotypic processes, selected AARS polypeptides were incubated with the cell types, and the conditions provided in the general methods section, and **Tables E5** and **E6**.

Table E6 Key to Assays and criteria for indicating a hit	
Proliferation assays	
Source and cell type	Assay Number
Human megakaryocytic leukemia cells / Mo7e	A1
Human acute promyelocytic leukemia cells / HL60	A2
Human lymphoblast (cancer cell line) / RPMI8226	A3
Human mesenchymal stem cells / hMSC	A4
Human astrocytes	A5
Human bone marrow aspirate cells / Bone Marrow Cells	A6
Human bone marrow aspirate cells/ Bone Marrow Cells (Long Term Culture)	A7
Human Synoviocyte / HFLS-SynRA	A8
Human pre-adipocyte cells /hPAD	A9
Human pulmonary artery smooth muscle cell /hPASMC	A10
Human skeletal muscle cell /hSKMC	A11
Data analysis for proliferation assays was performed by dividing the numerical value in the assay well by the average PBS value for the assay plate. AARS polypeptides were considered to be proliferative if greater than 3 SD away from the PBS value in the positive direction. AARS polypeptide was considered to be cytotoxic if greater than 3 SD away from the PBS value in the negative direction. A cytotoxic compound was utilized as a negative control and the average value for this was always greater than 3 SD away from PBS average value.	
Cellular differentiation and phenotype assays	
Assay Description	Assay Number
Human hepatocyte (HepG2C3a cells) acetylated LDL uptake	B1
Data analysis for ac-LDL uptake assay was performed by dividing the numerical value in the assay well by the average PBS value for the assay plate. AARS polypeptides were considered to be a modulator of ac-LDL uptake if greater than 2 SD away from the PBS value in the positive or negative direction. A visual check to confirm plate reader results was made using a fluorescent microscope.	



<b>Table E6</b> <b>Key to Assays and criteria for indicating a hit</b>	
<b>Human Neutrophil assays</b>	
<b>Assay Description</b>	<b>Assay Number</b>
Neutrophil Elastase	C1
Neutrophil oxidative burst (agonist)	C2
Neutrophil oxidative burst (antagonist)	C3
Data analysis for neutrophil assays was performed by dividing the numerical value in the assay well by the average PBS value for the assay plate. AAR polypeptides were considered to be a modulator of neutrophil elastase production or oxidative burst biology if greater than 2 SD away from the PBS value in the positive or negative direction.	
<b>Modulation of Toll-like receptors (TLR)</b>	
<b>Assay Description</b>	<b>Assay Number</b>
TLR activation in RAW BLUE cells	D1
TLR antagonism in RAW BLUE cells	D2
Activation of hTLR2	D3
Activation of hTLR4	D4
Data analysis for TLR modulation assays was performed by dividing the numerical value in the assay well by the average PBS value for the assay plate. AARS polypeptides were considered to be a modulator of TLR specific biology if greater than 3 SD away from the PBS value in the positive or negative direction. Positive controls, including LPS and detection reagent were always significantly distinct and > 3 SD from PBS average value.	
<b>Cytokine Release</b>	
<b>Assay Description</b>	<b>Assay Number</b>
Human Synoviocyte cytokine production (IL6 release)	E1
Human pulmonary artery smooth muscle cell (hPASMC) cytokine production (IL6 release)	E2
Human skeletal muscle cell (hSKMC) cytokine production (IL6 release)	E3
Human Astrocyte cytokine production (IL6 release)	E4
Whole blood IL6 release	E5
Human pulmonary artery smooth muscle cell (hPASMC) cytokine production (IL8release) 72 h Incubation	E6
<b>IL8 production</b>	
<b>Assay Description</b>	<b>Assay Number</b>
Human Synoviocyte cytokine production (IL8 release)	E7
Human pulmonary artery smooth muscle cell (hPASMC) cytokine production (IL8release)	E8
Human skeletal muscle cell (hSKMC) cytokine production (IL8 release)	E9
Human Astrocyte cytokine production (IL8 release)	E10
Human hepatocyte (HepG2C3a cells) IL8 release	E11
Human acute promyelocytic leukemia cells / HL60 (IL8 release)	E12

<b>Table E6</b>	
<b>Key to Assays and criteria for indicating a hit</b>	
Human lymphoblast (cancer cell line) / RPMI8226 (IL8 Release)	E13
TNF alpha production	
Human Synovioocyte cytokine production (TNF alpha release)	E14
Whole blood TNF alpha release	E15
IL10 Release	
Human acute promyelocytic leukemia cells / HL60 IL10 release	E16
Human Primary Blood Mononuclear cells (IL10 Release)	E17
Data analysis for cytokine release assays was performed by dividing the numerical value in the assay well by the average PBS value for the assay plate. AARS polypeptides were considered to be a modulator of cytokine production or cytokine related biology if greater than 2 SD away from the PBS value in the positive or negative direction. A protein standard (specific to each assay kit) was run on every plate to insure good assay quality. Only assays with protein standard curves that had an $R^2$ value of > than 0.9 were chosen for data analysis.	
<b>Cell Adhesion and Chemotaxis</b>	
<b>Assay Description</b>	<b>Assay Number</b>
Monocyte THP 1/ Human umbilical vein endothelial cell (HUVEC) cell adhesion	F1
Human hepatocyte (HepG2C3a cells) (ICAM release)	F2
Human lung microvascular endothelial cell (HLMVEC) cell adhesion regulation (ICAM release)	F3
Human umbilical vein endothelial cell (HUVEC) cell adhesion regulation (VCAM release)	F4
Human mesenchymal stem cell (hMSC) cell adhesion regulation (VCAM release)	F5
Human skeletal muscle cell (hSKMC) cell adhesion regulation (VCAM release)	F6
Human pulmonary artery smooth muscle cell (hPASMC) cell adhesion regulation (VCAM release)	F7
Data analysis for cell adhesion regulation assays was performed by dividing the numerical value in the assay well by the average PBS value for the assay plate. AARS polypeptides were considered to be a modulator of cell adhesion or a regulator of biology related to cell adhesion if a value of greater than 2 SD away from the PBS value in the positive or negative direction was noted. In the case of the ELISA assays, a protein standard (specific to each assay kit) was run on every plate to insure good assay quality. Only assays with protein standard curves that had an $R^2$ value of > than 0.9 were chosen for data analysis.	
<b>Cellular Differentiation</b>	
<b>Assay Description</b>	<b>Assay Number</b>
Human pre-adipocyte (hPAD) cell differentiation	G1
Human skeletal muscle (hSKMC) cell differentiation	G2
Human mesenchymal stem (hMSC) cell differentiation	G3
Human pulmonary artery smooth muscle cell (hPASMC) differentiation	G4
Data analysis for cellular differentiation assays was performed by dividing the	

<b>Table E6</b> <b>Key to Assays and criteria for indicating a hit</b>	
numerical value in the assay well by the average PBS value for the assay plate. Differentiation assays were scored based upon fluorescent intensity of particular antibodies as described in the methods section. AARS polypeptides were considered to be a modulator of cellular differentiation if an intensity reading of greater than 2 SD away from the PBS value in the positive or negative direction was noted. For the hSKMC analysis, digital photos were taken of all wells and photos were scored in a blinded fashion by three people using a 4 point scoring system where 4 indicated intense skeletal muscle actin staining and obvious myotube formation. The average value from visual scoring was used and only wells with an average value of > 3 were considered hits. Differentiation control treated wells in this assay typically scored > 2, while PBS treated wells scored <2.	
Cell Binding	
Assay Description	Assay Number
PBMC	H1
Primary T cell	H2
Primary B cell	H3
Primary Monocyte	H4
HepG2	H5
3T3L1	H6
C2C12	H7
THP1	H8
Jurkat	H9
Raji	H10
AARS polypeptides were considered to be binding to a particular cell type if the mean cell bound fluorescence intensity was greater than 2 SD away from the reagent control values for that cell type.	

<b>Table E7</b> <b>Results of Functional Profiling studies of Representative AARS Polypeptides</b>			
AARS polypeptide	Location of Epitope Tag	Conc / [nM]	Assay Hits
AlaRS1 <sup>N1</sup>	C-terminal	139	D1 (Modulation of Toll-like receptors) E4, (Cytokine Release) G1 G3 (Cellular Differentiation)
AlaRS1 <sup>N3</sup>	N-terminal	114	G1 (Cellular Differentiation)
AlaRS1 <sup>N3</sup>	C-terminal	250	F2 (Cell Adhesion and Chemotaxis) G1 (Cellular Differentiation)
AlaRS1 <sup>N6</sup>	N-terminal	56	
AlaRS1 <sup>N6</sup>	C-terminal	40	F5 (Cell Adhesion and Chemotaxis)
AlaRS1 <sup>I1</sup>	N-terminal	172	A9 (Proliferation) C2 (Neutrophil activation) D1, D2, (Modulation of Toll-like receptors)

<b>Table E7</b> <b>Results of Functional Profiling studies of Representative AARS Polypeptides</b>			
<b>AARS polypeptide</b>	<b>Location of Epitope Tag</b>	<b>Conc / [nM]</b>	<b>Assay Hits</b>
			E1, E8 (Cytokine Release)
AlaRS1 <sup>C2</sup>	N-terminal	250	E4 (Cytokine Release) F5 (Cell Adhesion and Chemotaxis)
AlaRS1 <sup>C2</sup>	C-terminal	250	A6 (Proliferation)
AlaRS1 <sup>C6</sup>	C-terminal	250	A4 (Proliferation) C2 (Neutrophil activation) E15, E17 (Cytokine Release) H1-H9 (Cell binding)

[00649] It is concluded that AlaRS1<sup>N1</sup> and AlaRS1<sup>N3</sup> appear to be major regulators of cellular differentiation, cytokine release and the potentially modulators of Toll like receptor activity, while AlaRS1<sup>N6</sup> may play a role in regulating cell adhesion and chemotaxis. When viewed in light of the transcriptional profiling data, the phenotypic screening data demonstrates that the AARS polypeptides AlaRS1<sup>N1</sup>, (amino acids 1-401), AlaRS1<sup>N3</sup>, (amino acids 1-488) and AlaRS1<sup>N6</sup> (amino acids 1-497+24 amino acids) define the boundaries of a novel protein domain that is highly active in a broad array of phenotypic screening assays.

[00650] Accordingly it is concluded that AARS polypeptides comprising amino acids 1 to 497 amino acids of Alanyl tRNA synthetase (and optionally a variable c-terminal amino acid extension of about 1 to about 30 amino acids) defines the approximate boundaries of a novel, highly active AARS polypeptide domain, that is i) highly functionally active, ii) can be readily made and produced in E.coli, and iii) exhibits favorable protein stability and aggregation characteristics. It will be appreciated by those of skill in the art that any AARS polypeptides comprising as few as about the first 400 amino acids of the Alanyl tRNA synthetase, to as large as about the first 500 amino acids of Alanyl tRNA synthetase represent functional equivalents of the specific AARS polypeptides described.

[00651] Moreover AlaRS1<sup>C2</sup> and AlaRS1<sup>C6</sup>, appear to be major regulators of proliferation and cellular differentiation, as well as neutrophil activation and cytokine release. When viewed in light of the transcriptional profiling data, the phenotypic screening data demonstrates that the AARS polypeptides AlaRS1<sup>C2</sup>, (amino acids 758-968) and AlaRS1<sup>C6</sup>, (amino acids 747-968) define the boundaries of a second novel AARS polypeptide domain that is highly active in a broad array of phenotypic screening assays.

[00652] Accordingly it is concluded that AARS polypeptides comprising amino acids 747-968 amino acids of Alanyl tRNA synthetase defines the approximate boundaries of a novel, highly active AARS polypeptide domain, that is i) highly functionally active, ii) can be readily made and produced in E.coli, and iii) exhibits favorable protein stability and aggregation characteristics. It will be appreciated by those of skill in the art that any AARS polypeptides comprising as few as about the last 210 amino acids of the Alanyl tRNA synthetase, to as large as about the last 221 amino acids of Alanyl tRNA synthetase represent functional equivalents of the specific AARS polypeptides described.

### EXAMPLE 13

#### ANIMAL STUDIES

[00653] To test for the ability of the AARS polypeptides to modulate hematopoiesis and inflammatory responses, a representative AARS polypeptide, AlaRS1<sup>C6</sup>, prepared as described in **Table E2**, was administered to wild type and DIO mice, as described in the general methods. Results with respect to Cell Blood Counts and Cytokine release are shown in **Tables E8** and **E9** respectively.

<b>Table E8</b> <b>Influence of AlaRS1<sup>C6</sup> on hematopoiesis in wild type and diet induced obese (DIO) mice</b>		
<b>Cell type</b>	<b>Wild type Mice</b>	<b>DIO Mice</b>
White Blood Cells	-	+
Neutrophils	0	0
Lymphocytes	-	+
Monocytes	0	0
HGB	0	0
HCT	0	0
MCV	0	0
MCH	0	0
MCHC	0	0
Platelets	-	-

<b>Table E9</b> <b>Influence of AlaRS1<sup>C6</sup> on Cell Blood Counts and cytokine release in wild type and diet induced obese mice</b>			
C57BL/6 Mice on Normal Diet			
AARS Polypeptide	CBC	Acute Cytokine	Chronic Cytokine
AlaRS1 <sup>C6</sup>	+	15	3
C57BL/6 Mice on High Fat Diet			

AlaRS1 <sup>C6</sup>	+	16	2
----------------------	---	----	---

[00654] The results from the *in vivo* animal studies show that daily administration of AlaRS1<sup>C6</sup> at 10 mg/kg for 4 days promoted white blood cell and lymphocyte levels, while reducing platelet counts in whole animals, in both mice and diet induced obese mice. Moreover the animal studies demonstrated that AlaRS1<sup>C6</sup> was safe and well tolerated at this dose, and produced no overt negative effects in treated animals. AlaRS1<sup>C6</sup> also promoted cell blood counts and acute cytokine release, suggesting a role in hematopoiesis, immune function and inflammatory responses.

2011248489 14 Sep 2016

## THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A therapeutic composition, comprising an isolated aminoacyl-tRNA synthetase (AARS) polypeptide that is at least 95%, 98%, or 100% identical to SEQ ID NO: 81 or 142, wherein the polypeptide has an extra-cellular signalling activity and has a solubility of at least about 5 mg/mL, and wherein the composition has a purity of at least about 95% on a protein basis and less than about 10 EU endotoxin/mg protein.
2. The therapeutic composition of claim 1, wherein the AARS polypeptide consist of SEQ ID NO: 81 or 142 and differs from SEQ ID NO: 81 or 142, by substitution, deletion, and/or addition of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or 11 amino acids.
3. The therapeutic composition of claim 1 or claim 2, wherein the AARS polypeptide is fused to a heterologous polypeptide.
4. The therapeutic composition of claim 3, wherein the heterologous polypeptide is selected from the group consisting of purification tags, epitope tags, targeting sequences, signal peptides, membrane translocating sequences, and PK modifiers.
5. The therapeutic composition of any one of claims 1-4, wherein at least one moiety is covalently or non-covalently attached to said polypeptide, or wherein a solid substrate is attached to said polypeptide.
6. A cellular composition or growth device, comprising an aminoacyl- tRNA synthetase (AARS) polypeptide that is at least 95%, 98% or 100% identical to SEQ ID NO:81 or 142, or a fragment thereof which is 200 or more contiguous amino acids of SEQ ID NO:81 or 142, and an element selected from the group consisting of

2011248489 14 Sep 2016

- (i) an engineered population of cells in which at least one cell comprises a polynucleotide encoding said AARS polypeptide, wherein the cells are capable of growing in a serum-free medium,
  - (ii) an engineered population of cells in which at least one cell comprises a polynucleotide encoding said AARS polypeptide, at least about 10 liters of a serum- free growth medium, and a sterile container.
- 7. A method of identifying a compound that specifically binds to an aminoacyl-tRNA synthetase (AARS) polypeptide that is at least 95%, 98%, or 100% identical to SEQ ID NO: 81 or 142, or a fragment thereof which is 200 or more contiguous amino acids of SEQ ID NO: 81 or 142, comprising
  - a) combining the AARS polypeptide with at least one test compound under suitable conditions, and
  - b) detecting binding of the AARS polypeptide to the test compound, thereby identifying a compound that specifically binds to the AARS polypeptide.
- 8. An engineered full length aminoacyl-tRNA synthetase (AARS) protein comprising a heterologous proteolysis site to enable the proteolytic generation of an AARS polypeptide that is at least 95%, 98% or 100% identical to SEQ ID NO:81 or 142, or a fragment thereof which is 200 or more contiguous amino acids of SEQ ID NO:81 or 142.
- 9. A method for treating a neurological disease in a subject in need thereof comprising administering to the subject an effective amount of a therapeutic composition of any one of claims 1-5.
- 10. The method of claim 14, wherein the neurological disease is neuro-inflammation, neurodegeneration, or brain ischemia.
- 11. A method for treating chronic granulomatous disease (CGD) or a condition where the oxidative burst of granulocytes is impaired in a subject in need thereof



2011248489 14 Sep 2016

comprising administering to the subject an effective amount of a therapeutic composition of any one of claims 1-5.

12. A method for treating an inflammatory disease in a subject in need thereof comprising administering to the subject an effective amount of a therapeutic composition of any one of claims 1-5.
13. The method of claim 17, wherein the inflammatory disease is selected from cardiovascular diseases, atherosclerosis, autoimmunity, and pulmonary hypertension.

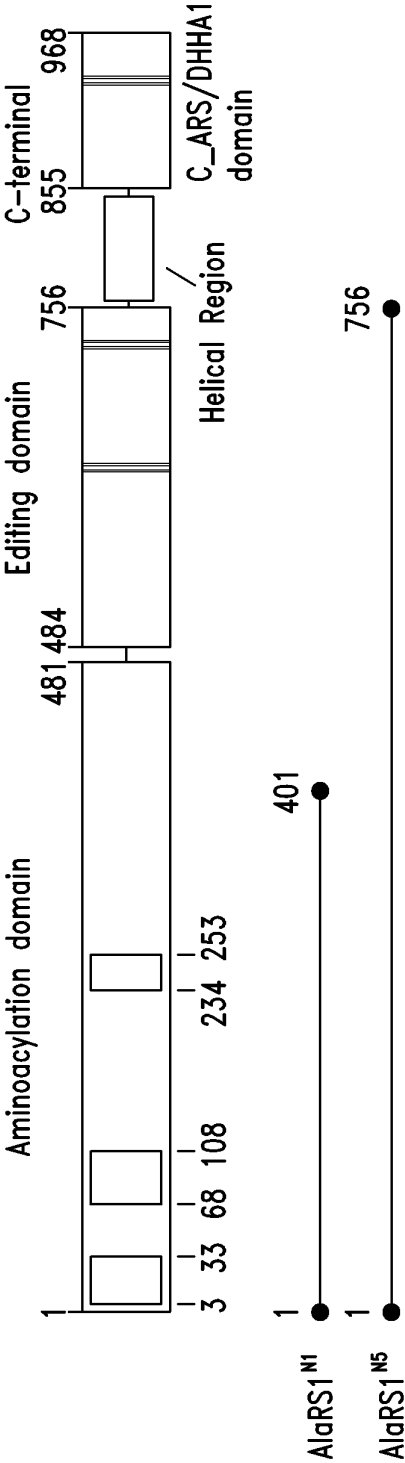


FIG. 1A

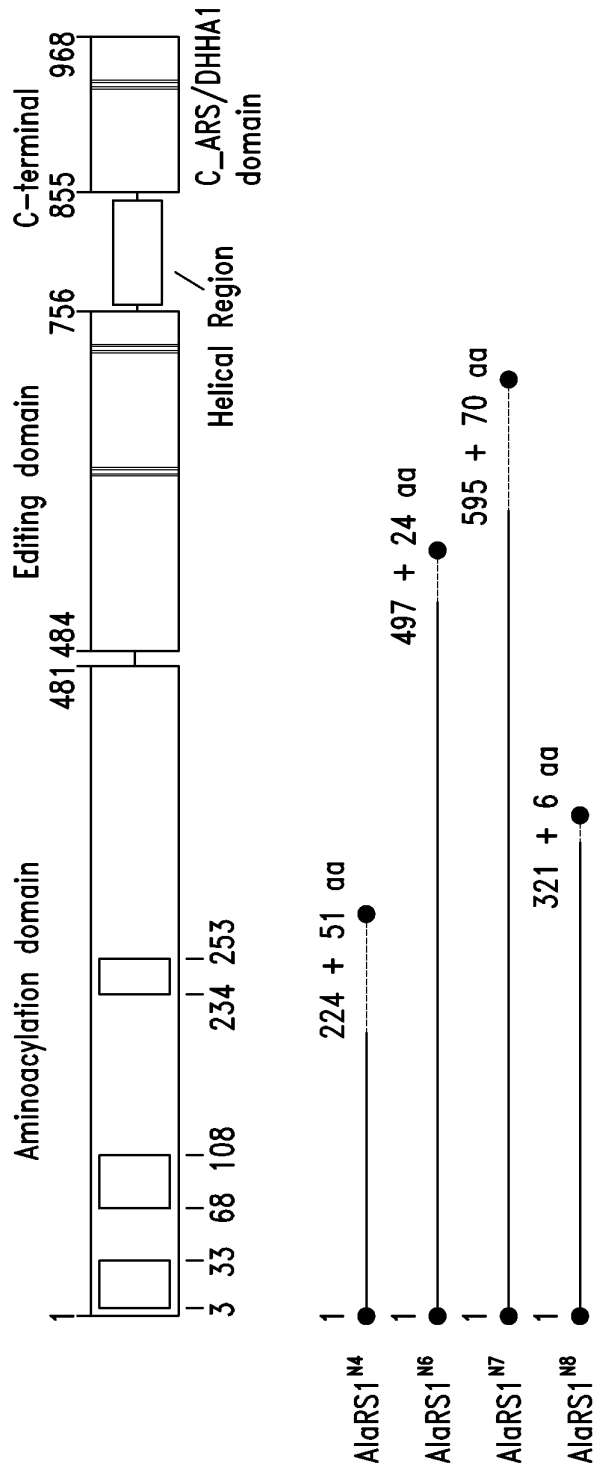


FIG. 1B

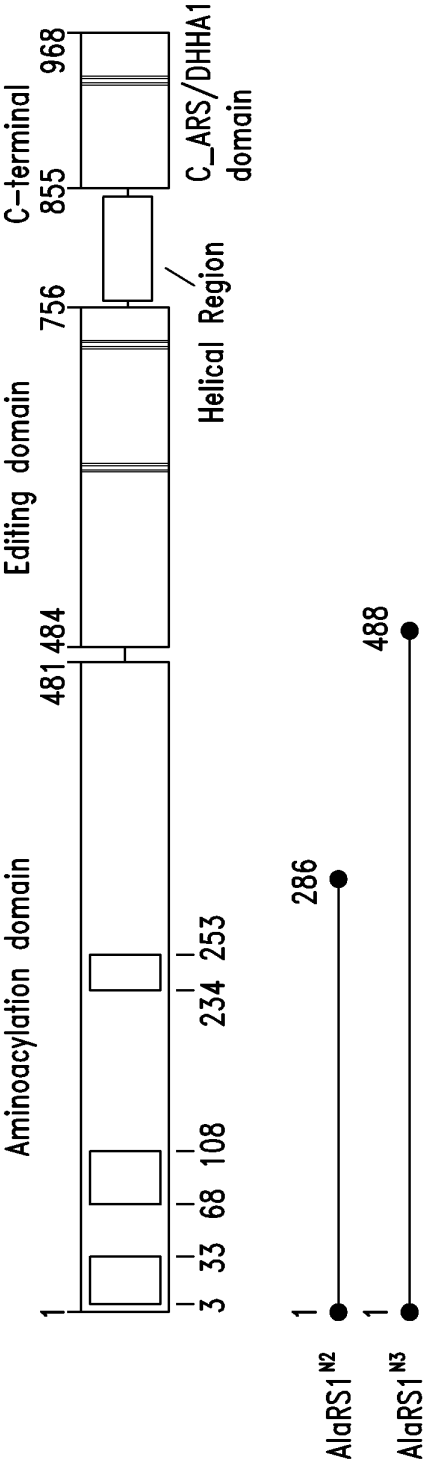


FIG. 1C

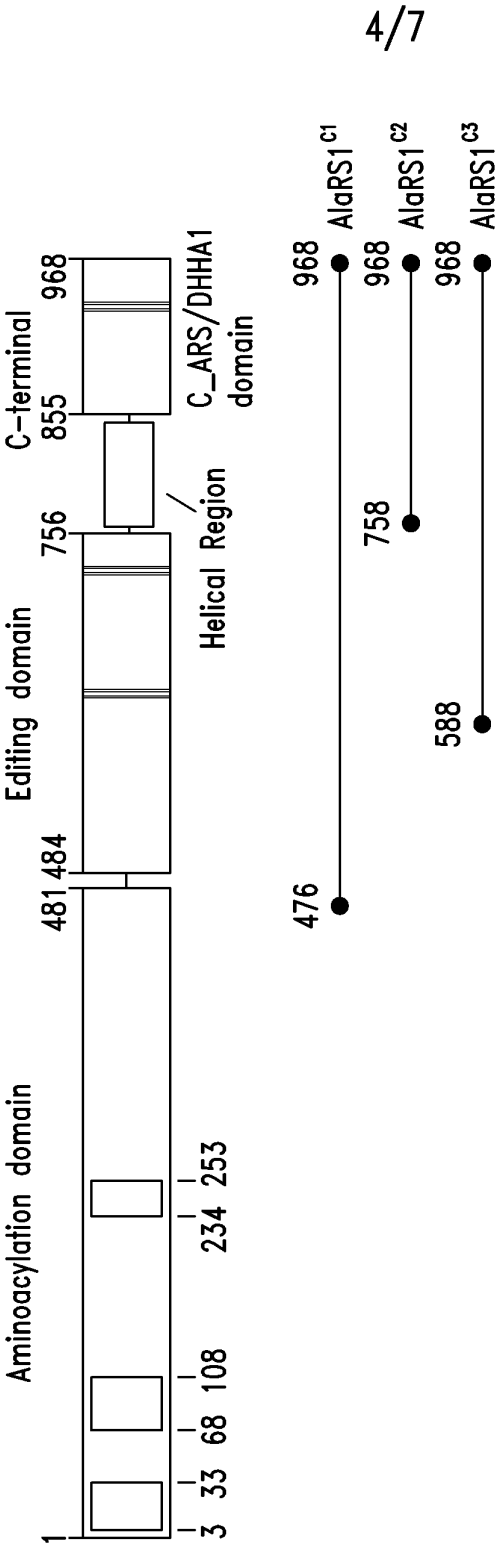


FIG. 2A

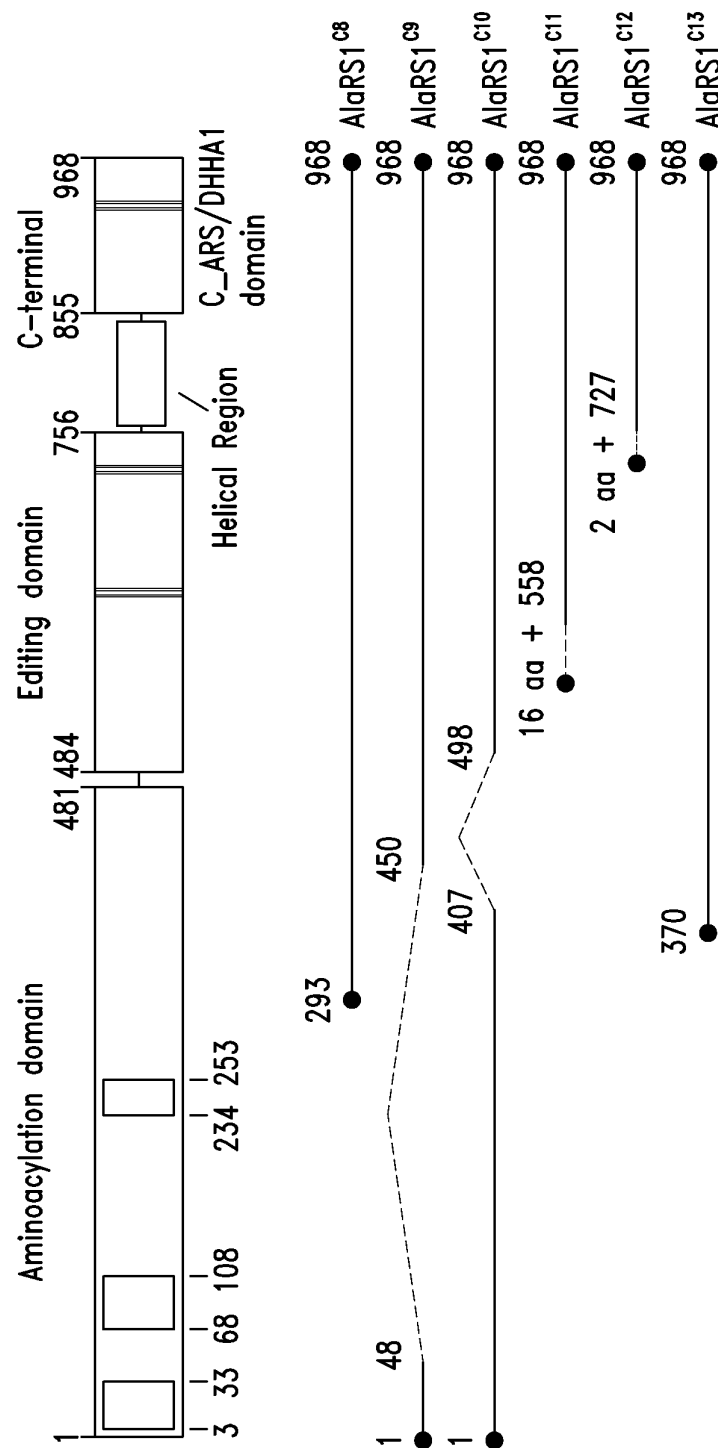


FIG. 2B

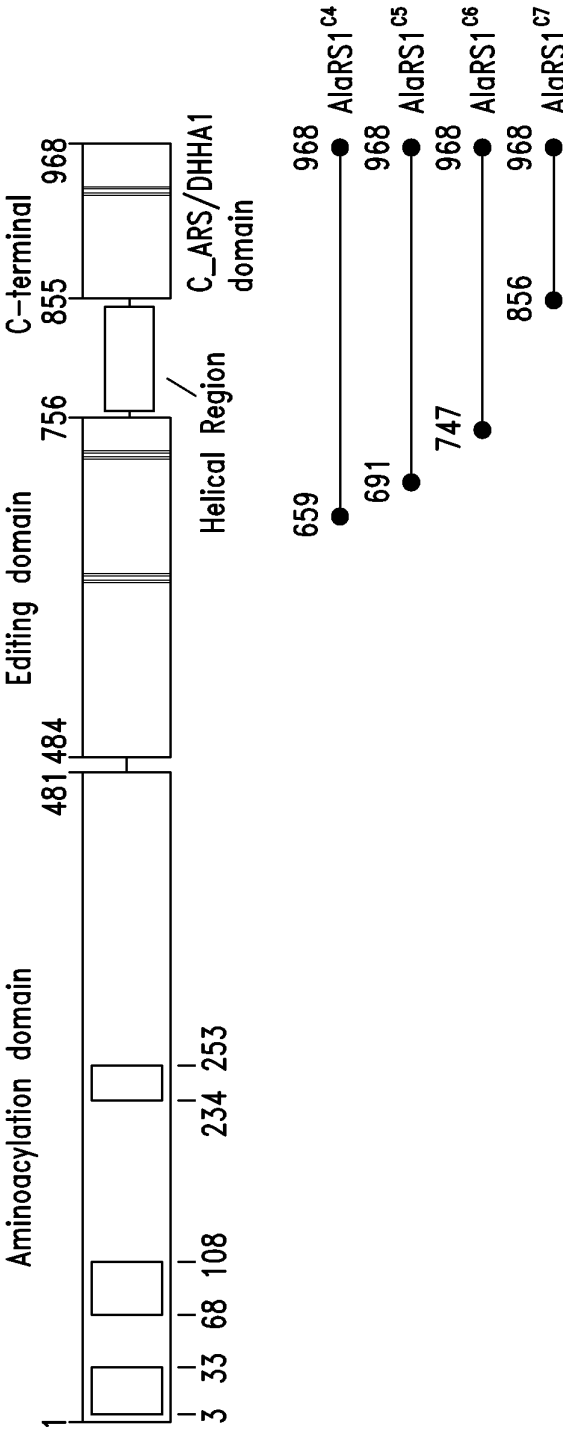


FIG. 2C

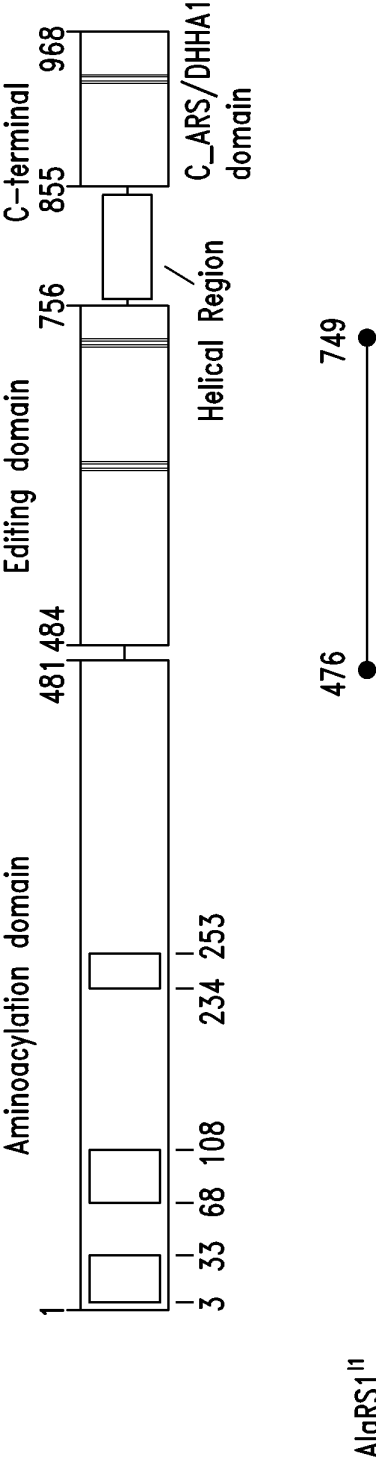


FIG. 3



## SEQUENCE LISTING

<110> aTyr Pharma Inc.  
Pangu Biopharma Limited  
Greene, Leslie Ann  
Chiang, Kyle P.  
Hong, Fei  
Vasserot, Alain P.  
Lo, Wing-Sze  
Watkins, Jeffry D.  
Mendlein, John D.  
Quinn, Cheryl L.

<120> INNOVATIVE DISCOVERY OF THERAPEUTIC,  
DIAGNOSTIC, AND ANTIBODY COMPOSITIONS RELATED TO PROTEIN FRAGMENTS  
OF ALANYL-TRNA SYNTHETASE

<130> 120161.426PC

<140> PCT  
<141> 2011-04-28

<150> US 61/329,048  
<151> 2010-04-28

<150> US 61/328,867  
<151> 2010-04-28

<150> US 61/328,871  
<151> 2010-04-28

<160> 186

<170> FastSEQ for Windows Version 4.0

<210> 1  
<211> 77  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide

<400> 1  
aggaggtaaa acatatgcat catcatcatc atcacggtaa gcctatccct aaccctttgc 60  
tcggtctcga ttctacg 77

<210> 2  
<211> 12  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide

<400> 2  
taatgactcg ag 12

<210> 3  
<211> 17  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide

<400> 3  
 aggagataaa acatatg 17  
  
 <210> 4  
 <211> 14  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Oligonucleotide  
  
 <400> 4  
 aggaggtaaa acat 14  
  
 <210> 5  
 <211> 14  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Oligonucleotide  
  
 <400> 5  
 aggagataaa acat 14  
  
 <210> 6  
 <211> 15  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Oligonucleotide  
  
 <400> 6  
 gaaggagata tacat 15  
  
 <210> 7  
 <211> 72  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Oligonucleotide  
  
 <400> 7  
 ggtaagccta tccctaacc tctcctcggc ctcgattcta cgcaccacca tcatcaccat 60  
 taatgactcg ag 72  
  
 <210> 8  
 <211> 72  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Oligonucleotide  
  
 <400> 8  
 catatgcatc atcatcatca tcacggtaag cctatcccta accctctcct cggctctcgat 60  
 tctacgggat cc 72  
  
 <210> 9  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Oligonucleotide

<400> 9  
ctcgagtaat ga 12

<210> 10  
<211> 12  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide

<400> 10  
catatgggat cc 12

<210> 11  
<211> 72  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide

<400> 11  
ctcgagggta agcctatccc taaccctctc ctcggtctcg attctacgca ccaccaccac 60  
caccactaat ga 72

<210> 12  
<211> 401  
<212> PRT  
<213> Homo sapiens

<400> 12  
Met Asp Ser Thr Leu Thr Ala Ser Glu Ile Arg Gln Arg Phe Ile Asp  
1 5 10 15  
Phe Phe Lys Arg Asn Glu His Thr Tyr Val His Ser Ser Ala Thr Ile  
20 25 30  
Pro Leu Asp Asp Pro Thr Leu Leu Phe Ala Asn Ala Gly Met Asn Gln  
35 40 45  
Phe Lys Pro Ile Phe Leu Asn Thr Ile Asp Pro Ser His Pro Met Ala  
50 55 60  
Lys Leu Ser Arg Ala Ala Asn Thr Gln Lys Cys Ile Arg Ala Gly Gly  
65 70 75 80  
Lys His Asn Asp Leu Asp Asp Val Gly Lys Asp Val Tyr His His Thr  
85 90 95  
Phe Phe Glu Met Leu Gly Ser Trp Ser Phe Gly Asp Tyr Phe Lys Glu  
100 105 110  
Leu Ala Cys Lys Met Ala Leu Glu Leu Leu Thr Gln Glu Phe Gly Ile  
115 120 125  
Pro Ile Glu Arg Leu Tyr Val Thr Tyr Phe Gly Gly Asp Glu Ala Ala  
130 135 140  
Gly Leu Glu Ala Asp Leu Glu Cys Lys Gln Ile Trp Gln Asn Leu Gly  
145 150 155 160  
Leu Asp Asp Thr Lys Ile Leu Pro Gly Asn Met Lys Asp Asn Phe Trp  
165 170 175  
Glu Met Gly Asp Thr Gly Pro Cys Gly Pro Cys Ser Glu Ile His Tyr  
180 185 190  
Asp Arg Ile Gly Gly Arg Asp Ala Ala His Leu Val Asn Gln Asp Asp  
195 200 205  
Pro Asn Val Leu Glu Ile Trp Asn Leu Val Phe Ile Gln Tyr Asn Arg  
210 215 220  
Glu Ala Asp Gly Ile Leu Lys Pro Leu Pro Lys Lys Ser Ile Asp Thr  
225 230 235 240  
Gly Met Gly Leu Glu Arg Leu Val Ser Val Leu Gln Asn Lys Met Ser  
245 250 255  
Asn Tyr Asp Thr Asp Leu Phe Val Pro Tyr Phe Glu Ala Ile Gln Lys  
260 265 270  
Gly Thr Gly Ala Arg Pro Tyr Thr Gly Lys Val Gly Ala Glu Asp Ala  
275 280 285

120161\_426PC\_Sequence\_Listing.TXT

Asp	Gly	Ile	Asp	Met	Ala	Tyr	Arg	Val	Leu	Ala	Asp	His	Ala	Arg	Thr
290						295					300				
Ile	Thr	Val	Ala	Leu	Ala	Asp	Gly	Gly	Arg	Pro	Asp	Asn	Thr	Gly	Arg
305					310					315					320
Gly	Tyr	Val	Leu	Arg	Arg	Ile	Leu	Arg	Arg	Ala	Val	Arg	Tyr	Ala	His
				325					330					335	
Glu	Lys	Leu	Asn	Ala	Ser	Arg	Gly	Phe	Phe	Ala	Thr	Leu	Val	Asp	Val
			340					345					350		
Val	Val	Gln	Ser	Leu	Gly	Asp	Ala	Phe	Pro	Glu	Leu	Lys	Lys	Asp	Pro
		355					360					365			
Asp	Met	Val	Lys	Asp	Ile	Ile	Asn	Glu	Glu	Glu	Val	Gln	Phe	Leu	Lys
370						375					380				
Thr	Leu	Ser	Arg	Gly	Arg	Arg	Ile	Leu	Asp	Arg	Lys	Ile	Gln	Ser	Leu
385					390					395					400
Gly															

<210> 13  
 <211> 1203  
 <212> DNA  
 <213> Homo sapiens

<400> 13

atggactcta	ctctaacagc	aagtgaatc	cggcagcgat	ttatagattt	cttcaagagg	60
aacgagcata	cgatgttca	ctcgtctgcc	accatcccat	tggatgacct	cactttgctc	120
tttgccaatg	caggcatgaa	ccagtttaaa	cccattttcc	tgaacacaat	tgacctatct	180
caccccatgg	caaagctgag	cagagctgcc	aataccaga	agtgcattcg	ggctgggggc	240
aaacataatg	acctggacga	tgtgggcaag	gatgtctatc	atcacacctt	cttcgagatg	300
ctgggctctt	ggtcttttgg	agattacttt	aaggaattgg	catgtaagat	ggctctggaa	360
ctccacacc	aagagtttgg	cattcccatt	gaaagacttt	atgttactta	ctttggcggg	420
gatgaagcag	ctggcttaga	agcagatctg	gaatgcaaac	agatctggca	aaatttgggg	480
ctggatgaca	ccaaaatcct	cccaggcaac	atgaaggata	acttctggga	gatgggtgac	540
acggggcccct	gtggctcttg	cagtgaatc	cactacgacc	ggattggtgg	tcgggacgcc	600
gcacatcttg	tcaaccagga	cgaccctaatt	gtgctggaga	tctggaacct	tgtgttcatt	660
cagtataaca	gggaagctga	tggcattctg	aaacctcttc	ccaagaaaag	cattgacaca	720
gggatggggc	tgaacgact	ggtatctgtg	ctgcagaata	agatgtccaa	ctatgacact	780
gacctttttg	tcccttactt	tgaagccatt	cagaaggcca	caggtgcccg	accatacact	840
gggaaagtgt	gtgctgagga	tgccgatggg	attgacatgg	cctaccgggt	gctggctgac	900
cacgctcgga	ccatcactgt	ggcactggct	gatggtggcc	ggcctgacaa	cacagggcgt	960
ggatatttgt	tgagacggat	tctccgccga	gctgtccgat	acgcccata	aaagctcaat	1020
gccagcaggg	gcttctttgc	tacgttagtg	gatgttgtcg	tccagtcctt	gggagatgca	1080
tttcttgagc	tgaagaagga	cccagacatg	gtgaaggaca	tcattaatga	agaagagggtg	1140
cagtttctca	agactctcag	cagagggcgt	cgcattcctg	acaggaaaat	tcagagcctg	1200
gga						1203

<210> 14  
 <211> 756  
 <212> PRT  
 <213> Homo sapiens

<400> 14

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

120161\_426PC\_Sequence\_Listing.TXT

Pro	Ile	Glu	Arg	Leu	Tyr	Val	Thr	Tyr	Phe	Gly	Gly	Asp	Glu	Ala	Ala
130						135					140				
Gly	Leu	Glu	Ala	Asp	Leu	Glu	Cys	Lys	Gln	Ile	Trp	Gln	Asn	Leu	Gly
145					150					155					160
Leu	Asp	Asp	Thr	Lys	Ile	Leu	Pro	Gly	Asn	Met	Lys	Asp	Asn	Phe	Trp
				165					170					175	
Glu	Met	Gly	Asp	Thr	Gly	Pro	Cys	Gly	Pro	Cys	Ser	Glu	Ile	His	Tyr
				180				185					190		
Asp	Arg	Ile	Gly	Gly	Arg	Asp	Ala	Ala	His	Leu	Val	Asn	Gln	Asp	Asp
		195					200					205			
Pro	Asn	Val	Leu	Glu	Ile	Trp	Asn	Leu	Val	Phe	Ile	Gln	Tyr	Asn	Arg
	210					215					220				
Glu	Ala	Asp	Gly	Ile	Leu	Lys	Pro	Leu	Pro	Lys	Lys	Ser	Ile	Asp	Thr
225					230					235					240
Gly	Met	Gly	Leu	Glu	Arg	Leu	Val	Ser	Val	Leu	Gln	Asn	Lys	Met	Ser
				245					250					255	
Asn	Tyr	Asp	Thr	Asp	Leu	Phe	Val	Pro	Tyr	Phe	Glu	Ala	Ile	Gln	Lys
			260					265					270		
Gly	Thr	Gly	Ala	Arg	Pro	Tyr	Thr	Gly	Lys	Val	Gly	Ala	Glu	Asp	Ala
		275					280					285			
Asp	Gly	Ile	Asp	Met	Ala	Tyr	Arg	Val	Leu	Ala	Asp	His	Ala	Arg	Thr
	290					295					300				
Ile	Thr	Val	Ala	Leu	Ala	Asp	Gly	Gly	Arg	Pro	Asp	Asn	Thr	Gly	Arg
305					310					315					320
Gly	Tyr	Val	Leu	Arg	Arg	Ile	Leu	Arg	Arg	Ala	Val	Arg	Tyr	Ala	His
				325					330					335	
Glu	Lys	Leu	Asn	Ala	Ser	Arg	Gly	Phe	Phe	Ala	Thr	Leu	Val	Asp	Val
			340					345					350		
Val	Val	Gln	Ser	Leu	Gly	Asp	Ala	Phe	Pro	Glu	Leu	Lys	Lys	Asp	Pro
		355					360					365			
Asp	Met	Val	Lys	Asp	Ile	Ile	Asn	Glu	Glu	Glu	Val	Gln	Phe	Leu	Lys
	370					375					380				
Thr	Leu	Ser	Arg	Gly	Arg	Arg	Ile	Leu	Asp	Arg	Lys	Ile	Gln	Ser	Leu
385					390					395					400
Gly	Asp	Ser	Lys	Thr	Ile	Pro	Gly	Asp	Thr	Ala	Trp	Leu	Leu	Tyr	Asp
				405					410					415	
Thr	Tyr	Gly	Phe	Pro	Val	Asp	Leu	Thr	Gly	Leu	Ile	Ala	Glu	Glu	Lys
			420					425					430		
Gly	Leu	Val	Val	Asp	Met	Asp	Gly	Phe	Glu	Glu	Glu	Arg	Lys	Leu	Ala
		435					440					445			
Gln	Leu	Lys	Ser	Gln	Gly	Lys	Gly	Ala	Gly	Gly	Glu	Asp	Leu	Ile	Met
	450					455					460				
Leu	Asp	Ile	Tyr	Ala	Ile	Glu	Glu	Leu	Arg	Ala	Arg	Gly	Leu	Glu	Val
465					470					475					480
Thr	Asp	Asp	Ser	Pro	Lys	Tyr	Asn	Tyr	His	Leu	Asp	Ser	Ser	Gly	Ser
				485					490					495	
Tyr	Val	Phe	Glu	Asn	Thr	Val	Ala	Thr	Val	Met	Ala	Leu	Arg	Arg	Glu
			500					505					510		
Lys	Met	Phe	Val	Glu	Glu	Val	Ser	Thr	Gly	Gln	Glu	Cys	Gly	Val	Val
		515					520					525			
Leu	Asp	Lys	Thr	Cys	Phe	Tyr	Ala	Glu	Gln	Gly	Gly	Gln	Ile	Tyr	Asp
	530					535					540				
Glu	Gly	Tyr	Leu	Val	Lys	Val	Asp	Asp	Ser	Ser	Glu	Asp	Lys	Thr	Glu
545					550					555					560
Phe	Thr	Val	Lys	Asn	Ala	Gln	Val	Arg	Gly	Gly	Tyr	Val	Leu	His	Ile
				565					570					575	
Gly	Thr	Ile	Tyr	Gly	Asp	Leu	Lys	Val	Gly	Asp	Gln	Val	Trp	Leu	Phe
			580					585					590		
Ile	Asp	Glu	Pro	Arg	Arg	Arg	Pro	Ile	Met	Ser	Asn	His	Thr	Ala	Thr
	595						600					605			
His	Ile	Leu	Asn	Phe	Ala	Leu	Arg	Ser	Val	Leu	Gly	Glu	Ala	Asp	Gln
	610					615					620				
Lys	Gly	Ser	Leu	Val	Ala	Pro	Asp	Arg	Leu	Arg	Phe	Asp	Phe	Thr	Ala
625					630					635					640
Lys	Gly	Ala	Met	Ser	Thr	Gln	Gln	Ile	Lys	Lys	Ala	Glu	Glu	Ile	Ala
				645					650					655	
Asn	Glu	Met	Ile	Glu	Ala	Ala	Lys	Ala	Val	Tyr	Thr	Gln	Asp	Cys	Pro
			660					665					670		

120161\_426PC\_Sequence\_Listing.TXT

Leu Ala Ala Ala Lys Ala Ile Gl n Gly Leu Arg Ala Val Phe Asp Gl u  
 675 680 685  
 Thr Tyr Pro Asp Pro Val Arg Val Val Ser Ile Gly Val Pro Val Ser  
 690 695 700  
 Gl u Leu Leu Asp Asp Pro Ser Gly Pro Ala Gly Ser Leu Thr Ser Val  
 705 710 715 720  
 Gl u Phe Cys Gly Gly Thr Hi s Leu Arg Asn Ser Ser Hi s Ala Gly Ala  
 725 730 735  
 Phe Val Ile Val Thr Gl u Gl u Ala Ile Ala Lys Gly Ile Arg Arg Ile  
 740 745 750  
 Val Ala Val Thr  
 755

<210> 15  
 <211> 2268  
 <212> DNA  
 <213> Homo sapi ens

<400> 15  
 atggactcta ctctaacagc aagtgaatc cggcagcgat ttatagattt cttcaagagg 60  
 aacgagcata cgtatgttca ctcgctcgcc accatcccat tggatgaccc cactttgctc 120  
 ttgccaatg caggcatgaa ccagtttaaa cccattttcc tgaacacaat tgacccatct 180  
 caccatcatg caaagctgag cagagctgcc aataccaga agtgcacccg ggctgggggc 240  
 aaacataatg acctggacga tgtgggcaag gatgtctatc atcacacctt cttcgagatg 300  
 ctgggctctt ggtcttttgg agattacttt aaggaattgg catgtaagat ggctctggaa 360  
 ctctcaccac aagagtttgg cattccattt gaaagacttt atgttactta ctttggcggg 420  
 gatgaagcag ctggcttaga agcagatctg gaatgcaaac agatctggca aaatttgggg 480  
 ctggatgaca ccaaaatcct cccaggcaac atgaaggata acttctggga gatgggtgac 540  
 acgggcccct gtggtccttg cagtgaatc cactacgacc ggattgggtg tggggacgcc 600  
 gcacatcttg tcaaccagga cgacccta atgtctggaga tctggaacct tgtgttcac 660  
 cagtataaca gggaagctga tggcattctg aaacctcttc ccaagaaaag cattgacaca 720  
 gggatggggc tggaaagact ggtatctgtg ctgcagaata agatgtccaa ctatgacact 780  
 gaccttttgg tcccttactt tgaagccatt cagaagggca cagggtgccg accatacact 840  
 gggaaagtgt gtgctgagga tgccgatggg attgacatgg cctaccgggt gctggctgac 900  
 cagctcggga ccatcactgt ggcactggct gatggtggcc ggcctgacaa cacagggcgt 960  
 ggaatgtgtg tgagacggat tctccgccga gctgtccgat acgcccataa aaagctcaat 1020  
 gccagcaggg gcttctttgc tacgttagtg gatgttgtcg tccagtccct gggagatgca 1080  
 tttctgagc tgaagaagga cccagacatg gtgaaggaca tcattaatga agaagaggtg 1140  
 cagtttctca agactctcag cagagggcgt cgcacctctg acaggaaaat tcagagcctg 1200  
 ggagacagca agaccattcc cggagacact gcttggctcc tctatgacac ctatgggttt 1260  
 ccagtggatc tgactggact gattgctgaa gagaagggcc tgggtgtaga catggaatggc 1320  
 ttgtaagagg agaggaaact ggcccagctg aaatcacagg gcaagggagc tgggtggggaa 1380  
 gacctatta tgcaggacat ttacgctatc gaagagctcc gggcacgggg tctggaggtc 1440  
 acagatgatt ccccaaagta caattacat tggactcca gtggtagcta tgtatttgag 1500  
 aacacagtgg ctacggtgat ggctctgctc agggagaaga tgttctgtga agaggtgtcc 1560  
 acaggccagg agtgtggagt ggtgctggac aagacctgtt tctatgctga gcaaggaggc 1620  
 cagatctatg acgaaggcta cctggtgaag gtggatgaca gcagtgaaga taaaacagag 1680  
 ttacagtga agaagtcta ggtccgagga gggtatgtgc tacacatttg aacctctac 1740  
 ggtgacctga aagtggggga tcaggtctgg ctgtttattg atgagccccg acgaagacct 1800  
 atcatgagca accacacagc tacgcacatt ctgaacttcg ccctgcgctc agtgcttggg 1860  
 gaagctgacc agaaaggctc attggttgct cctgaccgcc tcagatttga ctttactgcc 1920  
 aaggagacca tgtccaccca acagatcaag aaggctgaag agattgctaa tgagatgatt 1980  
 gaggcagcca aggccgtcta taccaggat tgccccctgg cagcagcgaa agccatccag 2040  
 ggcctacggg ctgtgtttga tgagacctat cctgaccctg tgcgagtcgt ctccattggg 2100  
 gtccccgtgt ccgagttgct ggatgacccc tctgggcctg ctggctccct gacttctgtt 2160  
 gatttctgtg ggggaacgca cctgcggaac tcgagtcatg caggagcttt tgtgatcgtg 2220  
 acggaagaag ccattgcca gggtatccgg aggtattgtg ctgtcaca 2268

<210> 16  
 <211> 21  
 <212> PRT  
 <213> Mus muscul us

<400> 16  
 Leu Tyr Val Thr Tyr Phe Gly Gly Asp Gl u Ala Ala Gly Leu Gl u Pro  
 1 5 10 15  
 Asp Leu Gl u Cys Arg

<210> 17  
 <211> 150  
 <212> PRT  
 <213> Mus musculus

<400> 17  
 Gln Ile Trp Gln Asn Leu Gly Leu Asp Glu Ala Arg Ile Leu Pro Gly  
 1 5 10 15  
 Asn Met Lys Asp Asn Phe Trp Glu Met Gly Asp Thr Gly Pro Cys Gly  
 20 25 30  
 Pro Cys Ser Glu Ile His Tyr Asp Arg Ile Gly Gly Arg Asp Ala Ala  
 35 40 45  
 His Leu Val Asn Gln Asp Asp Pro Asn Val Leu Glu Ile Trp Asn Leu  
 50 55 60  
 Val Phe Ile Gln Tyr Asn Arg Glu Ser Asp Gly Val Leu Lys Pro Leu  
 65 70 75 80  
 Pro Lys Lys Ser Ile Asp Thr Gly Met Gly Leu Glu Arg Leu Val Ser  
 85 90 95  
 Val Leu Gln Asn Lys Met Ser Asn Tyr Asp Thr Asp Leu Phe Met Pro  
 100 105 110  
 Tyr Phe Glu Ala Ile Gln Lys Gly Thr Gly Ala Arg Pro Tyr Thr Gly  
 115 120 125  
 Lys Val Gly Ala Glu Asp Ala Asp Gly Ile Asp Met Ala Tyr Arg Val  
 130 135 140  
 Leu Ala Asp His Ala Arg  
 145 150

<210> 18  
 <211> 17  
 <212> PRT  
 <213> Mus musculus

<400> 18  
 Thr Ile Thr Val Ala Leu Ala Asp Gly Gly Arg Pro Asp Asn Thr Gly  
 1 5 10 15  
 Arg

<210> 19  
 <211> 52  
 <212> PRT  
 <213> Mus musculus

<400> 19  
 Gly Tyr Val Leu Arg Arg Ile Leu Arg Arg Ala Val Arg Tyr Ser His  
 1 5 10 15  
 Glu Lys Leu Asn Ala Ser Arg Gly Phe Phe Ala Thr Leu Val Asp Val  
 20 25 30  
 Val Val Gln Ser Leu Gly Asp Ala Phe Pro Glu Leu Lys Lys Asp Pro  
 35 40 45  
 Glu Met Val Lys  
 50

<210> 20  
 <211> 12  
 <212> PRT  
 <213> Mus musculus

<400> 20  
 Asp Ile Ile Asn Glu Glu Glu Val Gln Phe Leu Lys  
 1 5 10

120161\_426PC\_Sequence\_Listing.TXT

<210> 21  
 <211> 252  
 <212> PRT  
 <213> Mus musculus

<400> 21  
 Leu Tyr Val Thr Tyr Phe Gly Gly Asp Glu Ala Ala Gly Leu Glu Pro  
 1 5 10 15  
 Asp Leu Glu Cys Arg Gln Ile Trp Gln Asn Leu Gly Leu Asp Glu Ala  
 20 25 30  
 Arg Ile Leu Pro Gly Asn Met Lys Asp Asn Phe Trp Glu Met Gly Asp  
 35 40 45  
 Thr Gly Pro Cys Gly Pro Cys Ser Glu Ile His Tyr Asp Arg Ile Gly  
 50 55 60  
 Gly Arg Asp Ala Ala His Leu Val Asn Gln Asp Asp Pro Asn Val Leu  
 65 70 75 80  
 Glu Ile Trp Asn Leu Val Phe Ile Gln Tyr Asn Arg Glu Ser Asp Gly  
 85 90 95  
 Val Leu Lys Pro Leu Pro Lys Lys Ser Ile Asp Thr Gly Met Gly Leu  
 100 105 110  
 Glu Arg Leu Val Ser Val Leu Gln Asn Lys Met Ser Asn Tyr Asp Thr  
 115 120 125  
 Asp Leu Phe Met Pro Tyr Phe Glu Ala Ile Gln Lys Gly Thr Gly Ala  
 130 135 140  
 Arg Pro Tyr Thr Gly Lys Val Gly Ala Glu Asp Ala Asp Gly Ile Asp  
 145 150 155 160  
 Met Ala Tyr Arg Val Leu Ala Asp His Ala Arg Thr Ile Thr Val Ala  
 165 170 175  
 Leu Ala Asp Gly Gly Arg Pro Asp Asn Thr Gly Arg Gly Tyr Val Leu  
 180 185 190  
 Arg Arg Ile Leu Arg Arg Ala Val Arg Tyr Ser His Glu Lys Leu Asn  
 195 200 205  
 Ala Ser Arg Gly Phe Phe Ala Thr Leu Val Asp Val Val Val Gln Ser  
 210 215 220  
 Leu Gly Asp Ala Phe Pro Glu Leu Lys Lys Asp Pro Glu Met Val Lys  
 225 230 235 240  
 Asp Ile Ile Asn Glu Glu Glu Val Gln Phe Leu Lys  
 245 250

<210> 22  
 <211> 16  
 <212> PRT  
 <213> Mus musculus

<400> 22  
 Met Ala Leu Glu Leu Leu Thr Gln Glu Phe Gly Ile Pro Val Glu Arg  
 1 5 10 15

<210> 23  
 <211> 33  
 <212> PRT  
 <213> Mus musculus

<400> 23  
 Leu Tyr Val Thr Tyr Phe Gly Gly Asp Glu Ala Ala Gly Leu Glu Pro  
 1 5 10 15  
 Asp Leu Glu Cys Arg Gln Ile Trp Gln Asn Leu Gly Leu Asp Glu Ala  
 20 25 30  
 Arg

<210> 24  
 <211> 89  
 <212> PRT



120161\_426PC\_Sequence\_Listing.TXT

<213> Mus musculus

<400> 24

```

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

```

<210> 25

<211> 18

<212> PRT

<213> Mus musculus

<400> 25

```

Met Ser Asn Tyr Asp Thr Asp Leu Phe Met Pro Tyr Phe Glu Ala Ile
 1          5          10          15
Gln Lys

```

<210> 26

<211> 31

<212> PRT

<213> Mus musculus

<400> 26

```

Gly Thr Gly Ala Arg Pro Tyr Thr Gly Lys Val Gly Ala Glu Asp Ala
 1          5          10          15
Asp Gly Ile Asp Met Ala Tyr Arg Val Leu Ala Asp His Ala Arg
          20          25          30

```

<210> 27

<211> 17

<212> PRT

<213> Mus musculus

<400> 27

```

Thr Ile Thr Val Ala Leu Ala Asp Gly Gly Arg Pro Asp Asn Thr Gly
 1          5          10          15
Arg

```

<210> 28

<211> 52

<212> PRT

<213> Mus musculus

<400> 28

```

Gly Tyr Val Leu Arg Arg Ile Leu Arg Arg Ala Val Arg Tyr Ser His
 1          5          10          15
Glu Lys Leu Asn Ala Ser Arg Gly Phe Phe Ala Thr Leu Val Asp Val
          20          25          30
Val Val Gln Ser Leu Gly Asp Ala Phe Pro Glu Leu Lys Lys Asp Pro
          35          40          45
Glu Met Val Lys
          50

```

120161\_426PC\_Sequence\_Listing.TXT

<210> 29  
 <211> 12  
 <212> PRT  
 <213> Mus musculus

<400> 29  
 Asp Ile Ile Asn Glu Glu Glu Val Gln Phe Leu Lys  
 1 5 10

<210> 30  
 <211> 20  
 <212> PRT  
 <213> Mus musculus

<400> 30  
 Thr Leu Ser Arg Gly Arg Arg Ile Leu Asp Arg Lys Ile Gln Ser Leu  
 1 5 10 15  
 Gly Asp Cys Lys  
 20

<210> 31  
 <211> 28  
 <212> PRT  
 <213> Mus musculus

<400> 31  
 Thr Ile Pro Gly Asp Thr Ala Trp Leu Leu Tyr Asp Thr Tyr Gly Phe  
 1 5 10 15  
 Pro Val Asp Leu Thr Gly Leu Ile Ala Glu Glu Lys  
 20 25

<210> 32  
 <211> 23  
 <212> PRT  
 <213> Mus musculus

<400> 32  
 Gly Leu Val Val Asp Met Asn Gly Phe Glu Glu Glu Arg Arg Leu Ala  
 1 5 10 15  
 Gln Leu Lys Ser Gln Gly Lys  
 20

<210> 33  
 <211> 19  
 <212> PRT  
 <213> Mus musculus

<400> 33  
 Gly Ala Gly Asp Glu Asp Leu Ile Met Leu Asp Ile Tyr Ala Ile Glu  
 1 5 10 15  
 Glu Leu Arg

<210> 34  
 <211> 39  
 <212> PRT  
 <213> Mus musculus

<400> 34  
 Ala Lys Gly Leu Glu Ala Thr Asp Asp Ser Pro Lys Tyr Asn Tyr Gln  
 1 5 10 15  
 Ser Asp Ser Ser Gly Ser Tyr Val Phe Glu Cys Thr Val Ala Thr Val  
 Page 10

120161\_426PC\_Sequence\_Listing.TXT

Leu Ala Leu Arg Arg Glu Lys  
20 25 30  
35

<210> 35  
<211> 18  
<212> PRT  
<213> Mus musculus

<400> 35  
Met Phe Val Asp Glu Val Val Thr Gly Gln Glu Cys Gly Val Val Leu  
1 5 10 15  
Asp Lys

<210> 36  
<211> 19  
<212> PRT  
<213> Mus musculus

<400> 36  
Thr Cys Phe Tyr Ala Glu Gln Gly Gly Gln Ile Tyr Asp Glu Gly Tyr  
1 5 10 15  
Leu Val Lys

<210> 37  
<211> 14  
<212> PRT  
<213> Mus musculus

<400> 37  
Val Asp Asp Ser Ser Glu Asp Lys Thr Glu Phe Thr Val Lys  
1 5 10

<210> 38  
<211> 86  
<212> PRT  
<213> Mus musculus

<400> 38  
Asn Ala Gln Val Arg Gly Gly Tyr Val Leu His Ile Gly Thr Ile Tyr  
1 5 10 15  
Gly Asn Leu Lys Val Gly Asp Gln Val Arg Leu Phe Ile Asp Glu Pro  
20 25 30  
Arg Arg Arg Pro Val Met Ser Asn His Thr Ala Thr His Ile Leu Asn  
35 40 45  
Phe Ala Leu Arg Ser Val Leu Gly Glu Ala Asp Gln Lys Gly Ser Leu  
50 55 60  
Val Ala Pro Asp Arg Leu Arg Phe Asp Phe Thr Ala Lys Gly Ala Met  
65 70 75 80  
Ser Thr Gln Gln Ile Lys  
85

<210> 39  
<211> 27  
<212> PRT  
<213> Mus musculus

<400> 39  
Lys Ala Glu Glu Ile Val Asn Gly Met Ile Glu Ala Ala Lys Pro Val  
1 5 10 15  
Tyr Thr Gln Asp Cys Pro Leu Ala Ala Lys

<210> 40  
 <211> 561  
 <212> PRT  
 <213> Mus musculus

<400> 40

Met	Ala	Leu	Glu	Leu	Leu	Thr	Gln	Glu	Phe	Gly	Ile	Pro	Val	Glu	Arg
1				5					10					15	
Leu	Tyr	Val	Thr	Tyr	Phe	Gly	Gly	Asp	Glu	Ala	Ala	Gly	Leu	Glu	Pro
			20					25					30		
Asp	Leu	Glu	Cys	Arg	Gln	Ile	Trp	Gln	Asn	Leu	Gly	Leu	Asp	Glu	Ala
		35					40					45			
Arg	Ile	Leu	Pro	Gly	Asn	Met	Lys	Asp	Asn	Phe	Trp	Glu	Met	Gly	Asp
	50					55					60				
Thr	Gly	Pro	Cys	Gly	Pro	Cys	Ser	Glu	Ile	His	Tyr	Asp	Arg	Ile	Gly
65					70					75					80
Gly	Arg	Asp	Ala	Ala	His	Leu	Val	Asn	Gln	Asp	Asp	Pro	Asn	Val	Leu
				85					90					95	
Glu	Ile	Trp	Asn	Leu	Val	Phe	Ile	Gln	Tyr	Asn	Arg	Glu	Ser	Asp	Gly
			100					105					110		
Val	Leu	Lys	Pro	Leu	Pro	Lys	Lys	Ser	Ile	Asp	Thr	Gly	Met	Gly	Leu
		115					120					125			
Glu	Arg	Leu	Val	Ser	Val	Leu	Gln	Asn	Lys	Met	Ser	Asn	Tyr	Asp	Thr
	130					135					140				
Asp	Leu	Phe	Met	Pro	Tyr	Phe	Glu	Ala	Ile	Gln	Lys	Gly	Thr	Gly	Ala
145					150					155					160
Arg	Pro	Tyr	Thr	Gly	Lys	Val	Gly	Ala	Glu	Asp	Ala	Asp	Gly	Ile	Asp
				165					170					175	
Met	Ala	Tyr	Arg	Val	Leu	Ala	Asp	His	Ala	Arg	Thr	Ile	Thr	Val	Ala
			180					185						190	
Leu	Ala	Asp	Gly	Gly	Arg	Pro	Asp	Asn	Thr	Gly	Arg	Gly	Tyr	Val	Leu
		195					200					205			
Arg	Arg	Ile	Leu	Arg	Arg	Ala	Val	Arg	Tyr	Ser	His	Glu	Lys	Leu	Asn
	210					215					220				
Ala	Ser	Arg	Gly	Phe	Phe	Ala	Thr	Leu	Val	Asp	Val	Val	Val	Gln	Ser
225					230					235					240
Leu	Gly	Asp	Ala	Phe	Pro	Glu	Leu	Lys	Lys	Asp	Pro	Glu	Met	Val	Lys
				245					250					255	
Asp	Ile	Ile	Asn	Glu	Glu	Glu	Val	Gln	Phe	Leu	Lys	Thr	Leu	Ser	Arg
			260					265					270		
Gly	Arg	Arg	Ile	Leu	Asp	Arg	Lys	Ile	Gln	Ser	Leu	Gly	Asp	Cys	Lys
		275					280					285			
Thr	Ile	Pro	Gly	Asp	Thr	Ala	Trp	Leu	Leu	Tyr	Asp	Thr	Tyr	Gly	Phe
	290					295					300				
Pro	Val	Asp	Leu	Thr	Gly	Leu	Ile	Ala	Glu	Glu	Lys	Gly	Leu	Val	Val
305					310					315					320
Asp	Met	Asn	Gly	Phe	Glu	Glu	Glu	Arg	Arg	Leu	Ala	Gln	Leu	Lys	Ser
				325					330					335	
Gln	Gly	Lys	Gly	Ala	Gly	Asp	Glu	Asp	Leu	Ile	Met	Leu	Asp	Ile	Tyr
			340					345					350		
Ala	Ile	Glu	Glu	Leu	Arg	Ala	Lys	Gly	Leu	Glu	Ala	Thr	Asp	Asp	Ser
		355					360					365			
Pro	Lys	Tyr	Asn	Tyr	Gln	Ser	Asp	Ser	Ser	Gly	Ser	Tyr	Val	Phe	Glu
	370					375					380				
Cys	Thr	Val	Ala	Thr	Val	Leu	Ala	Leu	Arg	Arg	Glu	Lys	Met	Phe	Val
385					390					395					400
Asp	Glu	Val	Val	Thr	Gly	Gln	Glu	Cys	Gly	Val	Val	Leu	Asp	Lys	Thr
				405					410					415	
Cys	Phe	Tyr	Ala	Glu	Gln	Gly	Gly	Gln	Ile	Tyr	Asp	Glu	Gly	Tyr	Leu
			420					425					430		
Val	Lys	Val	Asp	Asp	Ser	Ser	Glu	Asp	Lys	Thr	Glu	Phe	Thr	Val	Lys
		435					440					445			
Asn	Ala	Gln	Val	Arg	Gly	Gly	Tyr	Val	Leu	His	Ile	Gly	Thr	Ile	Tyr
	450					455					460				
Gly	Asn	Leu	Lys	Val	Gly	Asp	Gln	Val	Arg	Leu	Phe	Ile	Asp	Glu	Pro

120161\_426PC\_Sequence\_Listing.TXT

```

465      470      475      480
Arg Arg Arg Pro Val Met Ser Asn His Thr Ala Thr His Ile Leu Asn
      485      490      495
Phe Ala Leu Arg Ser Val Leu Gly Glu Ala Asp Gln Lys Gly Ser Leu
      500      505      510
Val Ala Pro Asp Arg Leu Arg Phe Asp Phe Thr Ala Lys Gly Ala Met
      515      520      525
Ser Thr Gln Gln Ile Lys Lys Ala Glu Glu Ile Val Asn Gly Met Ile
      530      535      540
Glu Ala Ala Lys Pro Val Tyr Thr Gln Asp Cys Pro Leu Ala Ala Ala
545      550      555      560
Lys

```

<210> 41  
 <211> 275  
 <212> PRT  
 <213> Homo sapiens

```

<400> 41
Met Asp Ser Thr Leu Thr Ala Ser Glu Ile Arg Gln Arg Phe Ile Asp
 1      5      10      15
Phe Phe Lys Arg Asn Glu His Thr Tyr Val His Ser Ser Ala Thr Ile
      20      25      30
Pro Leu Asp Asp Pro Thr Leu Leu Phe Ala Asn Ala Gly Met Asn Gln
      35      40      45
Phe Lys Pro Ile Phe Leu Asn Thr Ile Asp Pro Ser His Pro Met Ala
      50      55      60
Lys Leu Ser Arg Ala Ala Asn Thr Gln Lys Cys Ile Arg Ala Gly Gly
65      70      75      80
Lys His Asn Asp Leu Asp Asp Val Gly Lys Asp Val Tyr His His Thr
      85      90      95
Phe Phe Glu Met Leu Gly Ser Trp Ser Phe Gly Asp Tyr Phe Lys Glu
      100      105      110
Leu Ala Cys Lys Met Ala Leu Glu Leu Leu Thr Gln Glu Phe Gly Ile
      115      120      125
Pro Ile Glu Arg Leu Tyr Val Thr Tyr Phe Gly Gly Asp Glu Ala Ala
      130      135      140
Gly Leu Glu Ala Asp Leu Glu Cys Lys Gln Ile Trp Gln Asn Leu Gly
145      150      155      160
Leu Asp Asp Thr Lys Ile Leu Pro Gly Asn Met Lys Asp Asn Phe Trp
      165      170      175
Glu Met Gly Asp Thr Gly Pro Cys Gly Pro Cys Ser Glu Ile His Tyr
      180      185      190
Asp Arg Ile Gly Gly Arg Asp Ala Ala His Leu Val Asn Gln Asp Asp
      195      200      205
Pro Asn Val Leu Glu Ile Trp Asn Leu Val Phe Ile Gln Tyr Asn Arg
      210      215      220
Ala Gln Val Pro Asp His Thr Leu Gly Lys Leu Val Leu Arg Met Pro
225      230      235      240
Met Gly Leu Thr Trp Pro Thr Gly Cys Trp Leu Thr Thr Leu Gly Pro
      245      250      255
Ser Leu Trp His Trp Leu Met Val Ala Gly Leu Thr Thr Gln Gly Val
      260      265      270
Asp Met Cys
      275

```

<210> 42  
 <211> 828  
 <212> DNA  
 <213> Homo sapiens

```

<400> 42
atggactcta ctctaacagc aagtgaatc cggcagcgat ttatagattt cttcaagagg 60
aacgagcata cgatatgtca ctcgtctgcc accatcccat tggatgaccc cactttgctc 120
tttgccaatg caggcatgaa ccagtttaaa cccattttcc tgaacacaat tgacctatct 180

```

120161\_426PC\_Sequence\_Listing.TXT

```

caccatg ccaagctg cagagctg aataccaga agtgcaccc ggctgggggc 240
aaacataatg acctggacga tgtgggcaag gatgtctatc atcacacctt cttcgagatg 300
ctgggctctt ggtcttttgg agattacttt aaggaattgg catgtaagat ggctctggaa 360
ctcctcacc ccaagagtttgg cattccatt gaaagacttt atgttactta ctttggcggg 420
gatgaagcag ctggcttaga agcagatctg gaatgcaaac agatctggca aaatttgggg 480
ctggatgaca ccaaaatcct cccaggcaac atgaaggata acttctggga gatgggtgac 540
acgggcccct gtggtccttg cagtgaatc cactacgacc ggattgggtg tcgggacgcc 600
gcacatcttg tcaaccagga cgacccta atgtctggaga tctggaacct tgtgttcac 660
cagtataaca gggcacagg gcccaccat acactgggaa agttgggtgct gaggatgccg 720
atgggattga catggcctac cgggtgctgg ctgaccacgc tcggaccatc actgtggcac 780
tggctgatgg tggccggcct gacaacacag ggcgtggata tgtgttga 828

```

<210> 43  
 <211> 521  
 <212> PRT  
 <213> Homo sapiens

```

<400> 43
Met Asp Ser Thr Leu Thr Ala Ser Glu Ile Arg Glu Arg Phe Ile Asp
1      5      10      15
Phe Phe Lys Arg Asn Glu His Thr Tyr Val His Ser Ser Ala Thr Ile
20     25     30
Pro Leu Asp Asp Pro Thr Leu Leu Phe Ala Asn Ala Gly Met Asn Glu
35     40     45
Phe Lys Pro Ile Phe Leu Asn Thr Ile Asp Pro Ser His Pro Met Ala
50     55     60
Lys Leu Ser Arg Ala Ala Asn Thr Glu Lys Cys Ile Arg Ala Gly Gly
65     70     75     80
Lys His Asn Asp Leu Asp Asp Val Gly Lys Asp Val Tyr His His Thr
85     90     95
Phe Phe Glu Met Leu Gly Ser Trp Ser Phe Gly Asp Tyr Phe Lys Glu
100    105    110
Leu Ala Cys Lys Met Ala Leu Glu Leu Leu Thr Glu Glu Phe Gly Ile
115    120    125
Pro Ile Glu Arg Leu Tyr Val Thr Tyr Phe Gly Gly Asp Glu Ala Ala
130    135    140
Gly Leu Glu Ala Asp Leu Glu Cys Lys Glu Ile Trp Glu Asn Leu Gly
145    150    155    160
Leu Asp Asp Thr Lys Ile Leu Pro Gly Asn Met Lys Asp Asn Phe Trp
165    170    175
Glu Met Gly Asp Thr Gly Pro Cys Gly Pro Cys Ser Glu Ile His Tyr
180    185    190
Asp Arg Ile Gly Gly Arg Asp Ala Ala His Leu Val Asn Glu Asp Asp
195    200    205
Pro Asn Val Leu Glu Ile Trp Asn Leu Val Phe Ile Glu Tyr Asn Arg
210    215    220
Glu Ala Asp Gly Ile Leu Lys Pro Leu Pro Lys Lys Ser Ile Asp Thr
225    230    235    240
Gly Met Gly Leu Glu Arg Leu Val Ser Val Leu Glu Asn Lys Met Ser
245    250    255
Asn Tyr Asp Thr Asp Leu Phe Val Pro Tyr Phe Glu Ala Ile Glu Lys
260    265    270
Gly Thr Gly Ala Arg Pro Tyr Thr Gly Lys Val Gly Ala Glu Asp Ala
275    280    285
Asp Gly Ile Asp Met Ala Tyr Arg Val Leu Ala Asp His Ala Arg Thr
290    295    300
Ile Thr Val Ala Leu Ala Asp Gly Gly Arg Pro Asp Asn Thr Gly Arg
305    310    315    320
Gly Tyr Val Leu Arg Arg Ile Leu Arg Arg Ala Val Arg Tyr Ala His
325    330    335
Glu Lys Leu Asn Ala Ser Arg Gly Phe Phe Ala Thr Leu Val Asp Val
340    345    350
Val Val Glu Ser Leu Gly Asp Ala Phe Pro Glu Leu Lys Lys Asp Pro
355    360    365
Asp Met Val Lys Asp Ile Ile Asn Glu Glu Glu Val Glu Phe Leu Lys
370    375    380
Thr Leu Ser Arg Gly Arg Arg Ile Leu Asp Arg Lys Ile Glu Ser Leu
385    390    395    400

```

120161\_426PC\_Sequence\_Listing.TXT

Gly	Asp	Ser	Lys	Thr	Ile	Pro	Gly	Asp	Thr	Ala	Trp	Leu	Leu	Tyr	Asp
				405					410					415	
Thr	Tyr	Gly	Phe	Pro	Val	Asp	Leu	Thr	Gly	Leu	Ile	Ala	Glu	Glu	Lys
			420					425					430		
Gly	Leu	Val	Val	Asp	Met	Asp	Gly	Phe	Glu	Glu	Glu	Arg	Lys	Leu	Ala
		435					440					445			
Gln	Leu	Lys	Ser	Gln	Gly	Lys	Gly	Ala	Gly	Gly	Glu	Asp	Leu	Ile	Met
		450				455					460				
Leu	Asp	Ile	Tyr	Ala	Ile	Glu	Glu	Leu	Arg	Ala	Arg	Gly	Leu	Glu	Val
465					470					475					480
Thr	Asp	Asp	Ser	Pro	Lys	Tyr	Asn	Tyr	His	Leu	Asp	Ser	Ser	Gly	Ser
				485					490					495	
Tyr	Glu	Asn	Arg	Val	Tyr	Ser	Glu	Glu	Cys	Ser	Gly	Pro	Arg	Arg	Val
			500					505					510		
Cys	Ala	Thr	His	Trp	Asn	His	Leu	Arg							
		515					520								

<210> 44  
 <211> 1566  
 <212> DNA  
 <213> Homo sapiens

<400> 44

atggactcta	ctctaacagc	aagtgaatc	cggcagcgat	ttatagattt	cttcaagagg	60
aacgagcata	cgatatgtca	ctcgtctgcc	accatcccat	tggatgacct	cactttgctc	120
tttgccaatg	caggcatgaa	ccagtttaaa	cccattttcc	tgaacacaa	tgacctatct	180
caccccatgg	caaagctgag	cagagctgcc	aatacccaga	agtgcattcc	ggctgggggc	240
aaacataatg	acctggacga	tgtgggcaag	gatgtctatc	atcacacctt	cttcgagatg	300
ctggggctctt	ggtcttttgg	agattacttt	aaggaattgg	catgtaagat	ggctctggaa	360
ctccctaccc	aagagtttgg	cattcccatt	gaaagacttt	atgttactta	ctttggcggg	420
gatgaagcag	ctggcttaga	agcagatctg	gaatgcaaac	agatctggca	aaatttgggg	480
ctggaatgaca	ccaaaatcct	cccaggcaac	atgaaggata	acttctggga	gatgggtgac	540
acggggcccct	gtggctcctg	cagtgaatc	cactacgacc	ggattgggtg	tcgggacgcc	600
gcacatcttg	tcaaccagga	cgacccta	gtgctggaga	tctggaacct	tgtgttcac	660
cagtataaca	gggaagctga	tggcattctg	aaacctcttc	ccaagaaaag	cattgacaca	720
gggatggggc	tgaacgact	ggtatctgtg	ctgcagaata	agatgtccaa	ctatgacact	780
gacctttttg	tcccttactt	tgaagccatt	cagaaggcca	caggtgccc	accatacact	840
gggaaagtgt	gtgctgagga	tgccgatggg	attgacatgg	cctaccgggt	gctggctgac	900
cacgctcgga	ccatcactgt	ggcactggct	gatggtggcc	ggcctgacaa	cacagggcgt	960
ggatatgtgt	tgagacggat	tctccgccga	gctgtccgat	acgcccatga	aaagctcaat	1020
gccagcaggg	gcttctttgc	tacgttagtg	gatgttgtcg	tccagtcctt	gggagatgca	1080
tttcttgagc	tgaagaagga	cccagacatg	gtgaaggaca	tcattaatga	agaagagggtg	1140
cagtttctca	agactctcag	cagaggcgct	cgcattcctg	acaggaaaat	tcagagcctg	1200
ggagacagca	agaccattcc	cggagacact	gcttggctcc	tctatgacac	ctatgggttt	1260
ccagtggatc	tgactggact	gattgctgaa	gagaagggcc	tggtggtaga	catggatggc	1320
tttgaagagg	agaggaaact	ggcccagctg	aaatcacagg	gcaagggagc	tggtggggaa	1380
gacctcatta	tgctggacat	ttacgctatc	gaagagctcc	gggcacgggg	tctggaggtc	1440
acagatgatt	cccaaagta	caattaccat	ttggactcca	gtggtagcta	tgaaaacaga	1500
gtttacagtg	aagaatgctc	aggtccgagg	agggatgtgt	ctacacattg	gaaccatcta	1560
cgggtga						1566

<210> 45  
 <211> 665  
 <212> PRT  
 <213> Homo sapiens

<400> 45

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

120161\_426PC\_Sequence\_Listing.TXT

Lys	His	Asn	Asp	Leu	Asp	Asp	Val	Gly	Lys	Asp	Val	Tyr	His	His	Thr
				85					90					95	
Phe	Phe	Glu	Met	Leu	Gly	Ser	Trp	Ser	Phe	Gly	Asp	Tyr	Phe	Lys	Glu
			100					105					110		
Leu	Ala	Cys	Lys	Met	Ala	Leu	Glu	Leu	Leu	Thr	Gln	Glu	Phe	Gly	Ile
		115					120					125			
Pro	Ile	Glu	Arg	Leu	Tyr	Val	Thr	Tyr	Phe	Gly	Gly	Asp	Glu	Ala	Ala
	130					135					140				
Gly	Leu	Glu	Ala	Asp	Leu	Glu	Cys	Lys	Gln	Ile	Trp	Gln	Asn	Leu	Gly
	145				150					155					160
Leu	Asp	Asp	Thr	Lys	Ile	Leu	Pro	Gly	Asn	Met	Lys	Asp	Asn	Phe	Trp
				165					170					175	
Glu	Met	Gly	Asp	Thr	Gly	Pro	Cys	Gly	Pro	Cys	Ser	Glu	Ile	His	Tyr
			180					185					190		
Asp	Arg	Ile	Gly	Gly	Arg	Asp	Ala	Ala	His	Leu	Val	Asn	Gln	Asp	Asp
		195					200					205			
Pro	Asn	Val	Leu	Glu	Ile	Trp	Asn	Leu	Val	Phe	Ile	Gln	Tyr	Asn	Arg
	210					215					220				
Glu	Ala	Asp	Gly	Ile	Leu	Lys	Pro	Leu	Pro	Lys	Lys	Ser	Ile	Asp	Thr
	225				230					235					240
Gly	Met	Gly	Leu	Glu	Arg	Leu	Val	Ser	Val	Leu	Gln	Asn	Lys	Met	Ser
				245					250					255	
Asn	Tyr	Asp	Thr	Asp	Leu	Phe	Val	Pro	Tyr	Phe	Glu	Ala	Ile	Gln	Lys
			260					265					270		
Gly	Thr	Gly	Ala	Arg	Pro	Tyr	Thr	Gly	Lys	Val	Gly	Ala	Glu	Asp	Ala
			275				280					285			
Asp	Gly	Ile	Asp	Met	Ala	Tyr	Arg	Val	Leu	Ala	Asp	His	Ala	Arg	Thr
	290					295					300				
Ile	Thr	Val	Ala	Leu	Ala	Asp	Gly	Gly	Arg	Pro	Asp	Asn	Thr	Gly	Arg
	305				310					315					320
Gly	Tyr	Val	Leu	Arg	Arg	Ile	Leu	Arg	Arg	Ala	Val	Arg	Tyr	Ala	His
				325					330					335	
Glu	Lys	Leu	Asn	Ala	Ser	Arg	Gly	Phe	Phe	Ala	Thr	Leu	Val	Asp	Val
			340					345					350		
Val	Val	Gln	Ser	Leu	Gly	Asp	Ala	Phe	Pro	Glu	Leu	Lys	Lys	Asp	Pro
		355					360					365			
Asp	Met	Val	Lys	Asp	Ile	Ile	Asn	Glu	Glu	Glu	Val	Gln	Phe	Leu	Lys
	370					375					380				
Thr	Leu	Ser	Arg	Gly	Arg	Arg	Ile	Leu	Asp	Arg	Lys	Ile	Gln	Ser	Leu
	385				390					395					400
Gly	Asp	Ser	Lys	Thr	Ile	Pro	Gly	Asp	Thr	Ala	Trp	Leu	Leu	Tyr	Asp
				405					410					415	
Thr	Tyr	Gly	Phe	Pro	Val	Asp	Leu	Thr	Gly	Leu	Ile	Ala	Glu	Glu	Lys
			420					425					430		
Gly	Leu	Val	Val	Asp	Met	Asp	Gly	Phe	Glu	Glu	Glu	Arg	Lys	Leu	Ala
		435					440					445			
Gln	Leu	Lys	Ser	Gln	Gly	Lys	Gly	Ala	Gly	Gly	Glu	Asp	Leu	Ile	Met
	450					455					460				
Leu	Asp	Ile	Tyr	Ala	Ile	Glu	Glu	Leu	Arg	Ala	Arg	Gly	Leu	Glu	Val
	465				470					475					480
Thr	Asp	Asp	Ser	Pro	Lys	Tyr	Asn	Tyr	His	Leu	Asp	Ser	Ser	Gly	Ser
				485					490					495	
Tyr	Val	Phe	Glu	Asn	Thr	Val	Ala	Thr	Val	Met	Ala	Leu	Arg	Arg	Glu
			500					505					510		
Lys	Met	Phe	Val	Glu	Glu	Val	Ser	Thr	Gly	Gln	Glu	Cys	Gly	Val	Val
		515					520					525			
Leu	Asp	Lys	Thr	Cys	Phe	Tyr	Ala	Glu	Gln	Gly	Gly	Gln	Ile	Tyr	Asp
	530					535					540				
Glu	Gly	Tyr	Leu	Val	Lys	Val	Asp	Asp	Ser	Ser	Glu	Asp	Lys	Thr	Glu
	545				550					555					560
Phe	Thr	Val	Lys	Asn	Ala	Gln	Val	Arg	Gly	Gly	Tyr	Val	Leu	His	Ile
				565					570					575	
Gly	Thr	Ile	Tyr	Gly	Asp	Leu	Lys	Val	Gly	Asp	Gln	Val	Trp	Leu	Phe
			580					585					590		
Ile	Asp	Glu	Ala	Pro	Ala	Glu	Leu	Glu	Ser	Cys	Arg	Ser	Phe	Cys	Asp
		595					600					605			
Arg	Asp	Gly	Arg	Ser	His	Cys	Gln	Gly	Tyr	Pro	Glu	Asp	Cys	Gly	Cys
	610					615					620				



120161\_426PC\_Sequence\_Listing.TXT

His	Arg	Cys	Arg	Gly	Pro	Glu	Gly	Pro	Gln	Glu	Ser	Arg	Glu	Leu	Glu
625					630					635					640
Glu	Met	Ser	Leu	Cys	His	Gly	Ser	Gln	Ser	Glu	Gly	Ser	Asp	Cys	Ser
				645					650					655	
Lys	Gln	Gly	Cys	Ala	Glu	Gly	Asp	Arg							
			660					665							

<210> 46  
 <211> 1998  
 <212> DNA  
 <213> Homo sapiens

<400> 46  
 atggactcta ctctaacagc aagtgaatc cggcagcgat ttatagattt cttcaagagg 60  
 aacgagcata cgtatgttca ctctgtgcc accatcccat tggatgaccc cactttgctc 120  
 ttgccaatg caggcatgaa ccagtttaaa cccattttcc tgaacacaat tgacccatct 180  
 caccatcatg caaagctgag cagagctgcc aataccaga agtgcacccg ggctgggggc 240  
 aaacataatg acctggacga tgtgggcaag gatgtctatc atcacacctt cttcgagatg 300  
 ctgggctctt ggtcttttgg agattacttt aaggaattgg catgtaagat ggctctggaa 360  
 ctctcaccac aagagtttgg cattccattt gaaagacttt atgttactta ctttggcggg 420  
 gatgaagcag ctggcttaga agcagatctg gaatgcaaac agatctggca aaatttgggg 480  
 ctggatgaca ccaaaatcct cccaggcaac atgaaggata acttctggga gatgggtgac 540  
 acgggcccct gtggtccttg cagtgaatc cactacgacc ggattgggtg tctgggacgc 600  
 gcacatcttg tcaaccagga cgaccctaataa gtgctggaga tctggaacct tgtgttcatc 660  
 cagtataaca gggaagctga tggcattctg aaacctcttc ccaagaaaag cattgacaca 720  
 gggatggggc tggaaacgact ggtatctgtg ctgcagaata agatgtccaa ctatgacact 780  
 gacctttttg tcccttactt tgaagccatt cagaagggca cagggtgccg accatacact 840  
 gggaaagtgt gtgctgagga tgccgatggg attgacatgg cctaccgggt gctggctgac 900  
 cactgctcga ccactactgt ggcactgggt gatggtggcc ggcctgacaa cacagggcgt 960  
 ggatagtgtg tgagacggat tctccgccga gctgtccgat acgcccataa aaagctcaat 1020  
 gccagcaggg gcttctttgc tacgttagtg gatgtgtcgc tccagtccct gggagatgca 1080  
 tttcctgagc tgaagaagga cccagacatg gtgaaggaca tcattaatga agaagaggtg 1140  
 cagtttctca agactctcag cagagggcgt cgcacacctg acaggaaaat tcagagcctg 1200  
 ggagacagca agaccattcc cggagacact gcttggctcc tctatgacac ctatgggttt 1260  
 ccagtggatc tgactggact gattgctgaa gagaagggcc tggtagtaga catggatggc 1320  
 tttgaagagg agaggaaact ggcccagctg aaatcacagg gcaaggagagc tggtagggaa 1380  
 gacctcatta tgcctggacat ttacgctatc gaagagctcc gggcacgggg tctggaggtc 1440  
 acagatgatt ccccaaagta caattacatc ttggactcca gtgtagctta tgtatttgag 1500  
 aacacagtgg ctacggtgat ggctctgctc agggagaaga tgttctgtga agaggtgtcc 1560  
 acaggccagg agtgtggagt ggtgctggac aagacctgtt tctatgctga gcaaggaggc 1620  
 cagatctatg acgaaggcta cctggtgaag gtggatgaca gcagtgaaga taaaacagag 1680  
 ttacagtga agaattgctc ggtccgagga gggtatgtgc tacacatttg aacctctac 1740  
 ggtgacctga aagtggggga tcaggtctgg ctgtttattg atgaggcacc tgcggaactc 1800  
 gagtcatgca ggagcttttg tgatcgtgac ggaagaagcc attgccaagg gtatccggag 1860  
 gattgtggct gtcacaggtg ccgaggccca gaaggccctc aggaagcag agagcttgaa 1920  
 gaaatgtctc tctgtcatgg aagccaaagt gaaggctcag actgctccaa acaaggatgt 1980  
 gcagagggag atcgtctga 1998

<210> 47  
 <211> 327  
 <212> PRT  
 <213> Homo sapiens

<400> 47  
 Met Asp Ser Thr Leu Thr Ala Ser Glu Ile Arg Gln Arg Phe Ile Asp  
 1 5 10 15  
 Phe Phe Lys Arg Asn Glu His Thr Tyr Val His Ser Ser Ala Thr Ile  
 20 25 30  
 Pro Leu Asp Asp Pro Thr Leu Leu Phe Ala Asn Ala Gly Met Asn Gln  
 35 40 45  
 Phe Lys Pro Ile Phe Leu Asn Thr Ile Asp Pro Ser His Pro Met Ala  
 50 55 60  
 Lys Leu Ser Arg Ala Ala Asn Thr Gln Lys Cys Ile Arg Ala Gly Gly  
 65 70 75 80  
 Lys His Asn Asp Leu Asp Asp Val Gly Lys Asp Val Tyr His His Thr  
 85 90 95  
 Phe Phe Glu Met Leu Gly Ser Trp Ser Phe Gly Asp Tyr Phe Lys Glu

120161\_426PC\_Sequence\_Listing.TXT

Leu	Ala	Cys	100	Lys	Met	Ala	Leu	Glu	105	Leu	Leu	Thr	Gln	Glu	110	Phe	Gly	Ile
Pro	Ile	Glu	115	Arg	Leu	Tyr	Val	Thr	120	Tyr	Phe	Gly	Gly	Asp	125	Glu	Ala	Ala
Gly	Leu	Glu	130	Ala	Asp	Leu	Glu	Cys	135	Lys	Gln	Ile	Trp	Gln	140	Asn	Leu	Gly
Leu	Asp	Asp	145	Thr	Lys	Ile	Leu	Pro	150	Gly	Asn	Met	Lys	Asp	155	Asn	Phe	Trp
Glu	Met	Gly	160	Asp	Thr	Gly	Pro	Cys	165	Gly	Pro	Cys	Ser	Glu	170	Ile	His	Tyr
Asp	Arg	Ile	175	Gly	Gly	Arg	Asp	Ala	180	Ala	His	Leu	Val	Asn	185	Gln	Asp	Asp
Pro	Asn	Val	190	Leu	Glu	Ile	Trp	Asn	195	Leu	Val	Phe	Ile	Gln	200	Tyr	Asn	Arg
Glu	Ala	Asp	205	Gly	Ile	Leu	Lys	Pro	210	Leu	Pro	Lys	Lys	Ser	215	Ile	Asp	Thr
Gly	Met	Gly	220	Leu	Glu	Arg	Leu	Val	225	Ser	Val	Leu	Gln	Asn	230	Lys	Met	Ser
Asn	Tyr	Asp	235	Thr	Asp	Leu	Phe	Val	240	Pro	Tyr	Phe	Glu	Ala	245	Ile	Gln	Lys
Gly	Thr	Gly	250	Ala	Arg	Pro	Tyr	Thr	255	Gly	Lys	Val	Gly	Ala	260	Glu	Asp	Ala
Asp	Gly	Ile	265	Asp	Met	Ala	Tyr	Arg	270	Val	Leu	Ala	Asp	His	275	Ala	Arg	Thr
Ile	Thr	Val	280	Ala	Leu	Ala	Asp	Gly	285	Gly	Arg	Pro	Asp	Asn	290	Thr	Gly	Arg
Gly	Glu	Met	295	His	Phe	Leu	Ser		300						305			
			310						315						320			
			325															

<210> 48  
 <211> 984  
 <212> DNA  
 <213> Homo sapiens

<400> 48  
 atggactcta ctctaacagc aagtgaatc cggcagcgat ttatagattt cttcaagagg 60  
 aacgagcata cgtatgttca ctgctctgcc accatcccat tggatgaccc cactttgctc 120  
 ttgccaatg caggcatgaa ccagtttaaa cccattttcc tgaacacaat tgacccatct 180  
 caccatcatg caaagctgag cagagctgcc aataccaga agtgcacccg ggctgggggc 240  
 aaacataatg acctggacga tgtgggcaag gatgtctatc atcacacctt cttcgagatg 300  
 ctgggctctt ggtcttttgg agattacttt aaggaattgg catgtaagat ggctctggaa 360  
 ctctcaccac aagagtttgg cattccattt gaaagacttt atgttactta ctttggcggg 420  
 gatgaagcag ctggcttaga agcagatctg gaatgcaaac agatctggca aaatttgggg 480  
 ctggatgaca ccaaaatcct cccaggcaac atgaaggata acttctggga gatgggtgac 540  
 acgggcccct gtggctcctg cagtgaatc cactacgacc ggatttggtg tcgggacgcc 600  
 gcacatcttg tcaaccagga cgacccta atgtctggaga tctggaacct tgtgttcac 660  
 cagtataaca ggaagctga tggcattctg aaacctcttc ccaagaaaag cattgacaca 720  
 gggatggggc tggaaagact ggtatctgtg ctgcagaata agatgtccaa ctatgacact 780  
 gaccttttgg tcccttactt tgaagccatt cagaagggca cagggtgccg accatacact 840  
 gggaaagtgt gtgctgagga tgccgatggg attgacatgg cctaccgggt gctggctgac 900  
 cagctcggga ccatcactgt ggcactggct gatggtggcc ggcctgacaa cacagggcgt 960  
 ggggagatgc atttctgag ctga 984

<210> 49  
 <211> 50  
 <212> DNA  
 <213> Homo sapiens

<400> 49  
 acctgtgtt catccagtat aacagggcac aggtgcccga ccatacactg 50

<210> 50  
 <211> 16  
 <212> PRT  
 <213> Homo sapiens

120161\_426PC\_Sequence\_Listing.TXT

<400> 50

Leu Val Phe Ile Gln Tyr Asn Arg Ala Gln Val Pro Asp His Thr Leu  
1 5 10 15

<210> 51

<211> 50

<212> DNA

<213> Homo sapiens

<400> 51

catttggtact ccagtggtag ctatgaaaac agagttttaca gtgaagaatg 50

<210> 52

<211> 16

<212> PRT

<213> Homo sapiens

<400> 52

His Leu Asp Ser Ser Gly Ser Tyr Glu Asn Arg Val Tyr Ser Glu Glu  
1 5 10 15

<210> 53

<211> 50

<212> DNA

<213> Homo sapiens

<400> 53

tcaggtctgg ctgtttattg atgaggcacc tgcggaactc gagtcatgca 50

<210> 54

<211> 16

<212> PRT

<213> Homo sapiens

<400> 54

Gln Val Trp Leu Phe Ile Asp Glu Ala Pro Ala Glu Leu Glu Ser Cys  
1 5 10 15

<210> 55

<211> 50

<212> DNA

<213> Homo sapiens

<400> 55

gccggcctga caacacaggg cgtggggaga tgcatttcct gagctgaaga 50

<210> 56

<211> 14

<212> PRT

<213> Homo sapiens

<400> 56

Arg Pro Asp Asn Thr Gly Arg Gly Glu Met His Phe Leu Ser  
1 5 10

<210> 57

<211> 286

<212> PRT

<213> Homo sapiens

<400> 57

Met Asp Ser Thr Leu Thr Ala Ser Glu Ile Arg Gln Arg Phe Ile Asp  
1 5 10 15  
Phe Phe Lys Arg Asn Glu His Thr Tyr Val His Ser Ser Ala Thr Ile

120161\_426PC\_Sequence\_Listing.TXT

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

<210> 58  
 <211> 858  
 <212> DNA  
 <213> Homo sapiens

<400> 58	atggactcta	ctctaacagc	aagtgaatc	cggcagcgat	ttatagattt	cttcaagagg	60
	aacgagcata	cgtatgttca	ctcgtctgcc	accatcccat	tgatgaccc	cactttgctc	120
	tttgccaatg	caggcatgaa	ccagtttaaa	cccattttcc	tgaacacaat	tgacccatct	180
	caccccatgg	caaagctgag	cagagctgcc	aatacccaga	agtgcacccg	ggctgggggc	240
	aaacataatg	acctggacga	tgtgggcaag	gatgtctatc	atcacacctt	cttcgagatg	300
	ctgggctctt	ggtcttttgg	agattacttt	aaggaattgg	catgtaagat	ggctctggaa	360
	ctccicaccc	aagagtttgg	cattcccatt	gaaagacttt	atgttactta	ctttggcggg	420
	gatgaagcag	ctggcttaga	agcagatctg	gaatgcaaac	agatctggca	aaatttgggg	480
	ctggatgaca	ccaaaatcct	cccaggcaac	atgaaggata	acttctggga	gatgggtgac	540
	acgggcccct	gtggtccttg	cagtgaatc	cactacgacc	ggattggttg	tcgggacgcc	600
	gcacatcttg	tcaaccagga	cgacccta	gtgctggaga	tctggaacct	tgtgttcac	660
	cagtataaca	gggaagctga	tggcattctg	aaacctcttc	ccaagaaaag	cattgacaca	720
	gggatgggcc	tggaaacgact	ggtatctgtg	ctgcagaata	agatgtccaa	ctatgacact	780
	gaccttttgg	tcccttactt	tgaagccatt	cagaagggca	caggtgccc	accatacact	840
	gggaaagtgg	gtgctgag					858

<210> 59  
 <211> 488  
 <212> PRT  
 <213> Homo sapiens

<400> 59	Met	Asp	Ser	Thr	Leu	Thr	Ala	Ser	Glu	Ile	Arg	Gln	Arg	Phe	Ile	Asp
1	Phe	Phe	Lys	Arg	Asn	Glu	His	Thr	Tyr	Val	His	Ser	Ser	Ala	Thr	Ile
	Pro	Leu	Asp	Asp	Pro	Thr	Leu	Leu	Phe	Ala	Asn	Ala	Gly	Met	Asn	Gln

120161\_426PC\_Sequence\_Listing.TXT

Phe	Lys	35	Pro	Ile	Phe	Leu	Asn	40	Thr	Ile	Asp	Pro	Ser	45	His	Pro	Met	Ala
Lys	50	Leu	Ser	Arg	Ala	Ala	Asn	55	Thr	Gln	Lys	Cys	Ile	60	Arg	Ala	Gly	Gly
65	Lys	His	Asn	Asp	Leu	Asp	Asp	70	Val	Gly	Lys	Asp	Val	75	Tyr	His	His	Thr
					85						90					95		
Phe	Phe	Glu	Met	100	Leu	Gly	Ser	Trp	Ser	105	Phe	Gly	Asp	Tyr	110	Phe	Lys	Glu
Leu	Ala	Cys	Lys	115	Met	Ala	Leu	Glu	Leu	120	Leu	Thr	Gln	Glu	125	Phe	Gly	Ile
Pro	Ile	Glu	Arg	Leu	Tyr	Val	135	Thr	Tyr	Phe	Gly	Gly	Asp	Glu	140	Ala	Ala	
130	Gly	Leu	Glu	Ala	Asp	Leu	Glu	Cys	Lys	Gln	Ile	155	Trp	Gln	Asn	Leu	Gly	160
145	Leu	Asp	Asp	Thr	Lys	Ile	Leu	Pro	Gly	Asn	Met	170	Lys	Asp	Asn	Phe	Trp	
					165							185				190		
Glu	Met	Gly	Asp	180	Thr	Gly	Pro	Cys	Gly	Pro	Cys	Ser	Glu	Ile	His	Tyr		
Asp	Arg	Ile	Gly	195	Gly	Arg	Asp	Ala	200	His	Leu	Val	Asn	Gln	Asp	Asp		
Pro	Asn	Val	Leu	Glu	Ile	Trp	Asn	Leu	Val	Phe	Ile	Gln	Tyr	Asn	Arg			
210	Glu	Ala	Asp	Gly	Ile	Leu	Lys	Pro	Leu	Pro	Lys	Lys	Ser	Ile	Asp	Thr		
225	Gly	Met	Gly	Leu	Glu	Arg	Leu	Val	Ser	Val	250	Leu	Gln	Asn	Lys	Met	Ser	
					245											255		
Asn	Tyr	Asp	Thr	260	Asp	Leu	Phe	Val	Pro	Tyr	Phe	Glu	Ala	Ile	Gln	Lys		
Gly	Thr	Gly	Ala	275	Arg	Pro	Tyr	Thr	Gly	Lys	Val	Gly	Ala	Glu	Asp	Ala		
Asp	Gly	Ile	Asp	Met	Ala	Tyr	295	Arg	Val	Leu	Ala	Asp	300	His	Ala	Arg	Thr	
Ile	Thr	Val	Ala	Leu	Ala	Asp	Gly	Gly	Arg	Pro	Asp	Asn	Thr	Gly	Arg			
305	Gly	Tyr	Val	Leu	Arg	Arg	Ile	Leu	Arg	Arg	Ala	Val	Arg	Tyr	Ala	His		
					325											335		
Glu	Lys	Leu	Asn	340	Ala	Ser	Arg	Gly	Phe	345	Phe	Ala	Thr	Leu	Val	Asp	Val	
Val	Val	Gln	Ser	Leu	Gly	Asp	Ala	Phe	Pro	Glu	Leu	Lys	Lys	Asp	Pro			
					355											365		
Asp	Met	Val	Lys	Asp	Ile	Ile	Asn	Glu	Glu	Glu	Val	Gln	Phe	Leu	Lys			
370	Thr	Leu	Ser	Arg	Gly	Arg	Arg	Ile	Leu	Asp	Arg	Lys	Ile	Gln	Ser	Leu		
385	Gly	Asp	Ser	Lys	Thr	Ile	Pro	Gly	Asp	Thr	Ala	Trp	Leu	Leu	Tyr	Asp		
					405											415		
Thr	Tyr	Gly	Phe	420	Pro	Val	Asp	Leu	Thr	Gly	Leu	Ile	Ala	Glu	Gly	Lys		
																430		
Gly	Leu	Val	Val	Asp	Met	Asp	Gly	Phe	Glu	Glu	Glu	Arg	Lys	Leu	Ala			
																445		
Gln	Leu	Lys	Ser	Gln	Gly	Lys	Gly	Ala	Gly	Gly	Glu	Asp	Leu	Ile	Met			
450	Leu	Asp	Ile	Tyr	Ala	Ile	Glu	Glu	Leu	Arg	Ala	Arg	Gly	Leu	Glu	Val		
465	Thr	Asp	Asp	Ser	Pro	Lys	Tyr	Asn								480		
					485													

<210> 60  
 <211> 1464  
 <212> DNA  
 <213> Homo sapiens

<400> 60  
 atggactcta ctctaacagc aagtgaatc cggcagcgat ttatagattt cttcaagagg 60  
 aacgagcata cgatatgtca ctgctctgcc accatcccat tggatgaccc cactttgctc 120  
 ttgccaatg caggcatgaa ccagtttaaa cccattttcc tgaacacaat tgaccatct 180

120161\_426PC\_Sequence\_Listing.TXT

```

caccatg ccaagctg cagagctg aataccaga agtgcaccc ggctgggggc 240
aaacataa acctggacga tgtgggcaag gatgtctatc atcacacctt cttcgagatg 300
ctgggctctt ggtctttttg agattacttt aaggaattgg catgtaagat ggctctggaa 360
ctcctcacc cagagtttgg cattcccat gaaagacttt atgttactta ctttggcggg 420
gatgaagcag ctggcttaga agcagatctg gaatgcaaac agatctggca aaatttgggg 480
ctggatgaca ccaaaatcct cccaggcaac atgaaggata acttctggga gatgggtgac 540
acggggccc gtggtccttg cagtgaatc cactacgacc ggattgggtg tcgggacgcc 600
gcacatcttg tcaaccagga cgacccta atgtctggaga tctggaacct tgtgttcac 660
cagtataaca gggaagctga tggcattctg aaacctctt ccaagaaaag cattgacaca 720
gggatggggc tggaaagact ggtatctgtg ctgcagaata agatgtccaa ctatgacact 780
gacctttttg tcccttactt tgaagccatt cagaagggca cagggtgccg accatacact 840
gggaaagtgt gtgctgagga tgccgatggg attgacatgg cctaccgggt gctggctgac 900
cacgctcggg ccatcactgt ggcactggct gatggtggcc ggcctgacaa cacagggcgt 960
ggatattgtg tgagacggat tctccgcca gctgtccgat acgcccataa aaagctcaat 1020
gccagcaggg gcttctttgc tacgttagtg gatgtgtcgc tccagtcctt gggagatgca 1080
tttctgagc tgaagaagga cccagacatg gtgaaggaca tcattaatga agaagaggtg 1140
cagtttctca agactctcag cagagggcgt cgcactctgg acaggaaaat tcagagcctg 1200
ggagacagca agaccattcc cggagacact gcttggctcc tctatgacac ctatgggttt 1260
ccagtggatc tgactggact gattgctgaa gagaagggcc tgggtggtaga catggatggc 1320
tttgaagagg agaggaaact ggccagctg aaatcacagg gcaagggagc tgggtgggaa 1380
gacctatta tgctggacat ttacgctatc gaagagctcc gggcacgggg tctggaggtc 1440
acagatgatt ccccaaagta caat
1464

```

<210> 61

<211> 421

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanine tRNA Synthetase polypeptide with N-terminal  
6xHis affinity tag

<400> 61

```

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu
 1          5          10          15
Gly Leu Asp Ser Thr Asp Ser Thr Leu Thr Ala Ser Glu Ile Arg Gln
 20          25          30
Arg Phe Ile Asp Phe Phe Lys Arg Asn Glu His Thr Tyr Val His Ser
 35          40          45
Ser Ala Thr Ile Pro Leu Asp Asp Pro Thr Leu Leu Phe Ala Asn Ala
 50          55          60
Gly Met Asn Gln Phe Lys Pro Ile Phe Leu Asn Thr Ile Asp Pro Ser
 65          70          75          80
His Pro Met Ala Lys Leu Ser Arg Ala Ala Asn Thr Gln Lys Cys Ile
 85          90          95
Arg Ala Gly Gly Lys His Asn Asp Leu Asp Asp Val Gly Lys Asp Val
100          105          110
Tyr His His Thr Phe Phe Glu Met Leu Gly Ser Trp Ser Phe Gly Asp
115          120          125
Tyr Phe Lys Glu Leu Ala Cys Lys Met Ala Leu Glu Leu Leu Thr Gln
130          135          140
Glu Phe Gly Ile Pro Ile Glu Arg Leu Tyr Val Thr Tyr Phe Gly Gly
145          150          155          160
Asp Glu Ala Ala Gly Leu Glu Ala Asp Leu Glu Cys Lys Gln Ile Trp
165          170          175
Gln Asn Leu Gly Leu Asp Asp Thr Lys Ile Leu Pro Gly Asn Met Lys
180          185          190
Asp Asn Phe Trp Glu Met Gly Asp Thr Gly Pro Cys Gly Pro Cys Ser
195          200          205
Glu Ile His Tyr Asp Arg Ile Gly Gly Arg Asp Ala Ala His Leu Val
210          215          220
Asn Gln Asp Asp Pro Asn Val Leu Glu Ile Trp Asn Leu Val Phe Ile
225          230          235          240
Gln Tyr Asn Arg Glu Ala Asp Gly Ile Leu Lys Pro Leu Pro Lys Lys
245          250          255
Ser Ile Asp Thr Gly Met Gly Leu Glu Arg Leu Val Ser Val Leu Gln
260          265          270
Asn Lys Met Ser Asn Tyr Asp Thr Asp Leu Phe Val Pro Tyr Phe Glu

```

120161\_426PC\_Sequence\_Listing.TXT

Ala	Ile	275	Gln	Lys	Gly	Thr	Gly	280	Ala	Arg	Pro	Tyr	Thr	285	Gly	Lys	Val	Gly
290	Ala	Glu	Asp	Ala	Asp	Gly	Ile	295	Asp	Met	Ala	Tyr	Arg	300	Val	Leu	Ala	Asp
305	His	Ala	Arg	Thr	Ile	310	Thr	Val	Ala	Leu	Ala	Asp	Gly	315	Gly	Arg	Pro	Asp
	Asn	Thr	Gly	Arg	325	Gly	Tyr	Val	Leu	Arg	330	Ile	Leu	Arg	335	Ala	Val	
	Arg	Tyr	Ala	His	340	Glu	Lys	Leu	Asn	Ala	Ser	Arg	Gly	345	Phe	Phe	Ala	Thr
	Leu	Val	Asp	Val	355	Val	Val	Gln	Ser	Leu	Gly	Asp	Ala	360	Phe	Pro	Glu	Leu
	Lys	Lys	Asp	Pro	375	Met	Val	Lys	Asp	Ile	Ile	380	Asn	Glu	Glu	Glu	Val	
385	Gln	Phe	Leu	Lys	390	Thr	Leu	Ser	Arg	Gly	Arg	400	Ile	Leu	Asp	Arg	Lys	
	Ile	Gln	Ser	Leu	405	Gly					410					415		
				420														

<210> 62

<211> 421

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanine tRNA Synthetase polypeptide with C-terminal  
6xHis affinity tag

<400> 62

Met	Asp	Ser	Thr	Leu	Thr	Ala	Ser	Glu	Ile	Arg	Gln	Arg	Phe	Ile	Asp
1	Phe	Phe	Lys	Arg	Asn	Glu	His	Thr	Tyr	Val	His	Ser	Ser	Ala	Thr
	20	Pro	Leu	Asp	Asp	Pro	Thr	Leu	Leu	Phe	Ala	Asn	Ala	Gly	Met
	35	Phe	Lys	Pro	Ile	Phe	Leu	Asn	Thr	Ile	Asp	Pro	Ser	His	Pro
	50	Lys	Leu	Ser	Arg	Ala	Ala	Asn	Thr	Gln	Lys	Cys	Ile	Arg	Ala
65	Lys	His	Asn	Asp	Leu	Asp	Asp	Val	Gly	Lys	Asp	Val	Tyr	His	His
	85	Phe	Phe	Glu	Met	Leu	Gly	Ser	Trp	Ser	Phe	Gly	Asp	Tyr	Phe
	100	Leu	Ala	Cys	Lys	Met	Ala	Leu	Glu	Leu	Leu	Thr	Gln	Glu	Phe
	115	Pro	Ile	Glu	Arg	Leu	Tyr	Val	Thr	Tyr	Phe	Gly	Gly	Asp	Glu
	130	Gly	Leu	Glu	Ala	Asp	Leu	Glu	Cys	Lys	Gln	Ile	Trp	Gln	Asn
145	Leu	Asp	Asp	Thr	Lys	Ile	Leu	Pro	Gly	Asn	Met	Lys	Asp	Asn	Phe
	165	Glu	Met	Gly	Asp	Thr	Gly	Pro	Cys	Gly	Pro	Cys	Ser	Glu	Ile
	180	Asp	Arg	Ile	Gly	Gly	Arg	Asp	Ala	Ala	His	Leu	Val	Asn	Gln
	195	Pro	Asn	Val	Leu	Glu	Ile	Trp	Asn	Leu	Val	Phe	Ile	Gln	Tyr
	210	Glu	Ala	Asp	Gly	Ile	Leu	Lys	Pro	Leu	Pro	Lys	Lys	Ser	Ile
225	Gly	Met	Gly	Leu	Glu	Arg	Leu	Val	Ser	Val	Leu	Gln	Asn	Lys	Met
	245	Asn	Tyr	Asp	Thr	Asp	Leu	Phe	Val	Pro	Tyr	Phe	Glu	Ala	Ile
	260	Gly	Thr	Gly	Ala	Arg	Pro	Tyr	Thr	Gly	Lys	Val	Gly	Ala	Glu
	275	Asp	Gly	Ile	Asp	Met	Ala	Tyr	Arg	Val	Leu	Ala	Asp	His	Ala
															Arg
															Thr

120161\_426PC\_Sequence\_Listing.TXT

```

      290
Ile Thr Val Ala Leu Ala Asp Gly Gly Arg Pro Asp Asn Thr Gly Arg
305 Gly Tyr Val Leu Arg Arg Ile Leu Arg Arg Ala Val Arg Tyr Ala His
      310
      325
Glu Lys Leu Asn Ala Ser Arg Gly Phe Phe Ala Thr Leu Val Asp Val
      340
Val Val Gln Ser Leu Gly Asp Ala Phe Pro Glu Leu Lys Lys Asp Pro
      355
Asp Met Val Lys Asp Ile Ile Asn Glu Glu Glu Val Gln Phe Leu Lys
      370
Thr Leu Ser Arg Gly Arg Arg Ile Leu Asp Arg Lys Ile Gln Ser Leu
385 Gly Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr His
      405
His His His His His
      420

```

<210> 63

<211> 310

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanyl tRNA Synthetase polypeptide with N-terminal  
6xHis affinity tag

<400> 63

```

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu
1 5 10 15
Gly Leu Asp Ser Thr Gly Ser Asp Ser Thr Leu Thr Ala Ser Glu Ile
      20
Arg Gln Arg Phe Ile Asp Phe Phe Lys Arg Asn Glu His Thr Tyr Val
      35
His Ser Ser Ala Thr Ile Pro Leu Asp Asp Pro Thr Leu Leu Phe Ala
      50
Asn Ala Gly Met Asn Gln Phe Lys Pro Ile Phe Leu Asn Thr Ile Asp
65 70 75 80
Pro Ser His Pro Met Ala Lys Leu Ser Arg Ala Ala Asn Thr Gln Lys
      85
Cys Ile Arg Ala Gly Gly Lys His Asn Asp Leu Asp Asp Val Gly Lys
      100
Asp Val Tyr His His Thr Phe Phe Glu Met Leu Gly Ser Trp Ser Phe
      115
Gly Asp Tyr Phe Lys Glu Leu Ala Cys Lys Met Ala Leu Glu Leu Leu
      130
Thr Gln Glu Phe Gly Ile Pro Ile Glu Arg Leu Tyr Val Thr Tyr Phe
145 150 155 160
Gly Gly Asp Glu Ala Ala Gly Leu Glu Ala Asp Leu Glu Cys Lys Gln
      165
Ile Trp Gln Asn Leu Gly Leu Asp Asp Thr Lys Ile Leu Pro Gly Asn
      180
Met Lys Asp Asn Phe Trp Glu Met Gly Asp Thr Gly Pro Cys Gly Pro
      195
Cys Ser Glu Ile His Tyr Asp Arg Ile Gly Gly Arg Asp Ala Ala His
      210
Leu Val Asn Gln Asp Asp Pro Asn Val Leu Glu Ile Trp Asn Leu Val
225 230 235 240
Phe Ile Gln Tyr Asn Arg Glu Ala Asp Gly Ile Leu Lys Pro Leu Pro
      245
Lys Lys Ser Ile Asp Thr Gly Met Gly Leu Glu Arg Leu Val Ser Val
      260
Leu Gln Asn Lys Met Ser Asn Tyr Asp Thr Asp Leu Phe Val Pro Tyr
      275
Phe Glu Ala Ile Gln Lys Gly Thr Gly Ala Arg Pro Tyr Thr Gly Lys
290 295 300
Val Gly Ala Glu Leu Glu

```



305

310

<210> 64  
 <211> 310  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanine tRNA Synthetase polypeptide with C-terminal  
 6xHis affinity tag

<400> 64  
 Met Gly Ser Asp Ser Thr Leu Thr Ala Ser Glu Ile Arg Gln Arg Phe  
 1 5 10 15  
 Ile Asp Phe Phe Lys Arg Asn Glu His Thr Tyr Val His Ser Ser Ala  
 20 25 30  
 Thr Ile Pro Leu Asp Asp Pro Thr Leu Leu Phe Ala Asn Ala Gly Met  
 35 40 45  
 Asn Gln Phe Lys Pro Ile Phe Leu Asn Thr Ile Asp Pro Ser His Pro  
 50 55 60  
 Met Ala Lys Leu Ser Arg Ala Ala Asn Thr Gln Lys Cys Ile Arg Ala  
 65 70 75 80  
 Gly Gly Lys His Asn Asp Leu Asp Asp Val Gly Lys Asp Val Tyr His  
 85 90 95  
 His Thr Phe Phe Glu Met Leu Gly Ser Trp Ser Phe Gly Asp Tyr Phe  
 100 105 110  
 Lys Glu Leu Ala Cys Lys Met Ala Leu Glu Leu Leu Thr Gln Glu Phe  
 115 120 125  
 Gly Ile Pro Ile Glu Arg Leu Tyr Val Thr Tyr Phe Gly Gly Asp Glu  
 130 135 140  
 Ala Ala Gly Leu Glu Ala Asp Leu Glu Cys Lys Gln Ile Trp Gln Asn  
 145 150 155 160  
 Leu Gly Leu Asp Asp Thr Lys Ile Leu Pro Gly Asn Met Lys Asp Asn  
 165 170 175  
 Phe Trp Glu Met Gly Asp Thr Gly Pro Cys Gly Pro Cys Ser Glu Ile  
 180 185 190  
 His Tyr Asp Arg Ile Gly Gly Arg Asp Ala Ala His Leu Val Asn Gln  
 195 200 205  
 Asp Asp Pro Asn Val Leu Glu Ile Trp Asn Leu Val Phe Ile Gln Tyr  
 210 215 220  
 Asn Arg Glu Ala Asp Gly Ile Leu Lys Pro Leu Pro Lys Lys Ser Ile  
 225 230 235 240  
 Asp Thr Gly Met Gly Leu Glu Arg Leu Val Ser Val Leu Gln Asn Lys  
 245 250 255  
 Met Ser Asn Tyr Asp Thr Asp Leu Phe Val Pro Tyr Phe Glu Ala Ile  
 260 265 270  
 Gln Lys Gly Thr Gly Ala Arg Pro Tyr Thr Gly Lys Val Gly Ala Glu  
 275 280 285  
 Leu Glu Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr  
 290 295 300  
 His His His His His His  
 305 310

<210> 65  
 <211> 508  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanine tRNA Synthetase polypeptide with N-terminal  
 6xHis affinity tag

<400> 65  
 Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15  
 Gly Leu Asp Ser Thr Asp Ser Thr Leu Thr Ala Ser Glu Ile Arg Gln

## 120161\_426PC\_Sequence\_Listing.TXT

Arg	Phe	Ile	Asp	Phe	Phe	Lys	Arg	Asn	Glu	His	Thr	Tyr	Val	His	Ser
Ser	Ala	Thr	Ile	Pro	Leu	Asp	Asp	Pro	Thr	Leu	Leu	Phe	Ala	Asn	Ala
Gly	Met	Asn	Gln	Phe	Lys	Pro	Ile	Phe	Leu	Asn	Thr	Ile	Asp	Pro	Ser
His	Pro	Met	Ala	Lys	Leu	Ser	Arg	Ala	Ala	Asn	Thr	Gln	Lys	Cys	Ile
Arg	Ala	Gly	Gly	Lys	His	Asn	Asp	Leu	Asp	Asp	Val	Gly	Lys	Asp	Val
Tyr	His	His	Thr	Phe	Phe	Glu	Met	Leu	Gly	Ser	Trp	Ser	Phe	Gly	Asp
Tyr	Phe	Lys	Glu	Leu	Ala	Cys	Lys	Met	Ala	Leu	Glu	Leu	Leu	Thr	Gln
Glu	Phe	Gly	Ile	Pro	Ile	Glu	Arg	Leu	Tyr	Val	Thr	Tyr	Phe	Gly	Gly
Asp	Glu	Ala	Ala	Gly	Leu	Glu	Ala	Asp	Leu	Glu	Cys	Lys	Gln	Ile	Trp
Gln	Asn	Leu	Gly	Leu	Asp	Asp	Thr	Lys	Ile	Leu	Pro	Gly	Asn	Met	Lys
Asp	Asn	Phe	Trp	Glu	Met	Gly	Asp	Thr	Gly	Pro	Cys	Gly	Pro	Cys	Ser
Glu	Ile	His	Tyr	Asp	Arg	Ile	Gly	Gly	Arg	Asp	Ala	Ala	His	Leu	Val
Asn	Gln	Asp	Asp	Pro	Asn	Val	Leu	Glu	Ile	Trp	Asn	Leu	Val	Phe	Ile
Gln	Tyr	Asn	Arg	Glu	Ala	Asp	Gly	Ile	Leu	Lys	Pro	Leu	Pro	Lys	Lys
Ser	Ile	Asp	Thr	Gly	Met	Gly	Leu	Glu	Arg	Leu	Val	Ser	Val	Leu	Gln
Asn	Lys	Met	Ser	Asn	Tyr	Asp	Thr	Asp	Leu	Phe	Val	Pro	Tyr	Phe	Glu
Ala	Ile	Gln	Lys	Gly	Thr	Gly	Ala	Arg	Pro	Tyr	Thr	Gly	Lys	Val	Gly
Ala	Glu	Asp	Ala	Asp	Gly	Ile	Asp	Met	Ala	Tyr	Arg	Val	Leu	Ala	Asp
His	Ala	Arg	Thr	Ile	Thr	Val	Ala	Leu	Ala	Asp	Gly	Gly	Arg	Pro	Asp
Asn	Thr	Gly	Arg	Gly	Tyr	Val	Leu	Arg	Arg	Ile	Leu	Arg	Arg	Ala	Val
Arg	Tyr	Ala	His	Glu	Lys	Leu	Asn	Ala	Ser	Arg	Gly	Phe	Phe	Ala	Thr
Leu	Val	Asp	Val	Val	Val	Gln	Ser	Leu	Gly	Asp	Ala	Phe	Pro	Glu	Leu
Lys	Lys	Asp	Pro	Asp	Met	Val	Lys	Asp	Ile	Ile	Asn	Glu	Glu	Glu	Val
Gln	Phe	Leu	Lys	Thr	Leu	Ser	Arg	Gly	Arg	Arg	Ile	Leu	Asp	Arg	Lys
Ile	Gln	Ser	Leu	Gly	Asp	Ser	Lys	Thr	Ile	Pro	Gly	Asp	Thr	Ala	Trp
Leu	Leu	Tyr	Asp	Thr	Tyr	Gly	Phe	Pro	Val	Asp	Leu	Thr	Gly	Leu	Ile
Ala	Glu	Glu	Lys	Gly	Leu	Val	Val	Asp	Met	Asp	Gly	Phe	Glu	Glu	Glu
Arg	Lys	Leu	Ala	Gln	Leu	Lys	Ser	Gln	Gly	Lys	Gly	Ala	Gly	Gly	Glu
Asp	Leu	Ile	Met	Leu	Asp	Ile	Tyr	Ala	Ile	Glu	Glu	Leu	Arg	Ala	Arg
Gly	Leu	Glu	Val	Thr	Asp	Asp	Ser	Pro	Lys	Tyr	Asn				

<210> 66  
<211> 508  
<212> PRT  
<213> Arti fi ci al Sequence

120161\_426PC\_Sequence\_Listing.TXT

<220>

<223> Alanyl tRNA Synthetase polypeptide with C-terminal  
6xHis affinity tag

<400> 66

Met	Asp	Ser	Thr	Leu	Thr	Ala	Ser	Glu	Ile	Arg	Gln	Arg	Phe	Ile	Asp
1				5					10					15	
Phe	Phe	Lys	Arg	Asn	Glu	His	Thr	Tyr	Val	His	Ser	Ser	Ala	Thr	Ile
			20					25					30		
Pro	Leu	Asp	Asp	Pro	Thr	Leu	Leu	Phe	Ala	Asn	Ala	Gly	Met	Asn	Gln
		35					40					45			
Phe	Lys	Pro	Ile	Phe	Leu	Asn	Thr	Ile	Asp	Pro	Ser	His	Pro	Met	Ala
	50					55					60				
Lys	Leu	Ser	Arg	Ala	Ala	Asn	Thr	Gln	Lys	Cys	Ile	Arg	Ala	Gly	Gly
65				70					75					80	
Lys	His	Asn	Asp	Leu	Asp	Asp	Val	Gly	Lys	Asp	Val	Tyr	His	His	Thr
			85					90					95		
Phe	Phe	Glu	Met	Leu	Gly	Ser	Trp	Ser	Phe	Gly	Asp	Tyr	Phe	Lys	Glu
			100					105					110		
Leu	Ala	Cys	Lys	Met	Ala	Leu	Glu	Leu	Leu	Thr	Gln	Glu	Phe	Gly	Ile
		115					120					125			
Pro	Ile	Glu	Arg	Leu	Tyr	Val	Thr	Tyr	Phe	Gly	Gly	Asp	Glu	Ala	Ala
	130					135					140				
Gly	Leu	Glu	Ala	Asp	Leu	Glu	Cys	Lys	Gln	Ile	Trp	Gln	Asn	Leu	Gly
145				150					155					160	
Leu	Asp	Asp	Thr	Lys	Ile	Leu	Pro	Gly	Asn	Met	Lys	Asp	Asn	Phe	Trp
			165					170						175	
Glu	Met	Gly	Asp	Thr	Gly	Pro	Cys	Gly	Pro	Cys	Ser	Glu	Ile	His	Tyr
			180					185					190		
Asp	Arg	Ile	Gly	Gly	Arg	Asp	Ala	Ala	His	Leu	Val	Asn	Gln	Asp	Asp
	195						200					205			
Pro	Asn	Val	Leu	Glu	Ile	Trp	Asn	Leu	Val	Phe	Ile	Gln	Tyr	Asn	Arg
	210					215					220				
Glu	Ala	Asp	Gly	Ile	Leu	Lys	Pro	Leu	Pro	Lys	Lys	Ser	Ile	Asp	Thr
225				230					235					240	
Gly	Met	Gly	Leu	Glu	Arg	Leu	Val	Ser	Val	Leu	Gln	Asn	Lys	Met	Ser
			245					250					255		
Asn	Tyr	Asp	Thr	Asp	Leu	Phe	Val	Pro	Tyr	Phe	Glu	Ala	Ile	Gln	Lys
			260					265					270		
Gly	Thr	Gly	Ala	Arg	Pro	Tyr	Thr	Gly	Lys	Val	Gly	Ala	Glu	Asp	Ala
		275				280					285				
Asp	Gly	Ile	Asp	Met	Ala	Tyr	Arg	Val	Leu	Ala	Asp	His	Ala	Arg	Thr
	290					295					300				
Ile	Thr	Val	Ala	Leu	Ala	Asp	Gly	Gly	Arg	Pro	Asp	Asn	Thr	Gly	Arg
305				310					315					320	
Gly	Tyr	Val	Leu	Arg	Arg	Ile	Leu	Arg	Arg	Ala	Val	Arg	Tyr	Ala	His
			325					330					335		
Glu	Lys	Leu	Asn	Ala	Ser	Arg	Gly	Phe	Phe	Ala	Thr	Leu	Val	Asp	Val
			340					345					350		
Val	Val	Gln	Ser	Leu	Gly	Asp	Ala	Phe	Pro	Glu	Leu	Lys	Lys	Asp	Pro
		355					360				365				
Asp	Met	Val	Lys	Asp	Ile	Ile	Asn	Glu	Glu	Glu	Val	Gln	Phe	Leu	Lys
	370					375					380				
Thr	Leu	Ser	Arg	Gly	Arg	Arg	Ile	Leu	Asp	Arg	Lys	Ile	Gln	Ser	Leu
385				390					395					400	
Gly	Asp	Ser	Lys	Thr	Ile	Pro	Gly	Asp	Thr	Ala	Trp	Leu	Leu	Tyr	Asp
			405					410					415		
Thr	Tyr	Gly	Phe	Pro	Val	Asp	Leu	Thr	Gly	Leu	Ile	Ala	Glu	Glu	Lys
			420					425					430		
Gly	Leu	Val	Val	Asp	Met	Asp	Gly	Phe	Glu	Glu	Glu	Arg	Lys	Leu	Ala
		435					440					445			
Gln	Leu	Lys	Ser	Gln	Gly	Lys	Gly	Ala	Gly	Gly	Glu	Asp	Leu	Ile	Met
	450					455					460				
Leu	Asp	Ile	Tyr	Ala	Ile	Glu	Glu	Leu	Arg	Ala	Arg	Gly	Leu	Glu	Val
465				470					475					480	
Thr	Asp	Asp	Ser	Pro	Lys	Tyr	Asn	Gly	Lys	Pro	Ile	Pro	Asn	Pro	Leu
			485					490					495		
Leu	Gly	Leu	Asp	Ser	Thr	His	His	His	His	His	His	His	His	His	His

<210> 67  
 <211> 299  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanine tRNA Synthetase polypeptide with N-terminal  
 6xHis affinity tag

<400> 67  
 Met His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15  
 Gly Leu Asp Ser Thr Gly Ser Asp Ser Thr Leu Thr Ala Ser Glu Ile  
 20 25 30  
 Arg Gln Arg Phe Ile Asp Phe Phe Lys Arg Asn Glu His Thr Tyr Val  
 35 40 45  
 His Ser Ser Ala Thr Ile Pro Leu Asp Asp Pro Thr Leu Leu Phe Ala  
 50 55 60  
 Asn Ala Gly Met Asn Gln Phe Lys Pro Ile Phe Leu Asn Thr Ile Asp  
 65 70 75 80  
 Pro Ser His Pro Met Ala Lys Leu Ser Arg Ala Ala Asn Thr Gln Lys  
 85 90 95  
 Cys Ile Arg Ala Gly Gly Lys His Asn Asp Leu Asp Asp Val Gly Lys  
 100 105 110  
 Asp Val Tyr His His Thr Phe Phe Glu Met Leu Gly Ser Trp Ser Phe  
 115 120 125  
 Gly Asp Tyr Phe Lys Glu Leu Ala Cys Lys Met Ala Leu Glu Leu Leu  
 130 135 140  
 Thr Gln Glu Phe Gly Ile Pro Ile Glu Arg Leu Tyr Val Thr Tyr Phe  
 145 150 155 160  
 Gly Gly Asp Glu Ala Ala Gly Leu Glu Ala Asp Leu Glu Cys Lys Gln  
 165 170 175  
 Ile Trp Gln Asn Leu Gly Leu Asp Asp Thr Lys Ile Leu Pro Gly Asn  
 180 185 190  
 Met Lys Asp Asn Phe Trp Glu Met Gly Asp Thr Gly Pro Cys Gly Pro  
 195 200 205  
 Cys Ser Glu Ile His Tyr Asp Arg Ile Gly Gly Arg Asp Ala Ala His  
 210 215 220  
 Leu Val Asn Gln Asp Asp Pro Asn Val Leu Glu Ile Trp Asn Leu Val  
 225 230 235 240  
 Phe Ile Gln Tyr Asn Arg Ala Gln Val Pro Asp His Thr Leu Gly Lys  
 245 250 255  
 Leu Val Leu Arg Met Pro Met Gly Leu Thr Trp Pro Thr Gly Cys Trp  
 260 265 270  
 Leu Thr Thr Leu Gly Pro Ser Leu Trp His Trp Leu Met Val Ala Gly  
 275 280 285  
 Leu Thr Thr Gln Gly Val Asp Met Cys Leu Glu  
 290 295

<210> 68  
 <211> 299  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanine tRNA Synthetase polypeptide with C-terminal  
 6xHis affinity tag

<400> 68  
 Met Gly Ser Asp Ser Thr Leu Thr Ala Ser Glu Ile Arg Gln Arg Phe  
 1 5 10 15  
 Ile Asp Phe Phe Lys Arg Asn Glu His Thr Tyr Val His Ser Ser Ala  
 20 25 30  
 Thr Ile Pro Leu Asp Asp Pro Thr Leu Leu Phe Ala Asn Ala Gly Met

120161\_426PC\_Sequence\_Listing.TXT

Asn	Gln	35	Lys	Pro	Ile	Phe	40	Leu	Asn	Thr	Ile	Asp	45	Pro	Ser	His	Pro
50	Met	Ala	Lys	Leu	Ser	Arg	55	Ala	Ala	Asn	Thr	Gln	60	Lys	Cys	Ile	Arg
65	Gly	Gly	Lys	His	Asn	Asp	70	Leu	Asp	Asp	Val	75	Gly	Lys	Asp	Val	Tyr
				85							90					95	
His	Thr	Phe	Phe	100	Glu	Met	Leu	Gly	Ser	105	Trp	Ser	Phe	Gly	Asp	110	Tyr
Lys	Glu	Leu	Ala	Cys	Lys	Met	Ala	Leu	Glu	Leu	Leu	Thr	Gln	Glu	Phe		
		115					120						125				
Gly	Ile	Pro	Ile	Glu	Arg	Leu	Tyr	Val	Thr	Tyr	Phe	140	Gly	Gly	Asp	Glu	
130	Ala	Ala	Gly	Leu	Glu	Ala	150	Asp	Leu	Glu	Cys	Lys	155	Gln	Ile	Trp	Gln
145	Leu	Gly	Leu	Asp	Asp	Thr	Lys	Ile	Leu	Pro	Gly	Asn	Met	Lys	Asp	Asn	
				165							170				175		
Phe	Trp	Glu	Met	Gly	Asp	Thr	Gly	Pro	Cys	Gly	Pro	Cys	Ser	190	Glu	Ile	
			180					185									
His	Tyr	Asp	Arg	Ile	Gly	Gly	Arg	200	Asp	Ala	Ala	His	Leu	Val	Asn	Gln	
		195											205				
Asp	Asp	Pro	Asn	Val	Leu	Glu	Ile	Trp	Asn	Leu	Val	Phe	Ile	Gln	Tyr		
210						215						220					
Asn	Arg	Ala	Gln	Val	Pro	Asp	His	Thr	Leu	Gly	Lys	Leu	Val	Leu	Arg		
225	Met	Pro	Met	Gly	Leu	Thr	Trp	Pro	Thr	Gly	Cys	Trp	Leu	Thr	Thr	Leu	
				245						250					255		
Gly	Pro	Ser	Leu	Trp	His	Trp	Leu	Met	Val	Ala	Gly	Leu	Thr	Thr	Gln		
			260					265						270			
Gly	Val	Asp	Met	Cys	Leu	Glu	Gly	Lys	Pro	Ile	Pro	Asn	Pro	Leu	Leu		
		275					280					285					
Gly	Leu	Asp	Ser	Thr	His	His	His	His	His	His							
290						295											

<210> 69

<211> 541

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanyl tRNA Synthetase polypeptide with N-terminal 6xHis affinity tag

<400> 69

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

120161\_426PC\_Sequence\_Listing.TXT

Asp	Asn	Phe	180	Trp	Glu	Met	Gly	Asp	185	Thr	Gly	Pro	Cys	Gly	190	Pro	Cys	Ser
Glu	Ile	His	195	Tyr	Asp	Arg	Ile	Gly	200	Gly	Arg	Asp	Ala	Ala	205	His	Leu	Val
Asn	Gln	Asp	210	Asp	Pro	Asn	Val	Leu	215	Glu	Ile	Trp	Asn	Leu	220	Val	Phe	Ile
225	Gln	Tyr	Asn	Arg	Glu	Ala	Asp	Gly	230	Ile	Leu	Lys	Pro	Leu	235	Pro	Lys	Lys
Ser	Ile	Asp	240	Thr	Gly	Met	Gly	Leu	245	Glu	Arg	Leu	Val	Ser	250	Val	Leu	Gln
Asn	Lys	Met	255	Ser	Asn	Tyr	Asp	Thr	260	Asp	Leu	Phe	Val	Pro	265	Tyr	Phe	Glu
Ala	Ile	Gln	270	Lys	Gly	Thr	Gly	Ala	275	Arg	Pro	Tyr	Thr	Gly	280	Lys	Val	Gly
Ala	Glu	Asp	285	Ala	Asp	Gly	Ile	Asp	290	Met	Ala	Tyr	Arg	Val	295	Leu	Ala	Asp
305	His	Ala	Arg	Thr	Ile	Thr	Val	Ala	310	Leu	Ala	Asp	Gly	Gly	315	Arg	Pro	Asp
Asn	Thr	Gly	320	Arg	Gly	Tyr	Val	Leu	325	Arg	Ile	Leu	Arg	Arg	330	Ala	Val	
Arg	Tyr	Ala	335	His	Glu	Lys	Leu	Asn	340	Ala	Ser	Arg	Gly	Phe	345	Phe	Ala	Thr
Leu	Val	Asp	350	Val	Val	Val	Gln	Ser	355	Leu	Gly	Asp	Ala	Phe	360	Pro	Glu	Leu
Lys	Lys	Asp	365	Pro	Asp	Met	Val	Lys	370	Asp	Ile	Ile	Asn	Glu	375	Glu	Glu	Val
Gln	Phe	Leu	380	Lys	Thr	Leu	Ser	Arg	385	Gly	Arg	Arg	Ile	Leu	390	Asp	Arg	Lys
Ile	Gln	Ser	395	Leu	Gly	Asp	Ser	Lys	400	Thr	Ile	Pro	Gly	Asp	405	Thr	Ala	Trp
Leu	Leu	Tyr	410	Asp	Thr	Tyr	Gly	Phe	415	Pro	Val	Asp	Leu	Thr	420	Gly	Leu	Ile
Ala	Glu	Glu	425	Lys	Gly	Leu	Val	Val	430	Asp	Met	Asp	Gly	Phe	435	Glu	Glu	Glu
Arg	Lys	Leu	440	Ala	Gln	Leu	Lys	Ser	445	Gln	Gly	Lys	Gly	Ala	450	Gly	Gly	Glu
Asp	Leu	Ile	455	Met	Leu	Asp	Ile	Tyr	460	Ala	Ile	Glu	Glu	Leu	465	Arg	Ala	Arg
Gly	Leu	Glu	465	Val	Thr	Asp	Asp	Ser	470	Pro	Lys	Tyr	Asn	Tyr	475	His	Leu	Asp
Ser	Ser	Gly	480	Ser	Tyr	Glu	Asn	Arg	485	Val	Tyr	Ser	Glu	Glu	490	Cys	Ser	Gly
Pro	Arg	Arg	495	Val	Cys	Ala	Thr	His	500	Trp	Asn	His	Leu	Arg	505			
			510						515						520			
			525						530						535			

<210> 70

<211> 541

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanyl tRNA Synthetase polypeptide with C-terminal 6xHis affinity tag

<400> 70

Met	Asp	Ser	Thr	Leu	Thr	Ala	Ser	Glu	Ile	Arg	Gln	Arg	Phe	Ile	Asp
1	Phe	Phe	Lys	Arg	Asn	Glu	His	Thr	25	Val	His	Ser	Ser	Ala	Thr
			20	Asp	Pro	Thr	Leu	Leu	30	Phe	Ala	Asn	Ala	Gly	Met
			35	Pro	Ile	Phe	Leu	Asn	40	Ile	Asp	Pro	Ser	His	Pro
			50	Lys	Leu	Ser	Arg	Ala	55	Asn	Thr	Gln	Lys	Cys	Ile
			65	Lys	His	Asn	Asp	Leu	70	Asp	Val	Gly	Lys	Asp	Val
									75	Tyr	His	His	Thr		

## 120161\_426PC\_Sequence\_Listing.TXT

Phe	Phe	Gl u	Met 100	85	Leu	Gl y	Ser	Trp	Ser 105	90	Phe	Gl y	Asp	Tyr	Phe 110	95	Lys	Gl u
Leu	Al a	Cys 115	Lys	Met	Al a	Leu	Gl u	120	Leu	Leu	Thr	Gl n	Gl u	125	Phe	Gl y	I le	
Pro	I le	Gl u	Arg	Leu	Tyr	Val 135	Thr	Tyr	Phe	Gl y	Gl y	140	Asp	Gl u	Al a	Al a		
Gl y	Leu	Gl u	Al a	Asp	Leu 150	Gl u	Cys	Lys	Gl n	I le	155	Trp	Gl n	Asn	Leu	Gl y	160	
Leu	Asp	Asp	Thr	Lys 165	I le	Leu	Pro	Gl y	Asn 170	Met	Lys	Asp	Asn	Phe 175	Trp			
Gl u	Met	Gl y	Asp	Thr	Gl y	Pro	Cys	Gl y	185	Pro	Cys	Ser	Gl u	I le	Hi s	Tyr		
Asp	Arg	I le	Gl y	Gl y	Arg	Asp	Al a	200	Al a	Hi s	Leu	Val	Asn	205	Gl n	Asp	Asp	
Pro	Asn	Val	Leu	Gl u	I le	Trp 215	Asn	Leu	Val	Phe	I le	220	Gl n	Tyr	Asn	Arg		
Gl u	Al a	Asp	Gl y	I le	Leu 230	Lys	Pro	Leu	Pro	Lys	Lys	Ser	I le	Asp	Thr	240		
Gl y	Met	Gl y	Leu	Gl u	Arg	Leu	Val	Ser	Val	250	Leu	Gl n	Asn	Lys	Met	255	Ser	
Asn	Tyr	Asp	Thr	260	Asp	Leu	Phe	Val	Pro	265	Tyr	Phe	Gl u	Al a	I le	270	Lys	
Gl y	Thr	Gl y	Al a	Arg	Pro	Tyr	Thr	280	Gl y	Lys	Val	Gl y	Al a	285	Gl u	Asp	Al a	
Asp	Gl y	I le	Asp	Met	Al a	Tyr	295	Arg	Val	Leu	Al a	Asp	300	Hi s	Al a	Arg	Thr	
I le	Thr	Val	Al a	Leu	Al a	310	Asp	Gl y	Gl y	Arg	Pro	315	Asp	Asn	Thr	Gl y	Arg	320
Gl y	Tyr	Val	Leu	Arg	Arg	I le	Leu	Arg	Arg	330	Al a	Val	Arg	Tyr	Al a	Hi s		
Gl u	Lys	Leu	Asn	Al a	Ser	Arg	Gl y	Phe	345	Phe	Al a	Thr	Leu	Val	350	Asp	Val	
Val	Val	Gl n	Ser	Leu	Gl y	Asp	Al a	360	Phe	Pro	Gl u	Leu	Lys	365	Lys	Asp	Pro	
Asp	Met	Val	Lys	Asp	I le	I le	375	Asn	Gl u	Gl u	Gl u	Val	380	Gl n	Phe	Leu	Lys	
Thr	Leu	Ser	Arg	Gl y	Arg	390	I le	Leu	Asp	Arg	395	Lys	I le	Gl n	Ser	Leu	400	
Gl y	Asp	Ser	Lys	Thr	I le	Pro	Gl y	Asp	Thr	410	Al a	Trp	Leu	Leu	Tyr	415	Asp	
Thr	Tyr	Gl y	Phe	Pro	Val	Asp	Leu	Thr	425	Gl y	Leu	I le	Al a	Gl u	Gl u	Lys		
Gl y	Leu	Val	Val	Asp	Met	Asp	Gl y	440	Phe	Gl u	Gl u	Gl u	Arg	445	Lys	Leu	Al a	
Gl n	Leu	Lys	Ser	Gl n	Gl y	Lys	455	Gl y	Al a	Gl y	Gl y	Gl u	Asp	460	Leu	I le	Met	
Leu	Asp	I le	Tyr	Al a	I le	470	Gl u	Gl u	Leu	Arg	Al a	Arg	Gl y	475	Leu	Gl u	Val	
Thr	Asp	Asp	Ser	Pro	Lys	Tyr	Asn	Tyr	Hi s	490	Leu	Asp	Ser	Ser	Gl y	495	Ser	
Tyr	Gl u	Asn	Arg	Val	Tyr	Ser	Gl u	Gl u	505	Cys	Ser	Gl y	Pro	Arg	510	Arg	Val	
Cys	Al a	Thr	Hi s	Trp	Asn	Hi s	Leu	520	Arg	Gl y	Lys	Pro	I le	525	Pro	Asn	Pro	
Leu	Leu	Gl y	Leu	Asp	Ser	Thr	535	Hi s	Hi s	Hi s	Hi s	Hi s	Hi s	540				

<210> 71

<211> 351

<212> PRT

<213> Artificial Sequence

$\langle 220 \rangle$

<223> Alanyl tRNA Synthetase polypeptide with N-terminal 6xHis affinity tag

120161\_426PC\_Sequence\_Listing.TXT

<400> 71

```

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

```

<210> 72

<211> 351

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanyl tRNA Synthetase polypeptide with C-terminal  
6xHis affinity tag

<400> 72

```

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

```



## 120161\_426PC\_Sequence\_Listing.TXT

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

<210> 73

$\langle 211 \rangle$  1212

<212> DNA

## <213> Artificial Sequence

 $\langle 220 \rangle$ 

<223> Codon optimized Alanyl tRNA Synthetase  
polynucleotide

<400> 73

gacagcacgc	tgaccgcatac	cgaatatccgc	caacgcttca	ttgacttttt	caaacgtaac	60
gagcatactt	acgtccatag	cagcgcaacc	attccgctgg	atgacccaac	cctgctgttt	120
gccaacgctg	gcataaatca	attcaagccg	attttcctga	atactatcga	cccagaccac	180
ccgatggcga	aactgtctcg	tgcggcgaat	acgcaaaagt	gcataccgtgc	gggtggtaag	240
cacaatgact	tggatgatgt	tggtaaggac	gtttaccacc	ataccttttt	cgaaatgctg	300
ggttctctga	gcttttgtga	ctactttaaa	gagctggcgt	gtaaaatggc	tttggagttg	360
ctgacgcagg	agttcggcat	tccgattgag	cgctgtacg	tgacctactt	cgttggcgat	420
gaggcggcag	gtctggaggc	agacctggag	tgcaaacaga	tttggcagaa	tctgggcctg	480
gacgacacca	aaatcttgcc	gggtaacatg	aaagataact	tctgggaaat	gggcgatacc	540
ggtccgtgcg	gtccgtgtag	cgagattcac	tatgatcgta	ttggtggtcg	cgatgccgcc	600
cacctggtta	accaagatga	tcctaacgtc	ctggagattt	ggaatctggt	gtttatccaa	660
tacaaccgcg	aggcggatgg	cattctgaag	ccgctgccaa	agaaaagcat	cgatacgggt	720
atgggcctgg	aacgcctgtg	gagcgtgtcg	cagaataaag	tgagcaatta	tgacacggat	780
ctgtttgtcc	cgtatttcga	agcaattccag	aagggtaccg	gtgcgcgtcc	gtacaccggt	840
aaagtgggcg	ctgaggacgc	ggacgggtatt	gacatggcct	accgtgtgtt	ggctgatcac	900
gcgcgcacca	tcacggtcgc	cctggcagat	ggtggccgtc	cggacaacac	cggtcgtggt	960
tatgtactgc	gtcgtatcct	gcgtcgcgcg	gtccggttatg	cacatgaaaa	gctgaacgcg	1020
tcgcgtggct	tcttcgccac	gctggtcgat	gttgttgtgc	agagcctggg	cgatgcgttt	1080
ccggagctga	agaaagaccc	ggacatggtt	aaggacatca	tcaacgaaga	agaagttcag	1140
tttctgaaaa	ccttgagccg	tggccgctcg	attctggacc	gcaaaatcca	atctttgggc	1200
taatgactcg	ag					1212

<210> 74

<211> 867  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Codon optimized Alanyl tRNA Synthetase  
 polynucleotide

<400> 74  
 ggatccgact caaccctgac agcaagcgaa attcgtcagc gctttatcga cttcttcaaa 60  
 cgtaacgagc acacctatgt ccatagcagt gccacaattc cgctggatga ccctaccctg 120  
 ctgtttgcta atgccggcat gaatcagttc aaaccgattt ttctgaatac catcgaccgc 180  
 tcacatccaa tggcgaaact gtctcgtgcc gctaatactc aaaaatgtat ccgtgccggg 240  
 ggtaaacaca atgacctgga tgatgtgggc aaagacgtgt atcaccacac gttcttcgag 300  
 atgctgggtt cttggctctt tggggactat ttcaaagaac tggcctgtaa aatggccctg 360  
 gaactgctga cacaggagtt tgggaattccg attgaacgcc tgtatgtcac ctattttggc 420  
 ggggatgaag ctgccggtct ggaagcagac ctggagtgtg aacaaatctg gcagaacctg 480  
 ggactggatg atacgaaaat cctgccgggc aacatgaaag acaacttttg ggaaatgggc 540  
 gatactggcc catgtggacc ttgttccgaa atccattatg accgtatcgg tggctcgtgat 600  
 gccgctcatc tggtaataca ggaatgacctg aatgttctgg aaatctggaa cctgggtgtc 660  
 attcagtata accgtgaggg tgatggcatc ctgaaacctc tgccgaaaaa atccatcgat 720  
 accggtatgg ggctggaacg cctggtttct gtactgcaaa acaaaatgtc gaactatgac 780  
 accgacctgt ttgtcccgtt tttcgaagcc attcaaaaag gcaccgggtg ccgtccttat 840  
 acagggaag ttggcgccga actcgag 867

<210> 75  
 <211> 1473  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Codon optimized Alanyl tRNA Synthetase  
 polynucleotide

<400> 75  
 gacagcacc tgaccgcatc cgaaatccgc cagcgcttca ttgacttttt caagcgcaac 60  
 gagcatacct atgtgcatag cagcgaacc atcccgttg atgacacct gttgctgttt 120  
 gcgaatgcgg gcatgaatca gttcaaacc atctttttga acacgattga cccaagccat 180  
 ccgatggcta aactgagccg tgccgcta atccagaagt gcattcgtgc ggggtggaag 240  
 cacaacgact tggacgacgt gggttaaggat gtgtatcacc acacgttctt tgagatgctg 300  
 ggcagctggg ccttcggcga ttactttaaa gaactggcgt gcaagatggc actggaactg 360  
 ttgacccaag aatttggcat cccgattgag cgtctgtacg ttacctact tggcggcgat 420  
 gaggctgacg gcctggaggc ggacctggag tgcaaacaga tctggcaaaa tctgggcctg 480  
 gatgacacca agatcctgcc gggaatatg aagataact tctgggagat gggcgacacg 540  
 ggtccgtgtg gcccgtgttc ggagatccac tacgaccgca ttggtggtcg tgacgcagcg 600  
 catctgggtg atcaagatga cccgaatgtg ctggagatct ggaatttggg ttcatccaa 660  
 tacaaccgtg aggcggatgg tattctgaaa ccgttgccga aaaagtctat cgacaccggc 720  
 atgggtctgg agcgtctggg tagcgtgctg cagaacaaaa tgagcaacta cgatactgat 780  
 ctgtttgtgc cgtatttcga ggcgatccga aaggtaccg gtgccgtcc gtacacgggt 840  
 aaggttgggt cggaggatgc tgacggatc accgtgtctt ggccgatcac 900  
 gcccgacca ttaccgtcgc gttggctgat ggtggtcgcc cggacaacac cggtcgcggg 960  
 tatgttctgc gtcgtatcct gcgtcgtgca gtccgctatg cgcacgagaa actgaatgcg 1020  
 agccgtggct tctttgccac cctggctgac gttgttgttc agtccctggg cgatgcgttc 1080  
 ccggaactga agaaagatcc ggaatggta aaggacatta tcaacgaaga agaggtccag 1140  
 ttcctgaaaa cctgtctcg tggccgtcgt attctggacc gcaagatcca aagcctgggt 1200  
 gacagcaaaa cgattccagg tgatacggcg tggctgctgt acgacacgta tggtttccct 1260  
 gtcgacctga cgggtctgat tgcggaagaa aagggcctgg tcgttgacat ggacggtttt 1320  
 gaagaagagc gcaagctggc ccaattgaaa agccagggtg aaggtgcagg tggcgaagat 1380  
 ctgattatgc tggacattta tgcaattgaa gagctgcgtg cgcgtggcct ggaggtgacc 1440  
 gatgatagcc cgaaatacaa ctaatgactc gag 1473

<210> 76  
 <211> 834  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Codon optimized Alanyl tRNA Synthetase

## pol ynucl eoti d

&lt;400&gt; 76

```

ggatccgata gtactctgac ggcaagtga atccgtcaac gctttatcga cttcttttaa 60
cgtaatgagc acacgtatgt ccactctagt gccaccatcc ctctggatga tccaaccctg 120
ctgttcgcta acgctgggat gaaccagttt aaaccaatct tcctgaacac gattgatccg 180
agccatccta tggcgaaact gtctcgtgcc gctaatactc aaaaatgtat ccgtgctggc 240
ggaaaacaca atgatctgga tgacgtgggc aaagatgttt atcatcacac ctttttcgaa 300
atgctgggca gctggctctt cggcgattat ttcaaagaac tggcctgtaa gatggccctg 360
gagctgctga cacaagagtt cggtatccct atcgaacgtc tgtatgtcac gtattttggg 420
ggtgacgaag cagctggcct ggaggtgat ctggagtgt aacaaatctg gcagaatctg 480
ggtctggacg atacaaaaat cctgcctggc aatatgaaag acaatttctg ggagatgggc 540
gacacaggtc catgtggtcc atgctctgag attcactatg accgtatcgg aggacgtgac 600
gctgctcacc tggtaaacca agacgacctg aacgtcctgg aaatctggaa tctggctctt 660
attcagtata accgtgcccc agtgccggac catactctgg gcaaactggg actgctgatg 720
ccaatgggcc tgacttggcc tacaggttgt tggctgacaa ctctgggacc ttctctgtgg 780
cattggctga tggttgctgg tctgactacc cagggtgttg atatgtgtct cgag 834

```

&lt;210&gt; 77

&lt;211&gt; 1572

&lt;212&gt; DNA

&lt;213&gt; Arti f i c i a l   S e q u e n c e

&lt;220&gt;

<223> Codon opti mi zed Al anyl tRNA Synthetase  
pol ynucl eoti d

&lt;400&gt; 77

```

gacagcacgc tgaccgcatc cgaaatccgc caacgcttca ttgacttctt caaacgtaat 60
gagcacacgt acgttcatag cagcgcgacc atccctttgg acgatccgac cctgctgttc 120
gctaacgccg gatgaatca gttaaagccg atttttctga atacgatcga cccaagccac 180
ccgatggcga aactgtcccc tgccgcaaac acccagaaat gtattcgtgc cggtaggtaaa 240
cataacgatc tggatgacgt gggtaaagac gtttaccacc acaccttctt cgagatgctg 300
ggtagctggg cttttggcga ctatttcaaa gagctggcgt gcaagatggc cctggagctg 360
ctgacgcagg agtttggtat tccgattgag cgtctgtatg tcacctattt cggtaggtgat 420
gaagcagcgg gcttggaggc ggaatctggaa tgcaagcaaa tctggcagaa ttggggcctg 480
gatgacacca agattctgcc gggcaatatg aaagataact tttgggaaat gggtagacact 540
ggcccgctgt gcccgtgctc tgagattcac tatgatcgca tcggtaggccc cgacgcggct 600
catctgggta accaagatga cccaacgtg ttggagattt ggaatctggg gttcattcag 660
tacaaccgag aggcggatgg catcctgaag ccgctgccga aaaagagcat cgacacgggt 720
atgggtctgg agcgcctggg gagcgtcctg caaaacaaga tgagcaatta cgatacggat 780
ctgttcgtcc cgtacttcca ggcgattcaa aaggtaccg gtgcacgccc gtacacgggt 840
aaagtgtgtg ccgaggatgc agacggcatt gacatggctt accgtgttct ggctgaicat 900
gcgcgtacca ttaccgtggc gctggcagac ggtggctgcc cggacaacac cggtcgtggg 960
tatgttctgc gtgcacacct gcgtcgtgcc gtccgttatg cgcacgaaaa actgaacgca 1020
tcgctgtggc ttttcgcgac tctggttgac gtgggtggtc agtctttggg tgacgcaitt 1080
ccggagttag agaaagatcc ggacatggtc aaagacatta tcaatgaaga agaagtccag 1140
tttctgaaaa cgctgagccg tggccgtcgc atcttggatc gcaagatcca aagcctgggc 1200
gatagaaga cgattccggg cgataccgag tggctgctgt atgataccta cggttttccg 1260
gtggacctga ccggcctgat ccggaagag aagggcctgg tcgttgatat ggatggcttt 1320
gaagaagaac gtaagctggc gcagctgaaa agccaaggta aaggtgctgg tggcgaagat 1380
ctgatcatgc tggacatcta cgcaatcgaa gagctgcgtg cgcgtgggtt ggaggtcacc 1440
gacgacagcc cgaagtacaa ctaccacctg gacagcagcg gcagctacga gaatcgtgtt 1500
tatagcgaag agtgctccgg tcctcgtcgc gtgtgtgcc a ctcactggaa tcatttgcgt 1560
taatgactcg ag 1572

```

&lt;210&gt; 78

&lt;211&gt; 990

&lt;212&gt; DNA

&lt;213&gt; Arti f i c i a l   S e q u e n c e

&lt;220&gt;

<223> Codon opti mi zed Al anyl tRNA Synthetase  
pol ynucl eoti d

&lt;400&gt; 78

```

ggatccgact ctacactgac cgcctctgaa attcgccagc gctttatcga cttcttcaaa 60
cgtaacgagc acacctatgt tcattctagt gccacgattc ctctggatga tccgacactg 120

```

120161\_426PC\_Sequence\_Listing.TXT

```

ctgttcgcta atgccggtat gaatcagttt aaaccgatct tcctgaatac catcgatcca 180
agccacccta tggcaaaact gtctcgtgct gccaatactc aaaaatgtat ccgtgccggg 240
ggtaaacaca atgatctgga cgacgtgggc aaagatgtgt atcaccacac gttcttcgaa 300
atgctgggtt catggagctt tggggactat ttcaaagagc tggcctgtaa aatggctctg 360
gaactgctga cacaggagtt tgggaattccg attgaacgct tgtatgtgac gtaattttggc 420
ggggatgaag cagccggtct ggaggctgac ctggagtgtg aacaaatctg gcagaacctg 480
ggactggatg acacgaaaat cctgcctggc aacatgaaag acaacttttg ggagatgggt 540
gacactggtc cttgtggacc ttgctccgaa atccactatg accgtatcgg tggtcgtgat 600
gctgctcacc tgggtaatca ggatgaccct aacgttctgg aaatttggaa cctgggtttt 660
atccagtata accgtgaggg agatgggatt ctgaaaccgc tgcctaaaaa aagcatcgac 720
acgggtatgg gactggaacg cctgggttct gtactgcaga acaaaatgag caactatgat 780
accgacctgt ttgtgcctta tttcgaggcc attcagaaag gcactggagc acgtccttat 840
acagggaag tgggcgctga agatgctgat ggcattgata tggcgtatcg tgtactggct 900
gatcatgccc gtaccattac cgtagccctg gctgacggtg gacgtccgga taatacgggg 960
cgtggggaaa tgcattttct gacgctcgag

```

<210> 79  
 <211> 493  
 <212> PRT  
 <213> Homo sapiens

<400> 79

Arg	Gly	Leu	Glu	Val	Thr	Asp	Asp	Ser	Pro	Lys	Tyr	Asn	Tyr	His	Leu
1				5					10					15	
Asp	Ser	Ser	Gly	Ser	Tyr	Val	Phe	Glu	Asn	Thr	Val	Ala	Thr	Val	Met
			20					25					30		
Ala	Leu	Arg	Arg	Glu	Lys	Met	Phe	Val	Glu	Glu	Val	Ser	Thr	Gly	Gln
		35					40					45			
Glu	Cys	Gly	Val	Val	Leu	Asp	Lys	Thr	Cys	Phe	Tyr	Ala	Glu	Gln	Gly
	50					55					60				
Gly	Gln	Ile	Tyr	Asp	Glu	Gly	Tyr	Leu	Val	Lys	Val	Asp	Asp	Ser	Ser
65					70					75				80	
Glu	Asp	Lys	Thr	Glu	Phe	Thr	Val	Lys	Asn	Ala	Gln	Val	Arg	Gly	Gly
				85					90				95		
Tyr	Val	Leu	His	Ile	Gly	Thr	Ile	Tyr	Gly	Asp	Leu	Lys	Val	Gly	Asp
			100					105					110		
Gln	Val	Trp	Leu	Phe	Ile	Asp	Glu	Pro	Arg	Arg	Arg	Pro	Ile	Met	Ser
		115					120					125			
Asn	His	Thr	Ala	Thr	His	Ile	Leu	Asn	Phe	Ala	Leu	Arg	Ser	Val	Leu
	130				135					140					
Gly	Glu	Ala	Asp	Gln	Lys	Gly	Ser	Leu	Val	Ala	Pro	Asp	Arg	Leu	Arg
145					150					155				160	
Phe	Asp	Phe	Thr	Ala	Lys	Gly	Ala	Met	Ser	Thr	Gln	Gln	Ile	Lys	Lys
				165					170					175	
Ala	Glu	Glu	Ile	Ala	Asn	Glu	Met	Ile	Glu	Ala	Ala	Lys	Ala	Val	Tyr
			180				185						190		
Thr	Gln	Asp	Cys	Pro	Leu	Ala	Ala	Ala	Lys	Ala	Ile	Gln	Gly	Leu	Arg
		195					200					205			
Ala	Val	Phe	Asp	Glu	Thr	Tyr	Pro	Asp	Pro	Val	Arg	Val	Val	Ser	Ile
	210					215					220				
Gly	Val	Pro	Val	Ser	Glu	Leu	Leu	Asp	Asp	Pro	Ser	Gly	Pro	Ala	Gly
225					230					235				240	
Ser	Leu	Thr	Ser	Val	Glu	Phe	Cys	Gly	Gly	Thr	His	Leu	Arg	Asn	Ser
				245					250					255	
Ser	His	Ala	Gly	Ala	Phe	Val	Ile	Val	Thr	Glu	Glu	Ala	Ile	Ala	Lys
			260					265					270		
Gly	Ile	Arg	Arg	Ile	Val	Ala	Val	Thr	Gly	Ala	Glu	Ala	Gln	Lys	Ala
	275					280						285			
Leu	Arg	Lys	Ala	Glu	Ser	Leu	Lys	Lys	Cys	Leu	Ser	Val	Met	Glu	Ala
	290					295					300				
Lys	Val	Lys	Ala	Gln	Thr	Ala	Pro	Asn	Lys	Asp	Val	Gln	Arg	Glu	Ile
305					310					315				320	
Ala	Asp	Leu	Gly	Glu	Ala	Leu	Ala	Thr	Ala	Val	Ile	Pro	Gln	Trp	Gln
				325					330					335	
Lys	Asp	Glu	Leu	Arg	Glu	Thr	Leu	Lys	Ser	Leu	Lys	Lys	Val	Met	Asp
			340					345					350		
Asp	Leu	Asp	Arg	Ala	Ser	Lys	Ala	Asp	Val	Gln	Lys	Arg	Val	Leu	Glu
	355					360						365			

120161\_426PC\_Sequence\_Listing.TXT

Lys	Thr	Lys	Gln	Phe	Ile	Asp	Ser	Asn	Pro	Asn	Gln	Pro	Leu	Val	Ile
370						375					380				
Leu	Glu	Met	Glu	Ser	Gly	Ala	Ser	Ala	Lys	Ala	Leu	Asn	Glu	Ala	Leu
385					390					395					400
Lys	Leu	Phe	Lys	Met	His	Ser	Pro	Gln	Thr	Ser	Ala	Met	Leu	Phe	Thr
				405					410					415	
Val	Asp	Asn	Glu	Ala	Gly	Lys	Ile	Thr	Cys	Leu	Cys	Gln	Val	Pro	Gln
			420					425					430		
Asn	Ala	Ala	Asn	Arg	Gly	Leu	Lys	Ala	Ser	Glu	Trp	Val	Gln	Gln	Val
		435					440					445			
Ser	Gly	Leu	Met	Asp	Gly	Lys	Gly	Gly	Gly	Lys	Asp	Val	Ser	Ala	Gln
	450					455					460				
Ala	Thr	Gly	Lys	Asn	Val	Gly	Cys	Leu	Gln	Glu	Ala	Leu	Gln	Leu	Ala
465					470					475					480
Thr	Ser	Phe	Ala	Gln	Leu	Arg	Leu	Gly	Asp	Val	Lys	Asn			
				485					490						

<210> 80  
 <211> 1482  
 <212> DNA  
 <213> Homo sapiens

<400> 80  
 cggggctctgg aggtcacaga tgattcccca aagtacaatt accatttggga ctccagtgggt 60  
 agctatgtat ttgagaacac agtggctacg gtgatggctc tgcgcagggga gaagatgttc 120  
 gtggaagagg tgtccacagg ccaggagtgt ggagtgggtgc tggacaagac ctgtttctat 180  
 gctgagcaag gaggccagat ctatgacgaa ggctacctgg tgaaggtgga tgacagcagt 240  
 gaagataaaa cagagtttac agtgaagaat gctcaggtcc gaggagggtta tgtgctacac 300  
 attggaacca tctacggtga cctgaaagtg ggggatcagg tctggctgtt tattgatgag 360  
 ccccgacgaa gacctatcat gagcaaccac acagctacgc acattctgaa cttcgccctg 420  
 cgctcagtcg ttggggaagc tgaccagaaa ggcctattgg ttgctcctga ccgcctcaga 480  
 ttgacttta ctgccaaggg agccatgtcc acccaacaga tcaagaaggc tgaagagatt 540  
 gctaatgaga tgattgaggc agccaaggcc gtctataccc aggattgccc cctggcagca 600  
 gcgaaagcca tccagggcct acgggctgtg tttgatgaga cctatcctga ccctgtgcga 660  
 gtcgtctcca ttgggggtccc ggtgtccgag ttgctggatg acccctctgg gcctgttggc 720  
 tccctgactt ctgttgagtt ctgtggggga acgcacctgc ggaactcgag tcatgcagga 780  
 gcttttgtag tcgtgacgga agaagccatt gccaaaggta tccggaggat tgtggctgtc 840  
 acaggtgccg agggccagaa ggccctcagg aaagcagaga gcttgaagaa atgtctctct 900  
 gtcatggaag ccaaagtga ggctcagact gctccaaaca aggatgtgca gagggagatc 960  
 gctgaccttg gagaggccct ggccactgca gtcatcccc agtggcagaa ggatgaattg 1020  
 cgggagactc tcaaatccct aaagaaggta atggatgact tggaccgagc cagcaaagcc 1080  
 gatgtccaga aacgagtgtt agagaagacg aagcagttca tcgacagcaa cccaaccag 1140  
 cctcttgta tcctggagat ggagagcggc gcctcagcca aggccctgaa tgaagccttg 1200  
 aagctcttca agatgcactc ccctcagact tctgccatgc tcttcacggt ggacaatgag 1260  
 gctggcaaga tcacgtgcct gtgtcaagtt ccccagaatg cagccaatcg gggcttaaaa 1320  
 gccagcgagt ggggtcagca ggtgtcaggc ttgatggacg gtaaagggtg tggcaaggat 1380  
 gtgtctgcac aggccacagg caagaacgtt ggctgcctgc aggaggcgt gcagctggcc 1440  
 acttccttcg cccagctgcg cctcgggat gtaagaact ga 1482

<210> 81  
 <211> 211  
 <212> PRT  
 <213> Homo sapiens

<400> 81  
 Ala Glu Ala Gln Lys Ala Leu Arg Lys Ala Glu Ser Leu Lys Lys Cys  
 1 5 10 15  
 Leu Ser Val Met Glu Ala Lys Val Lys Ala Gln Thr Ala Pro Asn Lys  
 20 25 30  
 Asp Val Gln Arg Glu Ile Ala Asp Leu Gly Glu Ala Leu Ala Thr Ala  
 35 40 45  
 Val Ile Pro Gln Trp Gln Lys Asp Glu Leu Arg Glu Thr Leu Lys Ser  
 50 55 60  
 Leu Lys Lys Val Met Asp Asp Leu Asp Arg Ala Ser Lys Ala Asp Val  
 65 70 75 80  
 Gln Lys Arg Val Leu Glu Lys Thr Lys Gln Phe Ile Asp Ser Asn Pro  
 85 90 95

120161\_426PC\_Sequence\_Listing.TXT

Asn	Gln	Pro	Leu	Val	Ile	Leu	Glu	Met	Glu	Ser	Gly	Ala	Ser	Ala	Lys
			100					105					110		
Ala	Leu	Asn	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Met	His	Ser	Pro	Gln	Thr
		115					120					125			
Ser	Ala	Met	Leu	Phe	Thr	Val	Asp	Asn	Glu	Ala	Gly	Lys	Ile	Thr	Cys
		130				135					140				
Leu	Cys	Gln	Val	Pro	Gln	Asn	Ala	Ala	Asn	Arg	Gly	Leu	Lys	Ala	Ser
145					150				155						160
Glu	Trp	Val	Gln	Gln	Val	Ser	Gly	Leu	Met	Asp	Gly	Lys	Gly	Gly	Gly
				165					170					175	
Lys	Asp	Val	Ser	Ala	Gln	Ala	Thr	Gly	Lys	Asn	Val	Gly	Cys	Leu	Gln
			180					185					190		
Glu	Ala	Leu	Gln	Leu	Ala	Thr	Ser	Phe	Ala	Gln	Leu	Arg	Leu	Gly	Asp
		195					200					205			
Val	Lys	Asn													
		210													

<210> 82  
 <211> 636  
 <212> DNA  
 <213> Homo sapiens

<400> 82																
gccgaggccc	agaaggccct	caggaaagca	gagagcttga	agaaatgtct	ctctgtcatg	60										
gaagccaaag	tgaaggctca	gactgctcca	aacaaggatg	tgcagaggga	gatcgctgac	120										
cttgagaggg	ccctggccac	tgcatgcatc	ccccagtggc	agaaggatga	attgcgggag	180										
actctcaa	ccctaaagaa	ggtcatggat	gacttggacc	gagccagcaa	agccgatgtc	240										
cagaaacgag	tgtagagaa	gacgaagcag	ttcatcgaca	gcaaccccaa	ccagcctctt	300										
gtcatcctgg	agatggagag	cggcgccctca	gccaaaggccc	tgaatgaagc	cttgaagctc	360										
ttcaagatgc	actccccfca	gacttctgcc	atgctcttca	cggtggacaa	tgaggctggc	420										
aagatcacgt	gcctgtgtca	agttccccag	aatgcagcca	atcggggctt	aaaagccagc	480										
gagtgggtgc	agcaggtgtc	aggcttgatg	gacggtaaag	gtggtggcaa	ggatgtgtct	540										
gcacaggcca	caggcaagaa	cgttggctgc	ctgcaggagg	cgctgcagct	ggccacttcc	600										
ttcgcccagc	tgcgccctcg	ggatgtaaag	aactga			636										

<210> 83  
 <211> 381  
 <212> PRT  
 <213> Homo sapiens

<400> 83																
Gln	Val	Trp	Leu	Phe	Ile	Asp	Glu	Pro	Arg	Arg	Arg	Pro	Ile	Met	Ser	
1				5					10					15		
Asn	His	Thr	Ala	Thr	His	Ile	Leu	Asn	Phe	Ala	Leu	Arg	Ser	Val	Leu	
			20					25					30			
Gly	Glu	Ala	Asp	Gln	Lys	Gly	Ser	Leu	Val	Ala	Pro	Asp	Arg	Leu	Arg	
		35					40					45				
Phe	Asp	Phe	Thr	Ala	Lys	Gly	Ala	Met	Ser	Thr	Gln	Gln	Ile	Lys	Lys	
	50					55					60					
Ala	Glu	Glu	Ile	Ala	Asn	Glu	Met	Ile	Glu	Ala	Ala	Lys	Ala	Val	Tyr	
65					70				75						80	
Thr	Gln	Asp	Cys	Pro	Leu	Ala	Ala	Ala	Lys	Ala	Ile	Gln	Gly	Leu	Arg	
				85					90					95		
Ala	Val	Phe	Asp	Glu	Thr	Tyr	Pro	Asp	Pro	Val	Arg	Val	Val	Ser	Ile	
			100					105						110		
Gly	Val	Pro	Val	Ser	Glu	Leu	Leu	Asp	Asp	Pro	Ser	Gly	Pro	Ala	Gly	
		115					120					125				
Ser	Leu	Thr	Ser	Val	Glu	Phe	Cys	Gly	Gly	Thr	His	Leu	Arg	Asn	Ser	
	130					135					140					
Ser	His	Ala	Gly	Ala	Phe	Val	Ile	Val	Thr	Glu	Glu	Ala	Ile	Ala	Lys	
145					150					155					160	
Gly	Ile	Arg	Arg	Ile	Val	Ala	Val	Thr	Gly	Ala	Glu	Ala	Gln	Lys	Ala	
				165					170					175		
Leu	Arg	Lys	Ala	Glu	Ser	Leu	Lys	Lys	Cys	Leu	Ser	Val	Met	Glu	Ala	
			180					185					190			
Lys	Val	Lys	Ala	Gln	Thr	Ala	Pro	Asn	Lys	Asp	Val	Gln	Arg	Glu	Ile	
		195					200					205				

120161\_426PC\_Sequence\_Listing.TXT

Ala	Asp	Leu	Gly	Glu	Ala	Leu	Ala	Thr	Ala	Val	Ile	Pro	Gln	Trp	Gln
210						215					220				
Lys	Asp	Glu	Leu	Arg	Glu	Thr	Leu	Lys	Ser	Leu	Lys	Lys	Val	Met	Asp
225					230					235					240
Asp	Leu	Asp	Arg	Ala	Ser	Lys	Ala	Asp	Val	Gln	Lys	Arg	Val	Leu	Glu
				245					250					255	
Lys	Thr	Lys	Gln	Phe	Ile	Asp	Ser	Asn	Pro	Asn	Gln	Pro	Leu	Val	Ile
			260					265					270		
Leu	Glu	Met	Glu	Ser	Gly	Ala	Ser	Ala	Lys	Ala	Leu	Asn	Glu	Ala	Leu
		275					280					285			
Lys	Leu	Phe	Lys	Met	His	Ser	Pro	Gln	Thr	Ser	Ala	Met	Leu	Phe	Thr
	290					295					300				
Val	Asp	Asn	Glu	Ala	Gly	Lys	Ile	Thr	Cys	Leu	Cys	Gln	Val	Pro	Gln
305					310					315					320
Asn	Ala	Ala	Asn	Arg	Gly	Leu	Lys	Ala	Ser	Glu	Trp	Val	Gln	Gln	Val
				325					330					335	
Ser	Gly	Leu	Met	Asp	Gly	Lys	Gly	Gly	Gly	Lys	Asp	Val	Ser	Ala	Gln
			340					345					350		
Ala	Thr	Gly	Lys	Asn	Val	Gly	Cys	Leu	Gln	Glu	Ala	Leu	Gln	Leu	Ala
		355					360					365			
Thr	Ser	Phe	Ala	Gln	Leu	Arg	Leu	Gly	Asp	Val	Lys	Asn			
	370					375					380				

<210> 84  
 <211> 1146  
 <212> DNA  
 <213> Homo sapiens

<400> 84  
 caggctctggc tgtttattga tgagccccga cgaagaccca tcatgagcaa ccacacagct 60  
 acgcacattc tgaacttcgc cctgcgctca gtgcctgggg aagctgacca gaaaggctca 120  
 ttggttgctc ctgaccgcct cagatttgac ttactgcca agggagccat gtccacccaa 180  
 cagatcaaga aggtgaaga gattgctaata gagatgattg aggcagccaa ggccgtctat 240  
 acccaggatt gccccctggc agcagcgaaa gccatccagg gcctacgggc tgtgtttgat 300  
 gagacctatc ctgaccctgt gcgagtcgtc tccattgggg tcccgggtgtc cgagttgctg 360  
 gatgaccctt ctgggcctgc tggctccctg acttctgttg agttctgttg gggaacgcac 420  
 ctgcggaact cgagtcatgc aggagctttt gtgatcgtga cggaagaagc cattgccaag 480  
 ggtatccgga ggattgtggc tgtcacaggt gccgaggccc agaaggccct caggaaagca 540  
 gagagcttga agaaatgtct ctctgtcatg gaagccaaag tgaaggctca gactgtctca 600  
 aacaaggatg tgcagagggg gatcgctgac cttggagagg ccctggccac tgcagtcatc 660  
 cccagtgagg agaaggatga attgcgggag actctcaaat ccctaaagaa ggtcatggat 720  
 gacttgacc gagccagcaa agccgatgtc cagaaacgag tgtagagaa gacgaagcag 780  
 ttcacgcaca gcaaccccaa ccagcctctt gtcacctgg agatggagag cggcgctca 840  
 gccaaaggccc tgaatgaagc cttgaagctc ttcaagatgc actcccctca gacttctgcc 900  
 atgctcttca cgggtggaaa tgaggctggc aagatcacgt gcctgtgtca agttccccag 960  
 aatgcagcca atcggggctt aaaagccagc gagggtgtgc agcaggtgtc aggtttgatg 1020  
 gacggtaaag gtggtggcaa ggatgtgtct gcacaggcca caggcaagaa cgttggctgc 1080  
 ctgcaggagg cgctgcagct ggccacttcc ttcgcccagc tgcgcctcgg ggatgtaaag 1140  
 aactga 1146

<210> 85  
 <211> 19  
 <212> PRT  
 <213> Mus musculus

<400> 85  
 Thr Cys Phe Tyr Ala Glu Gln Gly Gly Gln Ile Tyr Asp Glu Gly Tyr  
 1 5 10 15  
 Leu Val Lys

<210> 86  
 <211> 243  
 <212> PRT  
 <213> Mus musculus

120161\_426PC\_Sequence\_Listing.TXT

<400> 86

Val Asp Asp Ser Ser Glu Asp Lys Thr Glu Phe Thr Val Lys Asn Ala  
 1 5 10 15  
 Gln Val Arg Gly Gly Tyr Val Leu His Ile Gly Thr Ile Tyr Gly Asn  
 20 25 30  
 Leu Lys Val Gly Asp Gln Val Arg Leu Phe Ile Asp Glu Pro Arg Arg  
 35 40 45  
 Arg Pro Val Met Ser Asn His Thr Ala Thr His Ile Leu Asn Phe Ala  
 50 55 60  
 Leu Arg Ser Val Leu Gly Glu Ala Asp Gln Lys Gly Ser Leu Val Ala  
 65 70 75 80  
 Pro Asp Arg Leu Arg Phe Asp Phe Thr Ala Lys Gly Ala Met Ser Thr  
 85 90 95  
 Gln Gln Ile Lys Lys Ala Glu Glu Ile Val Asn Gly Met Ile Glu Ala  
 100 105 110  
 Ala Lys Pro Val Tyr Thr Gln Asp Cys Pro Leu Ala Ala Ala Lys Ala  
 115 120 125  
 Ile Gln Gly Leu Arg Ala Val Phe Asp Glu Thr Tyr Pro Asp Pro Val  
 130 135 140  
 Arg Val Val Ser Ile Gly Val Pro Val Ser Glu Leu Leu Asp Asp Pro  
 145 150 155 160  
 Cys Gly Pro Ala Gly Ser Leu Thr Ser Val Glu Phe Cys Gly Gly Thr  
 165 170 175  
 His Leu Arg Asn Ser Ser His Ala Gly Ala Phe Val Ile Val Thr Glu  
 180 185 190  
 Glu Ala Ile Ala Lys Gly Ile Arg Arg Ile Val Ala Val Thr Gly Ala  
 195 200 205  
 Glu Ala Gln Lys Ala Leu Arg Lys Ser Glu Thr Leu Lys Lys Ser Leu  
 210 215 220  
 Ser Ala Met Glu Ala Lys Val Lys Ala Gln Thr Ala Pro Asn Lys Asp  
 225 230 235 240  
 Val Gln Arg

<210> 87

<211> 19

<212> PRT

<213> Mus musculus

<400> 87

Glu Ile Ala Asp Leu Gly Glu Ala Leu Ala Thr Ala Val Ile Pro Gln  
 1 5 10 15  
 Trp Gln Lys

<210> 88

<211> 18

<212> PRT

<213> Mus musculus

<400> 88

Met Phe Val Asp Glu Val Val Thr Gly Gln Glu Cys Gly Val Val Leu  
 1 5 10 15  
 Asp Lys

<210> 89

<211> 19

<212> PRT

<213> Mus musculus

<400> 89

Thr Cys Phe Tyr Ala Glu Gln Gly Gly Gln Ile Tyr Asp Glu Gly Tyr  
 1 5 10 15  
 Leu Val Lys



<210> 90  
 <211> 14  
 <212> PRT  
 <213> Mus musculus

<400> 90  
 Val Asp Asp Ser Ser Glu Asp Lys Thr Glu Phe Thr Val Lys  
 1 5 10

<210> 91  
 <211> 24  
 <212> PRT  
 <213> Mus musculus

<400> 91  
 Tyr Asn Tyr Gln Ser Asp Ser Ser Gly Ser Tyr Val Phe Glu Cys Thr  
 1 5 10 15  
 Val Ala Thr Val Leu Ala Leu Arg  
 20

<210> 92  
 <211> 21  
 <212> PRT  
 <213> Mus musculus

<400> 92  
 Arg Glu Lys Met Phe Val Asp Glu Val Val Thr Gly Gln Glu Cys Gly  
 1 5 10 15  
 Val Val Leu Asp Lys  
 20

<210> 93  
 <211> 19  
 <212> PRT  
 <213> Mus musculus

<400> 93  
 Thr Cys Phe Tyr Ala Glu Gln Gly Gly Gln Ile Tyr Asp Glu Gly Tyr  
 1 5 10 15  
 Leu Val Lys

<210> 94  
 <211> 365  
 <212> PRT  
 <213> Mus musculus

<400> 94  
 Val Asp Asp Ser Ser Glu Asp Lys Thr Glu Phe Thr Val Lys Asn Ala  
 1 5 10 15  
 Gln Val Arg Gly Gly Tyr Val Leu His Ile Gly Thr Ile Tyr Gly Asn  
 20 25 30  
 Leu Lys Val Gly Asp Gln Val Arg Leu Phe Ile Asp Glu Pro Arg Arg  
 35 40 45  
 Arg Pro Val Met Ser Asn His Thr Ala Thr His Ile Leu Asn Phe Ala  
 50 55 60  
 Leu Arg Ser Val Leu Gly Glu Ala Asp Gln Lys Gly Ser Leu Val Ala  
 65 70 75 80  
 Pro Asp Arg Leu Arg Phe Asp Phe Thr Ala Lys Gly Ala Met Ser Thr  
 85 90 95  
 Gln Gln Ile Lys Lys Ala Glu Glu Ile Val Asn Gly Met Ile Glu Ala

120161\_426PC\_Sequence\_Listing.TXT

Ala	Lys	Pro	100	Tyr	Thr	Gln	Asp	105	Cys	Pro	Leu	Ala	Ala	110	Lys	Ala
Ile	Gln	Gly	115	Leu	Arg	Ala	Val	120	Phe	Asp	Glu	Thr	Tyr	125	Pro	Asp
Arg	Val	Val	130	Ser	Ile	Gly	Val	135	Pro	Val	Ser	Glu	Leu	140	Leu	Asp
145	Cys	Gly	Pro	Ala	Gly	150	Ser	Leu	Thr	Ser	Val	Glu	Phe	155	Cys	Gly
His	Leu	Arg	Asn	Ser	Ser	His	Ala	Gly	170	Ala	Phe	Val	Ile	175	Val	Thr
Glu	Ala	Ile	180	Ala	Lys	Gly	Ile	Arg	185	Arg	Ile	Val	Ala	190	Val	Thr
Glu	Ala	195	Gln	Lys	Ala	Leu	Arg	200	Lys	Ser	Glu	Thr	Leu	205	Lys	Lys
Ser	Ala	Met	Glu	Ala	Lys	Val	Lys	Ala	Gln	Thr	Ala	Pro	Asn	Lys	Asp	240
225	Val	Gln	Arg	Glu	Ile	Ala	Asp	Leu	Gly	Glu	Ala	Leu	Ala	Thr	Ala	Val
Ile	Pro	Gln	Trp	Gln	Lys	Asp	Glu	Gln	Arg	Glu	Thr	Leu	Lys	255	Ser	Leu
Lys	Lys	Val	Met	Asp	Asp	Leu	Asp	Arg	Ala	Ser	Lys	Ala	Asp	270	Val	Gln
Lys	Arg	Val	Leu	Glu	Lys	Thr	Lys	Gln	Leu	Ile	Asp	Ser	Asn	Pro	Asn	300
Gln	Pro	Leu	Val	Ile	Leu	Glu	Met	Glu	Ser	Gly	Ala	Ser	Ala	Lys	Ala	320
305	Leu	Asn	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Thr	His	Ser	Pro	Gln	Thr	Ser
Ala	Met	Leu	Phe	Thr	Val	Asp	Asn	Glu	Ala	Gly	Lys	Ile	Thr	Cys	Leu	335
Cys	Gln	Val	Pro	Gln	Asn	Ala	Ala	Asn	Arg	Gly	Leu	Lys	350			
		355						360					365			

<210> 95  
 <211> 15  
 <212> PRT  
 <213> Mus musculus

<400> 95  
 Ala Ser Glu Trp Val Gln Gln Val Ser Gly Leu Met Asp Gly Lys  
 1 5 10 15

<210> 96  
 <211> 14  
 <212> PRT  
 <213> Mus musculus

<400> 96  
 Val Asp Asp Ser Ser Glu Asp Lys Thr Glu Phe Thr Val Lys  
 1 5 10

<210> 97  
 <211> 229  
 <212> PRT  
 <213> Mus musculus

<400> 97  
 Asn Ala Gln Val Arg Gly Gly Tyr Val Leu His Ile Gly Thr Ile Tyr  
 1 5 10 15  
 Gly Asn Leu Lys Val Gly Asp Gln Val Arg Leu Phe Ile Asp Glu Pro  
 20 25 30  
 Arg Arg Arg Pro Val Met Ser Asn His Thr Ala Thr His Ile Leu Asn  
 35 40 45  
 Phe Ala Leu Arg Ser Val Leu Gly Glu Ala Asp Gln Lys Gly Ser Leu

120161\_426PC\_Sequence\_Listing.TXT

```

50
Val Ala Pro Asp Arg Leu Arg Phe Asp Phe Thr Ala Lys Gly Ala Met
65 70 75 80
Ser Thr Gln Gln Ile Lys Lys Ala Glu Glu Ile Val Asn Gly Met Ile
85 90 95
Glu Ala Ala Lys Pro Val Tyr Thr Gln Asp Cys Pro Leu Ala Ala Ala
100 105 110
Lys Ala Ile Gln Gly Leu Arg Ala Val Phe Asp Glu Thr Tyr Pro Asp
115 120 125
Pro Val Arg Val Val Ser Ile Gly Val Pro Val Ser Glu Leu Leu Asp
130 135 140
Asp Pro Cys Gly Pro Ala Gly Ser Leu Thr Ser Val Glu Phe Cys Gly
145 150 155 160
Gly Thr His Leu Arg Asn Ser Ser His Ala Gly Ala Phe Val Ile Val
165 170 175
Thr Glu Glu Ala Ile Ala Lys Gly Ile Arg Arg Ile Val Ala Val Thr
180 185 190
Gly Ala Glu Ala Gln Lys Ala Leu Arg Lys Ser Glu Thr Leu Lys Lys
195 200 205
Ser Leu Ser Ala Met Glu Ala Lys Val Lys Ala Gln Thr Ala Pro Asn
210 215 220
Lys Asp Val Gln Arg
225

```

<210> 98  
 <211> 19  
 <212> PRT  
 <213> Mus musculus

<400> 98  
 Glu Ile Ala Asp Leu Gly Glu Ala Leu Ala Thr Ala Val Ile Pro Gln  
 1 5 10 15  
 Trp Gln Lys

<210> 99  
 <211> 67  
 <212> PRT  
 <213> Mus musculus

<400> 99  
 Asp Glu Gln Arg Glu Thr Leu Lys Ser Leu Lys Lys Val Met Asp Asp  
 1 5 10 15  
 Leu Asp Arg Ala Ser Lys Ala Asp Val Gln Lys Arg Val Leu Glu Lys  
 20 25 30  
 Thr Lys Gln Leu Ile Asp Ser Asn Pro Asn Gln Pro Leu Val Ile Leu  
 35 40 45  
 Glu Met Glu Ser Gly Ala Ser Ala Lys Ala Leu Asn Glu Ala Leu Lys  
 50 55 60  
 Leu Phe Lys  
 65

<210> 100  
 <211> 19  
 <212> PRT  
 <213> Mus musculus

<400> 100  
 Thr His Ser Pro Gln Thr Ser Ala Met Leu Phe Thr Val Asp Asn Glu  
 1 5 10 15  
 Ala Gly Lys

<210> 101

120161\_426PC\_Sequence\_Listing.TXT

<211> 14

<212> PRT

<213> Mus musculus

<400> 101

Ile Thr Cys Leu Cys Gln Val Pro Gln Asn Ala Ala Asn Arg  
1 5 10

<210> 102

<211> 3

<212> PRT

<213> Mus musculus

<400> 102

Gly Leu Lys  
1

<210> 103

<211> 15

<212> PRT

<213> Mus musculus

<400> 103

Ala Ser Glu Trp Val Gln Gln Val Ser Gly Leu Met Asp Gly Lys  
1 5 10 15

<210> 104

<211> 281

<212> PRT

<213> Mus musculus

<400> 104

Thr Cys Phe Tyr Ala Glu Gln Gly Gly Gln Ile Tyr Asp Glu Gly Tyr  
1 5 10 15  
Leu Val Lys Val Asp Asp Ser Ser Glu Asp Lys Thr Glu Phe Thr Val  
20 25 30  
Lys Asn Ala Gln Val Arg Gly Gly Tyr Val Leu His Ile Gly Thr Ile  
35 40 45  
Tyr Gly Asn Leu Lys Val Gly Asp Gln Val Arg Leu Phe Ile Asp Glu  
50 55 60  
Pro Arg Arg Arg Pro Val Met Ser Asn His Thr Ala Thr His Ile Leu  
65 70 75 80  
Asn Phe Ala Leu Arg Ser Val Leu Gly Glu Ala Asp Gln Lys Gly Ser  
85 90 95  
Leu Val Ala Pro Asp Arg Leu Arg Phe Thr Ala Lys Gly Ala  
100 105 110  
Met Ser Thr Gln Gln Ile Lys Lys Ala Glu Glu Ile Val Asn Gly Met  
115 120 125  
Ile Glu Ala Ala Lys Pro Val Tyr Thr Gln Asp Cys Pro Leu Ala Ala  
130 135 140  
Ala Lys Ala Ile Gln Gly Leu Arg Ala Val Phe Asp Glu Thr Tyr Pro  
145 150 155 160  
Asp Pro Val Arg Val Val Ser Ile Gly Val Pro Val Ser Glu Leu Leu  
165 170 175  
Asp Asp Pro Cys Gly Pro Ala Gly Ser Leu Thr Ser Val Glu Phe Cys  
180 185 190  
Gly Gly Thr His Leu Arg Asn Ser Ser His Ala Gly Ala Phe Val Ile  
195 200 205  
Val Thr Glu Glu Ala Ile Ala Lys Gly Ile Arg Arg Ile Val Ala Val  
210 215 220  
Thr Gly Ala Glu Ala Gln Lys Ala Leu Arg Lys Ser Glu Thr Leu Lys  
225 230 235 240  
Lys Ser Leu Ser Ala Met Glu Ala Lys Val Lys Ala Gln Thr Ala Pro  
245 250 255  
Asn Lys Asp Val Gln Arg Glu Ile Ala Asp Leu Gly Glu Ala Leu Ala

120161\_426PC\_Sequence\_Listing.TXT

Thr Ala Val Ile Pro Gln Trp Gln Lys  
260 265 270  
275 280

<210> 105  
<211> 51  
<212> PRT  
<213> Mus musculus

<400> 105  
Met Phe Val Asp Glu Val Val Thr Gly Gln Glu Cys Gly Val Val Leu  
1 5 10 15  
Asp Lys Thr Cys Phe Tyr Ala Glu Gln Gly Gln Ile Tyr Asp Glu  
20 25 30  
Gly Tyr Leu Val Lys Val Asp Asp Ser Ser Glu Asp Lys Thr Glu Phe  
35 40 45  
Thr Val Lys  
50

<210> 106  
<211> 444  
<212> PRT  
<213> Mus musculus

<400> 106  
Tyr Asn Tyr Gln Ser Asp Ser Ser Gly Ser Tyr Val Phe Glu Cys Thr  
1 5 10 15  
Val Ala Thr Val Leu Ala Leu Arg Arg Glu Lys Met Phe Val Asp Glu  
20 25 30  
Val Val Thr Gly Gln Glu Cys Gly Val Val Leu Asp Lys Thr Cys Phe  
35 40 45  
Tyr Ala Glu Gln Gly Gly Gln Ile Tyr Asp Glu Gly Tyr Leu Val Lys  
50 55 60  
Val Asp Asp Ser Ser Glu Asp Lys Thr Glu Phe Thr Val Lys Asn Ala  
65 70 75 80  
Gln Val Arg Gly Gly Tyr Val Leu His Ile Gly Thr Ile Tyr Gly Asn  
85 90 95  
Leu Lys Val Gly Asp Gln Val Arg Leu Phe Ile Asp Glu Pro Arg Arg  
100 105 110  
Arg Pro Val Met Ser Asn His Thr Ala Thr His Ile Leu Asn Phe Ala  
115 120 125  
Leu Arg Ser Val Leu Gly Glu Ala Asp Gln Lys Gly Ser Leu Val Ala  
130 135 140  
Pro Asp Arg Leu Arg Phe Asp Phe Thr Ala Lys Gly Ala Met Ser Thr  
145 150 155 160  
Gln Gln Ile Lys Lys Ala Glu Glu Ile Val Asn Gly Met Ile Glu Ala  
165 170 175  
Ala Lys Pro Val Tyr Thr Gln Asp Cys Pro Leu Ala Ala Ala Lys Ala  
180 185 190  
Ile Gln Gly Leu Arg Ala Val Phe Asp Glu Thr Tyr Pro Asp Pro Val  
195 200 205  
Arg Val Val Ser Ile Gly Val Pro Val Ser Glu Leu Leu Asp Asp Pro  
210 215 220  
Cys Gly Pro Ala Gly Ser Leu Thr Ser Val Glu Phe Cys Gly Gly Thr  
225 230 235 240  
His Leu Arg Asn Ser Ser His Ala Gly Ala Phe Val Ile Val Thr Glu  
245 250 255  
Glu Ala Ile Ala Lys Gly Ile Arg Arg Ile Val Ala Val Thr Gly Ala  
260 265 270  
Glu Ala Gln Lys Ala Leu Arg Lys Ser Glu Thr Leu Lys Lys Ser Leu  
275 280 285  
Ser Ala Met Glu Ala Lys Val Lys Ala Gln Thr Ala Pro Asn Lys Asp  
290 295 300  
Val Gln Arg Glu Ile Ala Asp Leu Gly Glu Ala Leu Ala Thr Ala Val  
305 310 315 320  
Ile Pro Gln Trp Gln Lys Asp Glu Gln Arg Glu Thr Leu Lys Ser Leu

120161\_426PC\_Sequence\_Listing.TXT

Lys	Lys	Val	Met	325	Asp	Asp	Leu	Asp	Arg	330	Ala	Ser	Lys	Ala	Asp	335	Val	Gln
			340						345						350			
Lys	Arg	Val	Leu	Glu	Lys	Thr	Lys	Gln	Leu	Ile	Asp	Ser	Asn	Pro	Asn			
		355					360						365					
Gln	Pro	Leu	Val	Ile	Leu	Glu	Met	Glu	Ser	Gly	Ala	Ser	Ala	Lys	Ala			
		370				375					380							
Leu	Asn	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Thr	His	Ser	Pro	Gln	Thr	Ser			
385					390					395								
Ala	Met	Leu	Phe	Thr	Val	Asp	Asn	Glu	Ala	Gly	Lys	Ile	Thr	Cys	Leu			
				405					410					415				
Cys	Gln	Val	Pro	Gln	Asn	Ala	Ala	Asn	Arg	Gly	Leu	Lys	Ala	Ser	Glu			
			420					425					430					
Trp	Val	Gln	Gln	Val	Ser	Gly	Leu	Met	Asp	Gly	Lys							
		435					440											

<210> 107  
 <211> 380  
 <212> PRT  
 <213> Mus musculus

<400> 107

Val	Asp	Asp	Ser	Ser	Glu	Asp	Lys	Thr	Glu	Phe	Thr	Val	Lys	Asn	Ala
1				5					10					15	
Gln	Val	Arg	Gly	Gly	Tyr	Val	Leu	His	Ile	Gly	Thr	Ile	Tyr	Gly	Asn
			20					25					30		
Leu	Lys	Val	Gly	Asp	Gln	Val	Arg	Leu	Phe	Ile	Asp	Glu	Pro	Arg	Arg
		35					40					45			
Arg	Pro	Val	Met	Ser	Asn	His	Thr	Ala	Thr	His	Ile	Leu	Asn	Phe	Ala
	50				55					60					
Leu	Arg	Ser	Val	Leu	Gly	Glu	Ala	Asp	Gln	Lys	Gly	Ser	Leu	Val	Ala
65				70					75						80
Pro	Asp	Arg	Leu	Arg	Phe	Asp	Phe	Thr	Ala	Lys	Gly	Ala	Met	Ser	Thr
			85					90					95		
Gln	Gln	Ile	Lys	Lys	Ala	Glu	Glu	Ile	Val	Asn	Gly	Met	Ile	Glu	Ala
			100					105					110		
Ala	Lys	Pro	Val	Tyr	Thr	Gln	Asp	Cys	Pro	Leu	Ala	Ala	Lys	Ala	
		115					120					125			
Ile	Gln	Gly	Leu	Arg	Ala	Val	Phe	Asp	Glu	Thr	Tyr	Pro	Asp	Pro	Val
	130				135						140				
Arg	Val	Val	Ser	Ile	Gly	Val	Pro	Val	Ser	Glu	Leu	Leu	Asp	Asp	Pro
145				150					155						160
Cys	Gly	Pro	Ala	Gly	Ser	Leu	Thr	Ser	Val	Glu	Phe	Cys	Gly	Gly	Thr
			165					170					175		
His	Leu	Arg	Asn	Ser	Ser	His	Ala	Gly	Ala	Phe	Val	Ile	Val	Thr	Glu
			180					185					190		
Glu	Ala	Ile	Ala	Lys	Gly	Ile	Arg	Arg	Ile	Val	Ala	Val	Thr	Gly	Ala
		195					200					205			
Glu	Ala	Gln	Lys	Ala	Leu	Arg	Lys	Ser	Glu	Thr	Leu	Lys	Lys	Ser	Leu
	210					215					220				
Ser	Ala	Met	Glu	Ala	Lys	Val	Lys	Ala	Gln	Thr	Ala	Pro	Asn	Lys	Asp
225				230				235							240
Val	Gln	Arg	Glu	Ile	Ala	Asp	Leu	Gly	Glu	Ala	Leu	Ala	Thr	Ala	Val
			245					250					255		
Ile	Pro	Gln	Trp	Gln	Lys	Asp	Glu	Gln	Arg	Glu	Thr	Leu	Lys	Ser	Leu
		260					265					270			
Lys	Lys	Val	Met	Asp	Asp	Leu	Asp	Arg	Ala	Ser	Lys	Ala	Asp	Val	Gln
		275					280					285			
Lys	Arg	Val	Leu	Glu	Lys	Thr	Lys	Gln	Leu	Ile	Asp	Ser	Asn	Pro	Asn
	290					295					300				
Gln	Pro	Leu	Val	Ile	Leu	Glu	Met	Glu	Ser	Gly	Ala	Ser	Ala	Lys	Ala
305					310					315					320
Leu	Asn	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Thr	His	Ser	Pro	Gln	Thr	Ser
			325					330						335	
Ala	Met	Leu	Phe	Thr	Val	Asp	Asn	Glu	Ala	Gly	Lys	Ile	Thr	Cys	Leu
			340					345					350		
Cys	Gln	Val	Pro	Gln	Asn	Ala	Ala	Asn	Arg	Gly	Leu	Lys	Ala	Ser	Glu

120161\_426PC\_Sequence\_Listing.TXT

Trp Val Gln Gln Val Ser Gly Leu Met Asp Gly Lys  
 355 360 365  
 370 375 380

<210> 108  
 <211> 28  
 <212> PRT  
 <213> Mus musculus

<400> 108  
 Asp Met Ser Ala Gln Ala Thr Gly Lys Asn Val Gly Cys Leu Gln Glu  
 1 5 10 15  
 Ala Leu Gln Leu Ala Thr Ser Phe Ala Gln Leu Arg  
 20 25

<210> 109  
 <211> 9  
 <212> PRT  
 <213> Mus musculus

<400> 109  
 Asp Met Ser Ala Gln Ala Thr Gly Lys  
 1 5

<210> 110  
 <211> 19  
 <212> PRT  
 <213> Mus musculus

<400> 110  
 Asn Val Gly Cys Leu Gln Glu Ala Leu Gln Leu Ala Thr Ser Phe Ala  
 1 5 10 15  
 Gln Leu Arg

<210> 111  
 <211> 51  
 <212> PRT  
 <213> Mus musculus

<400> 111  
 Met Phe Val Asp Glu Val Val Thr Gly Gln Glu Cys Gly Val Val Leu  
 1 5 10 15  
 Asp Lys Thr Cys Phe Tyr Ala Glu Gln Gly Gly Gln Ile Tyr Asp Glu  
 20 25 30  
 Gly Tyr Leu Val Lys Val Asp Asp Ser Ser Glu Asp Lys Thr Glu Phe  
 35 40 45  
 Thr Val Lys  
 50

<210> 112  
 <211> 19  
 <212> PRT  
 <213> Mus musculus

<400> 112  
 Glu Ile Ala Asp Leu Gly Glu Ala Leu Ala Thr Ala Val Ile Pro Gln  
 1 5 10 15  
 Trp Gln Lys

<210> 113

120161\_426PC\_Sequence\_Listing.TXT

<211> 86  
<212> PRT  
<213> Mus musculus

<400> 113  
Asp Glu Gln Arg Glu Thr Leu Lys Ser Leu Lys Lys Val Met Asp Asp  
1 5 10 15  
Leu Asp Arg Ala Ser Lys Ala Asp Val Gln Lys Arg Val Leu Glu Lys  
20 25 30  
Thr Lys Gln Leu Ile Asp Ser Asn Pro Asn Gln Pro Leu Val Ile Leu  
35 40 45  
Glu Met Glu Ser Gly Ala Ser Ala Lys Ala Leu Asn Glu Ala Leu Lys  
50 55 60  
Leu Phe Lys Thr His Ser Pro Gln Thr Ser Ala Met Leu Phe Thr Val  
65 70 75 80  
Asp Asn Glu Ala Gly Lys  
85

<210> 114  
<211> 14  
<212> PRT  
<213> Mus musculus

<400> 114  
Ile Thr Cys Leu Cys Gln Val Pro Gln Asn Ala Ala Asn Arg  
1 5 10

<210> 115  
<211> 119  
<212> PRT  
<213> Mus musculus

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

<210> 116  
<211> 676  
<212> PRT  
<213> Homo sapiens

<400> 116  
Met Ala Tyr Arg Val Leu Ala Asp His Ala Arg Thr Ile Thr Val Ala  
1 5 10 15  
Leu Ala Asp Gly Gly Arg Pro Asp Asn Thr Gly Arg Gly Tyr Val Leu  
20 25 30  
Arg Arg Ile Leu Arg Arg Ala Val Arg Tyr Ala His Glu Lys Leu Asn  
35 40 45  
Ala Ser Arg Gly Phe Phe Ala Thr Leu Val Asp Val Val Val Gln Ser  
50 55 60  
Leu Gly Asp Ala Phe Pro Glu Leu Lys Lys Asp Pro Asp Met Val Lys



120161\_426PC\_Sequence\_Listing.TXT

65	Asp	Ile	Ile	Asn	Glu	70	Glu	Glu	Val	Gln	75	Leu	Lys	Thr	Leu	80	Ser	Arg
				85	Leu	Asp	Arg	Lys	Ile	Gln	90	Ser	Leu	Gly	Asp	95	Ser	Lys
			100	Gly	Asp	Thr	Ala	Trp	105	Leu	Leu	Tyr	Asp	Thr	110	Gly	Phe	
		115	Pro	Gly	Asp	Thr	Ala	Trp	120	Leu	Leu	Tyr	Asp	Thr	125	Tyr	Gly	Phe
	Pro	Val	Asp	Leu	Thr	Gly	Leu	Ile	Ala	Glu	Glu	Lys	Gly	Leu	Val	Val		
	130	Met	Asp	Gly	Phe	Glu	Glu	Glu	Arg	Lys	Leu	Ala	Gln	Leu	Lys	Ser		
145	Gln	Gly	Lys	Gly	Ala	150	Gly	Gly	Glu	Asp	Leu	Ile	Met	Leu	Asp	Ile	Tyr	
				165	Ala	Gly	Gly	Glu	Asp	Leu	Ile	Met	Leu	Asp	Ile	Tyr		
	Ala	Ile	Glu	Glu	180	Leu	Arg	Ala	Arg	Gly	185	Leu	Glu	Val	Thr	Asp	Ser	
	Pro	Lys	Tyr	Asn	Tyr	His	Leu	Asp	Ser	Ser	Gly	Ser	Tyr	Val	Phe	Glu		
		195	Val	Ala	Thr	Val	Met	Ala	Leu	Arg	Arg	Glu	Lys	Met	Phe	Val		
	Asn	Thr	Val	Ala	Thr	Val	Met	Ala	Leu	Arg	Arg	Glu	Lys	Met	Phe	Val		
	210	Glu	Val	Ser	Thr	Gly	230	Gln	Glu	Cys	Gly	Val	Val	Leu	Asp	Lys	Thr	
225	Glu	Val	Ser	Thr	Gly	230	Gln	Glu	Cys	Gly	Val	Val	Leu	Asp	Lys	Thr		
	Cys	Phe	Tyr	Ala	Glu	Gln	Gly	Gly	Gln	Ile	Tyr	Asp	Glu	Gly	Tyr	Leu		
				245	Glu	Gln	Gly	Gly	Gln	Ile	Tyr	Asp	Glu	Gly	Tyr	Leu		
	Val	Lys	Val	Asp	Asp	Ser	Ser	Glu	Asp	Lys	Thr	Glu	Phe	Thr	Val	Lys		
			260	Asp	Asp	Ser	Ser	Glu	Asp	Lys	Thr	Glu	Phe	Thr	Val	Lys		
	Asn	Ala	Gln	Val	Arg	Gly	Gly	Tyr	280	Val	Leu	His	Ile	Gly	Thr	Ile	Tyr	
		275	Val	Arg	Gly	Gly	Tyr	280	Val	Leu	His	Ile	Gly	Thr	Ile	Tyr		
	Gly	Asp	Leu	Lys	Val	Gly	Asp	Gln	Val	Trp	Leu	Phe	Ile	Asp	Glu	Pro		
	290	Asp	Leu	Lys	Val	Gly	Asp	Gln	Val	Trp	Leu	Phe	Ile	Asp	Glu	Pro		
	Arg	Arg	Arg	Pro	Ile	Met	Ser	Asn	His	Thr	Ala	Thr	His	Ile	Leu	Asn		
305	Arg	Arg	Arg	Pro	Ile	Met	Ser	Asn	His	Thr	Ala	Thr	His	Ile	Leu	Asn		
	Phe	Ala	Leu	Arg	Ser	310	Val	Leu	Gly	Glu	Ala	Asp	Gln	Lys	Gly	Ser	Leu	
				325	Val	Leu	Gly	Glu	Ala	Asp	Gln	Lys	Gly	Ser	Leu			
	Val	Ala	Pro	Asp	Arg	Leu	Arg	Phe	Asp	Phe	Thr	Ala	Lys	Gly	Ala	Met		
			340	Arg	Leu	Arg	Phe	Asp	Phe	Thr	Ala	Lys	Gly	Ala	Met			
	Ser	Thr	Gln	Gln	Ile	Lys	Lys	Ala	Glu	Glu	Ile	Ala	Asn	Glu	Met	Ile		
			355	Gln	Gln	Ile	Lys	Lys	Ala	Glu	Glu	Ile	Ala	Asn	Glu	Met	Ile	
	Glu	Ala	Ala	Lys	Ala	Val	Tyr	375	Thr	Gln	Asp	Cys	Pro	Leu	Ala	Ala	Ala	
	370	Ala	Ala	Lys	Ala	Val	Tyr	375	Thr	Gln	Asp	Cys	Pro	Leu	Ala	Ala	Ala	
	Lys	Ala	Ile	Gln	Gly	Leu	Arg	Ala	Val	Phe	Asp	Glu	Thr	Tyr	Pro	Asp		
385	Lys	Ala	Ile	Gln	Gly	Leu	Arg	Ala	Val	Phe	Asp	Glu	Thr	Tyr	Pro	Asp		
	Pro	Val	Arg	Val	Val	Ser	Ile	Gly	Val	Pro	Val	Ser	Glu	Leu	Leu	Asp		
				405	Ser	Ile	Gly	Val	Pro	Val	Ser	Glu	Leu	Leu	Asp			
	Asp	Pro	Ser	Gly	Pro	Ala	Gly	Ser	Leu	Thr	Ser	Val	Glu	Phe	Cys	Gly		
			420	Pro	Ala	Gly	Ser	Leu	Thr	Ser	Val	Glu	Phe	Cys	Gly			
	Gly	Thr	His	Leu	Arg	Asn	Ser	Ser	His	Ala	Gly	Ala	Phe	Val	Ile	Val		
		435	Leu	Arg	Asn	Ser	Ser	His	Ala	Gly	Ala	Phe	Val	Ile	Val			
	Thr	Glu	Ala	Ile	Ala	Lys	Gly	Ile	Arg	Arg	Ile	Val	Ala	Val	Thr			
		450	Ala	Ile	Ala	Lys	Gly	Ile	Arg	Arg	Ile	Val	Ala	Val	Thr			
	Gly	Ala	Glu	Ala	Gln	Lys	Ala	Leu	Arg	Lys	Ala	Glu	Ser	Leu	Lys	Lys		
465	Gly	Ala	Glu	Ala	Gln	Lys	Ala	Leu	Arg	Lys	Ala	Glu	Ser	Leu	Lys	Lys		
				470	Ala	Lys	Val	Lys	Ala	Gln	Thr	Ala	Pro	Asn				
	Cys	Leu	Ser	Val	Met	Glu	Ala	Lys	Val	Lys	Ala	Gln	Thr	Ala	Pro	Asn		
				485	Glu	Ala	Lys	Val	Lys	Ala	Gln	Thr	Ala	Pro	Asn			
	Lys	Asp	Val	Gln	Arg	Glu	Ile	Ala	Asp	Leu	Gly	Glu	Ala	Leu	Ala	Thr		
			500	Arg	Glu	Ile	Ala	Asp	Leu	Gly	Glu	Ala	Leu	Ala	Thr			
	Ala	Val	Ile	Pro	Gln	Trp	Gln	Lys	Asp	Glu	Leu	Arg	Glu	Thr	Leu	Lys		
			515	Pro	Gln	Trp	Gln	Lys	Asp	Glu	Leu	Arg	Glu	Thr	Leu	Lys		
	Ser	Leu	Lys	Lys	Val	Met	Asp	Asp	Leu	Asp	Arg	Ala	Ser	Lys	Ala	Asp		
		530	Lys	Lys	Val	Met	Asp	Asp	Leu	Asp	Arg	Ala	Ser	Lys	Ala	Asp		
	Val	Gln	Lys	Arg	Val	Leu	Glu	Lys	Thr	Lys	Gln	Phe	Ile	Asp	Ser	Asn		
545	Val	Gln	Lys	Arg	Val	Leu	Glu	Lys	Thr	Lys	Gln	Phe	Ile	Asp	Ser	Asn		
				550	Glu	Lys	Thr	Lys	Gln	Phe	Ile	Asp	Ser	Asn				
	Pro	Asn	Gln	Pro	Leu	Val	Ile	Leu	Glu	Met	Glu	Ser	Gly	Ala	Ser	Ala		
				565	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Met	His	Ser	Pro	Gln		
	Lys	Ala	Leu	Asn	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Met	His	Ser	Pro	Gln		
			580	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Met	His	Ser	Pro	Gln			
	Thr	Ser	Ala	Met	Leu	Phe	Thr	Val	Asp	Asn	Glu	Ala	Gly	Lys	Ile	Thr		
			595	Met	Leu	Phe	Thr	Val	Asp	Asn	Glu	Ala	Gly	Lys	Ile	Thr		
	Cys	Leu	Cys	Gln	Val	Pro	Gln	Asn	Ala	Ala	Asn	Arg	Gly	Leu	Lys	Ala		

120161\_426PC\_Sequence\_Listing.TXT

610		615		620	
Ser Glu Trp Val Gln Gln Val Ser Gly Leu Met Asp Gly Lys Gly Gly					
625		630		635	640
Gly Lys Asp Val Ser Ala Gln Ala Thr Gly Lys Asn Val Gly Cys Leu					
	645		650		655
Gln Glu Ala Leu Gln Leu Ala Thr Ser Phe Ala Gln Leu Arg Leu Gly					
	660		665		670
Asp Val Lys Asn					
675					

<210> 117  
 <211> 2031  
 <212> DNA  
 <213> Homo sapiens

<400> 117

atggcctacc	gggtgctggc	tgaccacgct	cggaccatca	ctgtggcact	ggctgatggg	60
ggccggcctg	acaacacagg	gcgtggatat	gtgttgagac	ggatttctccg	ccgagctgtc	120
cgatacgccc	atgaaaagct	caatgccagc	aggggcttct	ttgctacgtt	agtggatggt	180
gtcgtccagt	ccctgggaga	tgcatttcct	gagctgaaga	aggaccaga	catggatgaag	240
gacatcatta	atgaagaaga	ggtgcagttt	ctcaagactc	tcagcagagg	gcgtcgcatac	300
ctggacagga	aaattcagag	cctgggagac	agcaagacca	ttcccggaga	caactgcttg	360
ctcctctatg	acacctatgg	gtttccagtg	gatctgactg	gactgattgc	tgaagagaag	420
ggcctggtgg	tagacatgga	tggctttgaa	gaggagagga	aactggccca	gctgaaatca	480
cagggcaagg	gagctggtgg	ggaagacctc	attatgctgg	acatttacgc	tatcgaagag	540
ctccgggcac	ggggtctgga	ggtcacagat	gattcccca	agtacaatta	ccatttggac	600
tccagtggta	gctatgtatt	tgagaacaca	gtggctacgg	tgatggctct	gcgcagggag	660
aagatgttcg	tggaagaggt	gtccacaggc	caggagtgtg	gagtgggtgct	ggacaagacc	720
tgtttctatg	ctgagcaagg	aggccagatc	tatgacgaag	gctacctggg	gaaggtggat	780
gacagcagtg	aagataaaac	agagtttaca	gtgaagaatg	ctcaggtccg	aggagggtat	840
gtgctacaca	ttggaacct	ctacggtgac	ctgaaagtgg	gggatcagggt	ctggctgttt	900
attgatgagc	cccgcgaag	acccatcatg	agcaaccaca	cagctacgca	catttctgaac	960
ttcgccctgc	gctcagtgct	tggggaagct	gaccagaaag	gctcattggg	tgctcctgac	1020
cgctcagat	ttgactttac	tgccaaggga	gccatgtcca	cccaacagat	caagaaggct	1080
gaagagattg	ctaagagat	gattgaggca	gccaaggccg	tctataccca	ggattgcccc	1140
ctggcagcag	cgaagccat	ccagggccta	cgggctgtgt	ttgatgagac	ctatcctgac	1200
cctgtgctgag	tggtctccat	tgggttccc	gtgtccgagt	tgctggatga	cccctctggg	1260
cctgtctggct	ccctgacttc	tggtgagttc	tggtggggaa	cgcacctgag	gaactcgagt	1320
catgcaggag	cttttgtgat	cgtgacggaa	gaagccattg	ccaagggtat	ccggaggatt	1380
gtggctgtca	caggtgccga	ggcccagaag	gccctcagga	aagcagagag	cttgaagaaa	1440
tgctctctg	tcatggaagc	caaagtgaag	gctcagactg	ctccaaacaa	ggatgtgcag	1500
agggagatcg	ctgaccttgg	agaggccctg	gccactgcag	tcatacccca	gtggcagaag	1560
gatgaattgc	gggagactct	caaattcccta	aagaagggtca	tggtgactt	ggaccgagcc	1620
agcaaagccg	atgtccagaa	acgagtgtta	gagaagacga	agcagttcat	cgacagcaac	1680
cccaaccagc	ctcttgtcat	cctggagatg	gagagcggcg	cctcagccaa	ggccctgaat	1740
gaagccttga	agctcttcaa	gatgcactcc	cctcagactt	ctgccatgct	cttcacgggtg	1800
gacaatgagg	ctggcaagat	cacgtgcctg	tgtcaagttc	cccagaatgc	agccaatcgg	1860
ggcttaaaag	ccagcgagtg	ggtgcagcag	gtgtcaggct	tgatggacgg	taaagggtgtg	1920
ggcaaggatg	tgtctgcaca	ggccacaggc	aagaacgttg	gctgcctgca	ggaggcgctg	1980
cagctggcca	cttcttctgc	ccagctgcgc	ctcggggatg	taaagaactg	a	2031

<210> 118  
 <211> 567  
 <212> PRT  
 <213> Homo sapiens

<400> 118

Met Asp Ser Thr Leu Thr Ala Ser Glu Ile Arg Gln Arg Phe Ile Asp	
1	5
Phe Phe Lys Arg Asn Glu His Thr Tyr Val His Ser Ser Ala Thr Ile	
	20
Pro Leu Asp Asp Pro Thr Leu Leu Phe Ala Asn Ala Gly Met Asn Gln	
	35
Leu Lys Ser Gln Gly Lys Gly Ala Gly Gly Glu Asp Leu Ile Met Leu	
	50
Asp Ile Tyr Ala Ile Glu Glu Leu Arg Ala Arg Gly Leu Glu Val Thr	
65	70
	75
	80

120161\_426PC\_Sequence\_Listing.TXT

Asp	Asp	Ser	Pro	Lys	Tyr	Asn	Tyr	His	Leu	Asp	Ser	Ser	Gly	Ser	Tyr
Val	Phe	Glu	Asn	Thr	Val	Ala	Thr	Val	Met	Ala	Leu	Arg	Arg	Glu	Lys
Met	Phe	Val	Glu	Glu	Val	Ser	Thr	Gly	Gln	Glu	Cys	Gly	Val	Val	Leu
Asp	Lys	Thr	Cys	Phe	Tyr	Ala	Glu	Gln	Gly	Gly	Gln	Ile	Tyr	Asp	Glu
Gly	Tyr	Leu	Val	Lys	Val	Asp	Asp	Ser	Ser	Glu	Asp	Lys	Thr	Glu	Phe
Thr	Val	Lys	Asn	Ala	Gln	Val	Arg	Gly	Gly	Tyr	Val	Leu	His	Ile	Gly
Thr	Ile	Tyr	Gly	Asp	Leu	Lys	Val	Gly	Asp	Gln	Val	Trp	Leu	Phe	Ile
Asp	Glu	Pro	Arg	Arg	Arg	Pro	Ile	Met	Ser	Asn	His	Thr	Ala	Thr	His
Ile	Leu	Asn	Phe	Ala	Leu	Arg	Ser	Val	Leu	Gly	Glu	Ala	Asp	Gln	Lys
Gly	Ser	Leu	Val	Ala	Pro	Asp	Arg	Leu	Arg	Phe	Asp	Phe	Thr	Ala	Lys
Gly	Ala	Met	Ser	Thr	Gln	Gln	Ile	Lys	Lys	Ala	Glu	Glu	Ile	Ala	Asn
Glu	Met	Ile	Glu	Ala	Ala	Lys	Ala	Val	Tyr	Thr	Gln	Asp	Cys	Pro	Leu
Ala	Ala	Ala	Lys	Ala	Ile	Gln	Gly	Leu	Arg	Ala	Val	Phe	Asp	Glu	Thr
Tyr	Pro	Asp	Pro	Val	Arg	Val	Val	Ser	Ile	Gly	Val	Pro	Val	Ser	Glu
Leu	Leu	Asp	Asp	Pro	Ser	Gly	Pro	Ala	Gly	Ser	Leu	Thr	Ser	Val	Glu
Phe	Cys	Gly	Gly	Thr	His	Leu	Arg	Asn	Ser	Ser	His	Ala	Gly	Ala	Phe
Val	Ile	Val	Thr	Glu	Glu	Ala	Ile	Ala	Lys	Gly	Ile	Arg	Arg	Ile	Val
Ala	Val	Thr	Gly	Ala	Glu	Ala	Gln	Lys	Ala	Leu	Arg	Lys	Ala	Glu	Ser
Leu	Lys	Lys	Cys	Leu	Ser	Val	Met	Glu	Ala	Lys	Val	Lys	Ala	Gln	Thr
Ala	Pro	Asn	Lys	Asp	Val	Gln	Arg	Glu	Ile	Ala	Asp	Leu	Gly	Glu	Ala
Leu	Ala	Thr	Ala	Val	Ile	Pro	Gln	Trp	Gln	Lys	Asp	Glu	Leu	Arg	Glu
Thr	Leu	Lys	Ser	Leu	Lys	Lys	Val	Met	Asp	Asp	Leu	Asp	Arg	Ala	Ser
Lys	Ala	Asp	Val	Gln	Lys	Arg	Val	Leu	Glu	Lys	Thr	Lys	Gln	Phe	Ile
Asp	Ser	Asn	Pro	Asn	Gln	Pro	Leu	Val	Ile	Leu	Glu	Met	Glu	Ser	Gly
Ala	Ser	Ala	Lys	Ala	Leu	Asn	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Met	His
Ser	Pro	Gln	Thr	Ser	Ala	Met	Leu	Phe	Thr	Val	Asp	Asn	Glu	Ala	Gly
Lys	Ile	Thr	Cys	Leu	Cys	Gln	Val	Pro	Gln	Asn	Ala	Ala	Asn	Arg	Gly
Leu	Lys	Ala	Ser	Glu	Trp	Val	Gln	Val	Ser	Gly	Leu	Met	Asp	Gly	
Lys	Gly	Gly	Gly	Lys	Asp	Val	Ser	Ala	Gln	Ala	Thr	Gly	Lys	Asn	Val
Gly	Cys	Leu	Gln	Glu	Ala	Leu	Gln	Leu	Ala	Thr	Ser	Phe	Ala	Gln	Leu
Arg	Leu	Gly	Asp	Val	Lys	Asn									

<210> 119  
 <211> 1704  
 <212> DNA  
 <213> Homo sapiens

120161\_426PC\_Sequence\_Listing.TXT

<400> 119

```

atggactcta ctctaacagc aagtgaatc cggcagcgat ttatagattt cttcaagagg 60
aacgagcata cgtatgttca ctctgtgcc accatcccat tggatgaccc cactttgtct 120
tttgccaatg caggcatgaa ccagctgaaa tcacagggca agggagctgg tggggaagac 180
ctcattatgc tggacattta cgctatcgaa gagctccggg cacgggggtct ggaggtcaca 240
gatgattccc caaagtacaa ttaccatttg gactccagtg gtagctatgt atttgagaac 300
acagtggcta cggatgatgg tctgcgagg gagaagatgt tcgtggaaga ggtgtccaca 360
ggccaggagt gtggagtggg gctggacaag acctgtttct atgctgagca aggaggccag 420
atctatgacg aaggctacct ggtgaagggt gatgacagca gtgaagataa aacagagttt 480
acagtgaaga atgctcaggt ccgaggaggg tatgtgtctac acattggaac catctacggt 540
gacctgaaag tgggggatca ggtctggctg tttattgatg agccccgacg aagacccatc 600
atgagcaacc acacagctac gcacattctg aacttcgccc tgcgctcagt gcttggggaa 660
gctgaccaga aaggctcatt ggttgctcct gaccgcctca gatttgactt tactgccaag 720
ggagccatgt ccaccaaca gatcaagaag gctgaagaga ttgctaataa gatgattgag 780
gcagccaagg ccgtctatac ccaggattgc cccctggcag cagcgaaagc catccagggc 840
ctacgggctg tgtttgatga gacctatcct gacctgtgct gagtcgtctc cattgggggtc 900
ccggtgtccg agttgctgga tgacccctct gggcctgctg gctccctgac ttctgttgag 960
ttctgtgggg gaacgcacct gcggaactcg agtcatgcag gagcttttgt gatcgtgacg 1020
gaagaagcca ttgccaaggg tatccggagg attgtggctg tcacaggtgc cgaggcccag 1080
aaggccctca ggaaagcaga gagcttgaag aaatgtctct ctgtcatgga agccaaagtg 1140
aaggctcaga ctgctccaaa caaggaatgt cagagggaga tcgctgacct tggagaggcc 1200
ctggccactg cagtcatccc ccagtggcag aaggatgaat tgcgggagac tctcaaatcc 1260
ctaaagaagg tcatggatga cttggaccga gccagcaaag ccgatgtcca gaaacgagtg 1320
ttagagaaga cgaagcagtt catcgacagc aaccccaacc agcctcttgt catcctggag 1380
atggagagcg gcgctcagc caaggccctg aatgaagcct tgaagctctt caagatgcac 1440
tcccctcaga cttctgccat gctcttcacg gtggacaatg aggctggcaa gatcacgtgc 1500
ctgtgtcaag ttccccagaa tgcagccaat cggggcttaa aagccagcga gtgggtgcag 1560
cagggtgtcag gcttgatgga cggtaaaggt ggtggcaagg atgtgtctgc acaggccaca 1620
ggcaagaacg ttggctgcct gcaggaggcg ctgcagctgg ccacttcctt cgcccagctg 1680
cgccctcgggg atgtaaagaa ctga                                     1704

```

<210> 120

<211> 878

<212> PRT

<213> Homo sapiens

<400> 120

```

Met Asp Ser Thr Leu Thr Ala Ser Glu Ile Arg Gl n Arg Phe Ile Asp
 1          5          10          15
Phe Phe Lys Arg Asn Glu His Thr Tyr Val His Ser Ser Ala Thr Ile
 20          25          30
Pro Leu Asp Asp Pro Thr Leu Leu Phe Ala Asn Ala Gly Met Asn Gl n
 35          40          45
Phe Lys Pro Ile Phe Leu Asn Thr Ile Asp Pro Ser His Pro Met Ala
 50          55          60
Lys Leu Ser Arg Ala Ala Asn Thr Gl n Lys Cys Ile Arg Ala Gly Gly
 65          70          75          80
Lys His Asn Asp Leu Asp Asp Val Gly Lys Asp Val Tyr His His Thr
 85          90          95
Phe Phe Glu Met Leu Gly Ser Trp Ser Phe Gly Asp Tyr Phe Lys Glu
100          105          110
Leu Ala Cys Lys Met Ala Leu Glu Leu Leu Thr Gl n Glu Phe Gly Ile
115          120          125
Pro Ile Glu Arg Leu Tyr Val Thr Tyr Phe Gly Gly Asp Glu Ala Ala
130          135          140
Gly Leu Glu Ala Asp Leu Glu Cys Lys Gl n Ile Trp Gl n Asn Leu Gly
145          150          155          160
Leu Asp Asp Thr Lys Ile Leu Pro Gly Asn Met Lys Asp Asn Phe Trp
165          170          175
Glu Met Gly Asp Thr Gly Pro Cys Gly Pro Cys Ser Glu Ile His Tyr
180          185          190
Asp Arg Ile Gly Gly Arg Asp Ala Ala His Leu Val Asn Gl n Asp Asp
195          200          205
Pro Asn Val Leu Glu Ile Trp Asn Leu Val Phe Ile Gl n Tyr Asn Arg
210          215          220
Glu Ala Asp Gly Ile Leu Lys Pro Leu Pro Lys Lys Ser Ile Asp Thr
225          230          235          240

```

120161\_426PC\_Sequence\_Listing.TXT

Gly	Met	Gly	Leu	Glu	Arg	Leu	Val	Ser	Val	Leu	Gln	Asn	Lys	Met	Ser
				245					250					255	
Asn	Tyr	Asp	Thr	Asp	Leu	Phe	Val	Pro	Tyr	Phe	Glu	Ala	Ile	Gln	Lys
			260					265					270		
Gly	Thr	Gly	Ala	Arg	Pro	Tyr	Thr	Gly	Lys	Val	Gly	Ala	Glu	Asp	Ala
		275					280					285			
Asp	Gly	Ile	Asp	Met	Ala	Tyr	Arg	Val	Leu	Ala	Asp	His	Ala	Arg	Thr
	290					295					300				
Ile	Thr	Val	Ala	Leu	Ala	Asp	Gly	Gly	Arg	Pro	Asp	Asn	Thr	Gly	Arg
305					310					315					320
Gly	Tyr	Val	Leu	Arg	Arg	Ile	Leu	Arg	Arg	Ala	Val	Arg	Tyr	Ala	His
				325					330					335	
Glu	Lys	Leu	Asn	Ala	Ser	Arg	Gly	Phe	Phe	Ala	Thr	Leu	Val	Asp	Val
			340					345					350		
Val	Val	Gln	Ser	Leu	Gly	Asp	Ala	Phe	Pro	Glu	Leu	Lys	Lys	Asp	Pro
		355					360					365			
Asp	Met	Val	Lys	Asp	Ile	Ile	Asn	Glu	Glu	Glu	Val	Gln	Phe	Leu	Lys
	370					375					380				
Thr	Leu	Ser	Arg	Gly	Arg	Arg	Ile	Leu	Asp	Arg	Lys	Ile	Gln	Ser	Leu
385					390					395					400
Gly	Asp	Ser	Lys	Thr	Ile	Pro	Val	Phe	Glu	Asn	Thr	Val	Ala	Thr	Val
				405					410					415	
Met	Ala	Leu	Arg	Arg	Glu	Lys	Met	Phe	Val	Glu	Glu	Val	Ser	Thr	Gly
			420					425					430		
Gln	Glu	Cys	Gly	Val	Val	Leu	Asp	Lys	Thr	Cys	Phe	Tyr	Ala	Glu	Gln
		435					440					445			
Gly	Gly	Gln	Ile	Tyr	Asp	Glu	Gly	Tyr	Leu	Val	Lys	Val	Asp	Asp	Ser
	450					455					460				
Ser	Glu	Asp	Lys	Thr	Glu	Phe	Thr	Val	Lys	Asn	Ala	Gln	Val	Arg	Gly
465					470					475					480
Gly	Tyr	Val	Leu	His	Ile	Gly	Thr	Ile	Tyr	Gly	Asp	Leu	Lys	Val	Gly
				485					490					495	
Asp	Gln	Val	Trp	Leu	Phe	Ile	Asp	Glu	Pro	Arg	Arg	Arg	Pro	Ile	Met
			500					505					510		
Ser	Asn	His	Thr	Ala	Thr	His	Ile	Leu	Asn	Phe	Ala	Leu	Arg	Ser	Val
		515					520					525			
Leu	Gly	Glu	Ala	Asp	Gln	Lys	Gly	Ser	Leu	Val	Ala	Pro	Asp	Arg	Leu
	530					535					540				
Arg	Phe	Asp	Phe	Thr	Ala	Lys	Gly	Ala	Met	Ser	Thr	Gln	Gln	Ile	Lys
545					550					555					560
Lys	Ala	Glu	Glu	Ile	Ala	Asn	Glu	Met	Ile	Glu	Ala	Ala	Lys	Ala	Val
				565					570					575	
Tyr	Thr	Gln	Asp	Cys	Pro	Leu	Ala	Ala	Lys	Ala	Ile	Gln	Gly	Leu	
			580					585					590		
Arg	Ala	Val	Phe	Asp	Glu	Thr	Tyr	Pro	Asp	Pro	Val	Arg	Val	Val	Ser
		595					600					605			
Ile	Gly	Val	Pro	Val	Ser	Glu	Leu	Leu	Asp	Asp	Pro	Ser	Gly	Pro	Ala
	610					615					620				
Gly	Ser	Leu	Thr	Ser	Val	Glu	Phe	Cys	Gly	Gly	Thr	His	Leu	Arg	Asn
625					630					635					640
Ser	Ser	His	Ala	Gly	Ala	Phe	Val	Ile	Val	Thr	Glu	Glu	Ala	Ile	Ala
				645					650					655	
Lys	Gly	Ile	Arg	Arg	Ile	Val	Ala	Val	Thr	Gly	Ala	Glu	Ala	Gln	Lys
			660					665					670		
Ala	Leu	Arg	Lys	Ala	Glu	Ser	Leu	Lys	Lys	Cys	Leu	Ser	Val	Met	Glu
		675					680					685			
Ala	Lys	Val	Lys	Ala	Gln	Thr	Ala	Pro	Asn	Lys	Asp	Val	Gln	Arg	Glu
	690					695					700				
Ile	Ala	Asp	Leu	Gly	Glu	Ala	Leu	Ala	Thr	Ala	Val	Ile	Pro	Gln	Trp
705					710					715					720
Gln	Lys	Asp	Glu	Leu	Arg	Glu	Thr	Leu	Lys	Ser	Leu	Lys	Lys	Val	Met
				725					730					735	
Asp	Asp	Leu	Asp	Arg	Ala	Ser	Lys	Ala	Asp	Val	Gln	Lys	Arg	Val	Leu
			740					745					750		
Glu	Lys	Thr	Lys	Gln	Phe	Ile	Asp	Ser	Asn	Pro	Asn	Gln	Pro	Leu	Val
		755					760					765			
Ile	Leu	Glu	Met	Glu	Ser	Gly	Ala	Ser	Ala	Lys	Ala	Leu	Asn	Glu	Ala
	770					775					780				

120161\_426PC\_Sequence\_Listing.TXT

Leu	Lys	Leu	Phe	Lys	Met	His	Ser	Pro	Gln	Thr	Ser	Ala	Met	Leu	Phe
785					790					795					800
Thr	Val	Asp	Asn	Glu	Ala	Gly	Lys	Ile	Thr	Cys	Leu	Cys	Gln	Val	Pro
				805					810					815	
Gln	Asn	Ala	Ala	Asn	Arg	Gly	Leu	Lys	Ala	Ser	Glu	Trp	Val	Gln	Gln
			820					825					830		
Val	Ser	Gly	Leu	Met	Asp	Gly	Lys	Gly	Gly	Gly	Lys	Asp	Val	Ser	Ala
		835					840					845			
Gln	Ala	Thr	Gly	Lys	Asn	Val	Gly	Cys	Leu	Gln	Glu	Ala	Leu	Gln	Leu
	850					855					860				
Ala	Thr	Ser	Phe	Ala	Gln	Leu	Arg	Leu	Gly	Asp	Val	Lys	Asn		
865					870					875					

<210> 121  
 <211> 2637  
 <212> DNA  
 <213> Homo sapiens

<400> 121  
 atggactcta ctctaacagc aagtgaatc cggcagcgat ttatagattt cttcaagagg 60  
 aacgagcata cgtatgttca ctctgtgccc accatcccat tggatgaccc cactttgtct 120  
 ttgccaatg caggcatgaa ccagtttaaa cccattttcc tgaacacaat tgacccatct 180  
 caccatcatg caaagctgag cagagctgcc aatacccaga agtgcacccg ggctgggggc 240  
 aaacataatg acctggacga tgtgggcaag gatgtctatc atcacacctt cttcgagatg 300  
 ctgggctctt ggtcttttgg agattacttt aaggaattgg catgtaagat ggctctggaa 360  
 ctctcaccac aagagtttgg cattccattt gaaagacttt atgttactta ctttggcggg 420  
 gatgaagcag ctggcttaga agcagatctg gaatgcaaac agatctggca aaatttgggg 480  
 ctggatgaca ccaaaatcct cccaggcaac atgaaggata acttctggga gatgggtgac 540  
 acgggcccct gtggtccttg cagtgaatc cactacgacc ggattggttg tcgggacgcc 600  
 gcacatcttg tcaaccagga cgaccctaata gtgctggaga tctggaacct tgtgttcac 660  
 cagtataaca ggaagctga tggcattctg aaacctcttc ccaagaaaag cattgacaca 720  
 gggatggggc tggaaagact ggtatctgtg ctgcagaata agatgtccaa ctatgacact 780  
 gaccttttgg tcccttactt tgaagccatt cagaaggcca cagggtcccc accatacact 840  
 gggaaagtgt gtgctgagga tgccgatggg attgacatgg cctaccgggt gctggctgac 900  
 cagctcggga ccatcactgt ggcactggct gatggtggcc ggcctgacaa cacaggcgct 960  
 ggaatgtgtg tgagacggat tctccgcca gctgtccgat acgcccata aaagctcaat 1020  
 gccagcaggg gcttctttgc tacgttagtg gatgtgtcgc tccagtccct gggagatgca 1080  
 tttctgagc tgaagaagga cccagacatg gtgaaggaca tcattaatga agaagagggt 1140  
 cagtttctca agactctcag cagagggcgt cgcacacctg acaggaaaat tcagagcctg 1200  
 ggagacagca agaccattcc cgtatttgag aacacagtgg ctacggtgat ggctctgcgc 1260  
 agggagaaga tgttcgtgga agaggtgtcc acaggccagg agtgtggagt ggtgctggac 1320  
 aagcactgtt tctatgtcta gcaaggaggc cagatctatg acgaaggcta cctggtgaag 1380  
 tggatgaca gcatggaaga taaaacagag ttacagtga agaatgctca ggtccgagga 1440  
 gggatgtgtc tacacatttg aaccatctac ggtgacctga aagtggggga tcaggtctgg 1500  
 ctgtttattg atgagccccg acgaagacc atcatgagca accacacagc tacgcacatt 1560  
 ctgaacttcg ccctgcgctc agtgcttggg gaagctgacc agaaaggctc attggttgct 1620  
 cctgaccgcc tcagatttga ctttactgcc aagggagcca tgtccacca acagatcaag 1680  
 aaggctgaag agattgctaa tgagatgatt gaggcagcca aggcgtcta taccaggat 1740  
 tgccccctgg cagcagcgaa agccatccag ggcctacggg ctgtgtttga tgagacctat 1800  
 cctgaccctg tgcgagtcgt ctccattggg gtcccgtgtt ccgagttgct ggatgacccc 1860  
 tctgggacct ctggctccct gacttctgtt gagttctgtg ggggaacgca cctgcggaac 1920  
 tcgagtcatt caggagcttt tgtgatcgtg acggaagaag ccattgccaa gggatccgg 1980  
 aggattgtgg ctgtcacagg tgccgagggc cagaaggccc tcaggaaagc agagagcttg 2040  
 aagaaatgtc tctctgtcat ggaagcctaa gtgaaggctc agactgctcc aaacaaggat 2100  
 gtgcagaggg agatcgctga ccttgagag ggcctggcca ctgcagtcac cccccagtg 2160  
 cagaaggatg aattgcggga gactctcaaa tccctaaaga aggtcatgga tgacttggac 2220  
 cgagccagca aagccgatgt ccagaaacga gtgttagaga agacgaagca gtcatcgac 2280  
 agcaacccca accagcctct tgtcatcctg gagatggaga gcgcgccctc agccaaggcc 2340  
 ctgaatgaag ccttgaagct cttcaagatg cactcccctc agacttctgc catgctcttc 2400  
 acggtggaca atgagctgg caagatcacg tgcctgtgtc aagttcccca gaatgcagc 2460  
 aatcggggct taaaagccag cgagtgggtg cagcaggtgt caggcttgat ggacggtaaa 2520  
 ggtgtgtgca aggatgtgtc tgcacaggcc acaggcaaga acgttggtct cctgcaggag 2580  
 gcgctgcagc tggccacttc cttcgcccag ctgcgcctcg gggatgtaaa gaactga 2637

<210> 122  
 <211> 427  
 <212> PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 122

```

Met 1 Ile Pro Gln Ser 5 Thr 11e Thr 11e Trp 10 Thr Pro Val Val Ala 15 Met
Lys Thr Glu Phe 20 Thr Val Lys Asn 25 Ala 10 Val Arg Gly Gly 30 Tyr Val
Leu His 35 Ile Gly Thr 11e Tyr 40 Gly Asp Leu Lys Val Gly 45 Asp Gln Val
Trp 50 Leu Phe 11e Asp Glu 55 Pro Arg Arg Arg Pro 60 Ile Met Ser Asn His
Thr 65 Ala Thr His 11e Leu 70 Asn Phe Ala Leu Arg 75 Ser Val Leu Gly Glu 80
Ala Asp Gln Lys 85 Gly Ser Leu Val Ala 90 Pro Asp Arg Leu Arg Phe 95 Asp
Phe Thr Ala Lys 100 Gly Ala Met Ser Thr 105 Gln Gln 11e Lys Lys Ala Glu
Glu 115 Ala Asn Glu Met 11e Glu 120 Ala Ala Lys Ala Val Tyr Thr Gln
Asp Cys 130 Pro Leu Ala Ala 135 Lys Ala 11e Gln Gly 140 Leu Arg Ala Val
Phe 145 Asp Glu Thr Tyr 150 Pro Asp Pro Val Arg Val Val Ser 11e Gly Val 160
Pro Val Ser Glu 165 Leu Leu Asp Asp Pro Ser 170 Gly Pro Ala Gly Ser 175 Leu
Thr Ser Val 180 Glu Phe Cys Gly Gly Thr 185 His Leu Arg Asn Ser 190 Ser His
Ala Gly Ala 195 Phe Val 11e Val Thr 200 Glu Glu Ala 11e Ala 205 Lys Gly 11e
Arg Arg 210 Ile Val Ala Val Thr 215 Gly Ala Glu Ala Gln Lys Ala Leu Arg
Lys 225 Ala Glu Ser Leu 230 Lys Cys Leu Ser Val 235 Met Glu Ala Lys Val 240
Lys Ala Gln Thr Ala 245 Pro Asn Lys Asp Val 250 Gln Arg Glu 11e Ala Asp 255
Leu Gly Glu Ala 260 Leu Ala Thr Ala Val 265 Ile Pro Gln Trp Gln Lys Asp 270
Glu Leu Arg 275 Glu Thr Leu Lys Ser 280 Leu Lys Lys Val Met 285 Asp Asp Leu
Asp Arg 290 Ala Ser Lys Ala 295 Asp Val Gln Lys Arg Val 300 Leu Glu Lys Thr
Lys 305 Gln Phe 11e Asp Ser 310 Asn Pro Asn Gln Pro 315 Leu Val 11e Leu Glu 320
Met Glu Ser Gly Ala 325 Ser Ala Lys Ala Leu Asn Glu Ala Leu Lys 335 Leu
Phe Lys Met His 340 Ser Pro Gln Thr Ser 345 Ala Met Leu Phe Thr 350 Val Asp
Asn Glu Ala 355 Gly Lys 11e Thr Cys 360 Leu Cys Gln Val Pro Gln Asn Ala
Ala Asn 370 Arg Gly Leu Lys Ala 375 Ser Glu Trp Val Gln Gln Val Ser Gly
Leu 385 Met Asp Gly Lys Gly 390 Gly Gly Lys Asp Val 395 Ser Ala Gln Ala Thr 400
Gly Lys Asn Val Gly 405 Cys Leu Gln Glu Ala 410 Leu Gln Leu Ala Thr 415 Ser
Phe Ala Gln Leu 420 Arg Leu Gly Asp Val 425 Lys Asn

```

&lt;210&gt; 123

&lt;211&gt; 1284

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 123

```

atgattcccc aaagtacaat taccatttgg actccagttg tagctatgaa aacagagttt 60
acagtgaaga atgctcaggt ccgaggagg tatgtgctac acattggaac catctacggt 120
gacctgaaag tgggggatca ggtctggctg tttattgatg agccccgacg aagacccatc 180

```

120161\_426PC\_Sequence\_Listing.TXT

```

atgagcaacc acacagctac gcacattctg aacttcgccc tgcgctcagt gcttggggaa 240
gctgaccaga aaggctcatt ggttgctcct gaccgcctca gatttgactt tactgccaag 300
ggagccatgt ccaccaaca gatcaagaag gctgaagaga ttgctaata gatgattgag 360
gcagccaagg ccgtctatac ccaggattgc cccctggcag cagcgaaagc catccagggc 420
ctacgggctg tgtttgatga gacctatcct gaccctgtgc gagtcgtctc cattgggggtc 480
ccggtgtccg agttgctgga tgacccctct gggcctgctg gctccctgac ttctgttgag 540
ttctgtgggg gaacgcacct gcggaactcg agtcatgcag gagcttttgt gatcgtgacg 600
gaagaagcca ttgccaaggg tatccggagg attgtggctg tcacagggtg cgaggcccag 660
aaggccctca ggaaagcaga gagcttgaag aaatgtctct ctgtcatgga agccaaagtg 720
aaggctcaga ctgctccaaa caaggatgtg cagagggaga tcgctgacct tggagaggcc 780
ctggccactg cagtcatccc ccagtggcag aaggatgaat tgcgggagac tcicaaatcc 840
ctaaagaagg tcatggatga cttggaccga gccagcaaag ccgatgtcca gaaacgagtg 900
ttagagaaga cgaagcagtt catcgacagc aaccccaacc agcctcttgt catcctggag 960
atggagagcg gcgcctcagc caaggccctt tgaagctctt caagatgcac 1020
tcccctcaga cttctgccat gctcttcacg gtggacaatg aggctggcaa gatcacgtgc 1080
ctgtgtcaag ttccccagaa tgcagccaat cggggcttaa aagccagcga gtgggtgcag 1140
cagggtgtcag gcttgatgga cggtaaaggt ggtggcaagg atgtgtctgc acaggccaca 1200
ggcaagaacg ttggctgcct gcaggaggcg ctgcagctgg ccacttcctt cgcccagctg 1260
cgccctcgggg atgtaaagaa ctga 1284

```

<210> 124  
 <211> 244  
 <212> PRT  
 <213> Homo sapiens

```

<400> 124
Met Arg His Leu Arg Asn Ser Ser His Ala Gly Ala Phe Val Ile Val
 1      5      10      15
Thr Glu Glu Ala Ile Ala Lys Gly Ile Arg Arg Ile Val Ala Val Thr
      20      25      30
Gly Ala Glu Ala Gln Lys Ala Leu Arg Lys Ala Glu Ser Leu Lys Lys
      35      40      45
Cys Leu Ser Val Met Glu Ala Lys Val Lys Ala Gln Thr Ala Pro Asn
      50      55      60
Lys Asp Val Gln Arg Glu Ile Ala Asp Leu Gly Glu Ala Leu Ala Thr
      65      70      75      80
Ala Val Ile Pro Gln Trp Gln Lys Asp Glu Leu Arg Glu Thr Leu Lys
      85      90      95
Ser Leu Lys Lys Val Met Asp Asp Leu Asp Arg Ala Ser Lys Ala Asp
      100      105      110
Val Gln Lys Arg Val Leu Glu Lys Thr Lys Gln Phe Ile Asp Ser Asn
      115      120      125
Pro Asn Gln Pro Leu Val Ile Leu Glu Met Glu Ser Gly Ala Ser Ala
      130      135      140
Lys Ala Leu Asn Glu Ala Leu Lys Leu Phe Lys Met His Ser Pro Gln
      145      150      155      160
Thr Ser Ala Met Leu Phe Thr Val Asp Asn Glu Ala Gly Lys Ile Thr
      165      170      175
Cys Leu Cys Gln Val Pro Gln Asn Ala Ala Asn Arg Gly Leu Lys Ala
      180      185      190
Ser Glu Trp Val Gln Gln Val Ser Gly Leu Met Asp Gly Lys Gly Gly
      195      200      205
Gly Lys Asp Val Ser Ala Gln Ala Thr Gly Lys Asn Val Gly Cys Leu
      210      215      220
Gln Glu Ala Leu Gln Leu Ala Thr Ser Phe Ala Gln Leu Arg Leu Gly
      225      230      235      240
Asp Val Lys Asn

```

<210> 125  
 <211> 735  
 <212> DNA  
 <213> Homo sapiens

```

<400> 125
atgaggcacc tgcggaactc gagtcatgca ggagcttttg tgatcgtgac ggaagaagcc 60
attgccaagg gtatccggag gattgtggct gtcacagggt ccgaggccca gaaggccctc 120

```



120161\_426PC\_Sequence\_Listing.TXT

```

aggaaagcag agagcttgaa gaaatgtctc tctgtcatgg aagccaaagt gaaggctcag 180
actgtctcaa acaaggatgt gcagagggag atcgctgacc ttggagaggc cctggccact 240
gcagtcatcc cccagtgcca gaaggatgaa ttgcgggaga ctctcaaadc cctaaagaag 300
gtcatggatg acttgaccg agccagcaaa gccgatgtcc agaaacgagt gttagagaag 360
acgaagcagt tcatcgacag caacccaac cagcctcttg tcatcctgga gatggagagc 420
ggcgcctcag ccaaggccct gaatgaagcc ttgaagctct tcaagatgca ctcccctcag 480
acttctgcca tgctcttcac ggtggacaat gaggctggca agatcacgtg cctgtgtcaa 540
gttccccaga atgcagccaa tcggggctta aaagccagcg agtgggtgca gcagggtgtca 600
ggccttgatg acggtaaagg tgggtggcaag gatgtgtctg cacaggccac aggcaagaac 660
gttggctgcc tgcaggaggc gctgcagctg gccacttcct tcgccagct gcgcctcggg 720
gatgtaaaga actga                                     735

```

<210> 126

<211> 599

<212> PRT

<213> Homo sapiens

<400> 126

```

Met Val Lys Asp Ile Ile Asn Glu Glu Glu Val Glu Phe Leu Lys Thr
1      5      10      15
Leu Ser Arg Gly Arg Arg Ile Leu Asp Arg Lys Ile Glu Ser Leu Gly
20     25     30
Asp Ser Lys Thr Ile Pro Gly Asp Thr Ala Trp Leu Leu Tyr Asp Thr
35     40     45
Tyr Gly Phe Pro Val Asp Leu Thr Gly Leu Ile Ala Glu Glu Lys Gly
50     55     60
Leu Val Val Asp Met Asp Gly Phe Glu Glu Glu Arg Lys Leu Ala Glu
65     70     75     80
Leu Lys Ser Glu Gly Lys Gly Ala Gly Gly Glu Asp Leu Ile Met Leu
85     90     95
Asp Ile Tyr Ala Ile Glu Glu Leu Arg Ala Arg Gly Leu Glu Val Thr
100    105    110
Asp Asp Ser Pro Lys Tyr Asn Tyr His Leu Asp Ser Ser Gly Ser Tyr
115    120    125
Val Phe Glu Asn Thr Val Ala Thr Val Met Ala Leu Arg Arg Glu Lys
130    135    140
Met Phe Val Glu Glu Val Ser Thr Gly Glu Glu Cys Gly Val Val Leu
145    150    155    160
Asp Lys Thr Cys Phe Tyr Ala Glu Glu Gly Gly Glu Ile Tyr Asp Glu
165    170    175
Gly Tyr Leu Val Lys Val Asp Asp Ser Ser Glu Asp Lys Thr Glu Phe
180    185    190
Thr Val Lys Asn Ala Glu Val Arg Gly Gly Tyr Val Leu His Ile Gly
195    200    205
Thr Ile Tyr Gly Asp Leu Lys Val Gly Asp Glu Val Trp Leu Phe Ile
210    215    220
Asp Glu Pro Arg Arg Arg Pro Ile Met Ser Asn His Thr Ala Thr His
225    230    235    240
Ile Leu Asn Phe Ala Leu Arg Ser Val Leu Gly Glu Ala Asp Glu Lys
245    250    255
Gly Ser Leu Val Ala Pro Asp Arg Leu Arg Phe Asp Phe Thr Ala Lys
260    265    270
Gly Ala Met Ser Thr Glu Glu Ile Lys Lys Ala Glu Glu Ile Ala Asn
275    280    285
Glu Met Ile Glu Ala Ala Lys Ala Val Tyr Thr Glu Asp Cys Pro Leu
290    295    300
Ala Ala Ala Lys Ala Ile Glu Gly Leu Arg Ala Val Phe Asp Glu Thr
305    310    315    320
Tyr Pro Asp Pro Val Arg Val Val Ser Ile Gly Val Pro Val Ser Glu
325    330    335
Leu Leu Asp Asp Pro Ser Gly Pro Ala Gly Ser Leu Thr Ser Val Glu
340    345    350
Phe Cys Gly Gly Thr His Leu Arg Asn Ser Ser His Ala Gly Ala Phe
355    360    365
Val Ile Val Thr Glu Glu Ala Ile Ala Lys Gly Ile Arg Arg Ile Val
370    375    380
Ala Val Thr Gly Ala Glu Ala Glu Lys Ala Leu Arg Lys Ala Glu Ser
385    390    395    400

```

120161\_426PC\_Sequence\_Listing.TXT

Leu	Lys	Lys	Cys	Leu	Ser	Val	Met	Glu	Ala	Lys	Val	Lys	Ala	Gln	Thr
				405					410					415	
Ala	Pro	Asn	Lys	Asp	Val	Gln	Arg	Glu	Ile	Ala	Asp	Leu	Gly	Glu	Ala
			420					425					430		
Leu	Ala	Thr	Ala	Val	Ile	Pro	Gln	Trp	Gln	Lys	Asp	Glu	Leu	Arg	Glu
		435					440					445			
Thr	Leu	Lys	Ser	Leu	Lys	Lys	Val	Met	Asp	Asp	Leu	Asp	Arg	Ala	Ser
	450					455					460				
Lys	Ala	Asp	Val	Gln	Lys	Arg	Val	Leu	Glu	Lys	Thr	Lys	Gln	Phe	Ile
465					470					475					480
Asp	Ser	Asn	Pro	Asn	Gln	Pro	Leu	Val	Ile	Leu	Glu	Met	Glu	Ser	Gly
				485					490					495	
Ala	Ser	Ala	Lys	Ala	Leu	Asn	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Met	His
			500					505					510		
Ser	Pro	Gln	Thr	Ser	Ala	Met	Leu	Phe	Thr	Val	Asp	Asn	Glu	Ala	Gly
		515					520					525			
Lys	Ile	Thr	Cys	Leu	Cys	Gln	Val	Pro	Gln	Asn	Ala	Ala	Asn	Arg	Gly
	530					535					540				
Leu	Lys	Ala	Ser	Glu	Trp	Val	Gln	Gln	Val	Ser	Gly	Leu	Met	Asp	Gly
545					550					555					560
Lys	Gly	Gly	Gly	Lys	Asp	Val	Ser	Ala	Gln	Ala	Thr	Gly	Lys	Asn	Val
				565					570					575	
Gly	Cys	Leu	Gln	Glu	Ala	Leu	Gln	Leu	Ala	Thr	Ser	Phe	Ala	Gln	Leu
			580					585					590		
Arg	Leu	Gly	Asp	Val	Lys	Asn									
		595													

<210> 127  
 <211> 1800  
 <212> DNA  
 <213> Homo sapiens

<400> 127

atggtgaagg	acatcattaa	tgaagaagag	gtgcagtttc	tcaagactct	cagcagaggg	60
cgtcgcatcc	tggacaggaa	aattcagagc	ctgggagaca	gcaagaccat	tcccggagac	120
actgcttgcc	tcctctatga	cacctatggg	tttccagtgg	atctgactgg	actgattgct	180
gaagagaagg	gcctggtggt	agacatggat	ggctttgaag	aggagaggaa	actggcccag	240
ctgaaatcac	ggggcaaggg	agctggtggg	gaagacctca	ttatgctgga	cattttacgt	300
atcgaagagc	tccgggcacg	gggtctggag	gtcacagatg	attccccaaa	gtacaattac	360
catitggact	ccagtggtag	ctatgtatit	gagaacacag	tggctacggg	gatggctctg	420
cgcagggaga	agatgttcgt	ggaagaggtg	tccacaggcc	aggagtgtgg	agtgggtgctg	480
gacaagacct	gtttctatgc	tgagcaagga	ggccagatct	atgacgaagg	ctacctgggtg	540
aaggtggatg	acagcagtga	agataaaaca	gagtttacag	tgaagaatgc	tcagggtccga	600
ggaggggatg	tgctacacat	tggaaaccatc	tacggtgacc	tgaagtggg	ggatcagggtc	660
tggctgttta	ttgatgagcc	ccgacgaaga	cccatcatga	gcaaccacac	agctacgcac	720
attctgaact	tcgccctgcg	ctcagtgcct	ggggaagctg	accagaaagg	ctcattgggt	780
gctcctgacc	gcctcagatt	tgactttact	gccaaaggag	ccatgtccac	ccaacagatc	840
aagaaggctg	aagagattgc	taatgagatg	attgaggcag	ccaaggccgt	ctataccag	900
gattgcccc	tggcagcagc	gaaagccatc	cagggcctac	gggctgtgtt	tgatgagacc	960
tatcctgacc	ctgtgcgagt	cgtctccatt	ggggtcccg	tgtccgagtt	gctggatgac	1020
ccctctgggc	ctgctggctc	cctgacttct	gttgagttct	gtgggggaac	gcacctgcgg	1080
aactcgagtc	atgcaggagc	ttttgtgatc	gtgacggaag	aagccattgc	caagggtatc	1140
cggaggattg	tggctgtcac	aggtgccgag	gccagaagg	ccctcaggaa	agcagagagc	1200
ttgaagaaat	gtctctctgt	catggaagcc	aaagtgaagg	ctcagactgc	tccaaacaag	1260
gatgtgcaga	gggagatcgc	tgaccttgga	gaggccctgg	ccactgcagt	catccccag	1320
tggcagaagg	atgaattgcg	ggagactctc	aaatccctaa	agaaggtcat	ggatgacttg	1380
gaccgagcca	gcaaagccga	tgtccagaaa	cgagtgttag	agaagacgaa	gcagttcatc	1440
gacagcaacc	ccaaccagcc	tcttgtcatc	ctggagatgg	agagcggcgc	ctcagccaag	1500
gccctgaatg	aagccttgaa	gctcttcaag	atgcactccc	ctcagacttc	tgccatgctc	1560
ttcacggtgg	acaatgaggc	tggcaagatc	acgtgcctgt	gtcaagttcc	ccagaatgca	1620
gccaatcggg	gcttaaaagc	cagcgagtgg	gtgcagcagg	tgtcaggctt	gatggacggg	1680
aaaggtggtg	gcaaggatgt	gtctgcacag	gccacaggca	agaacgttgg	ctgcctgcag	1740
gaggcgctgc	agctggccac	ttccttcgcc	cagctgcgc	tcggggatgt	aaagaactga	1800

<210> 128  
 <211> 50

120161\_426PC\_Sequence\_Listing.TXT

```

<212> DNA
<213> Homo sapiens

<400> 128
accttggtggt catccagtat aacagggcac aggtgcccga ccatacactg      50

<210> 129
<211> 50
<212> DNA
<213> Homo sapiens

<400> 129
ctttgccaat gcaggcatga accagctgaa atcacagggc aaggagctg      50

<210> 130
<211> 16
<212> PRT
<213> Homo sapiens

<400> 130
Phe Ala Asn Ala Gly Met Asn Gl n Leu Lys Ser Gl n Gly Lys Gly Ala
 1          5          10          15

<210> 131
<211> 50
<212> DNA
<213> Homo sapiens

<400> 131
ctgggagaca gcaagaccat tcccgtatatt gagaacacag tggctacggt      50

<210> 132
<211> 16
<212> PRT
<213> Homo sapiens

<400> 132
Leu Gly Asp Ser Lys Thr Ile Pro Val Phe Gl u Asn Thr Val Ala Thr
 1          5          10          15

<210> 133
<211> 50
<212> DNA
<213> Homo sapiens

<400> 133
catttggtgact ccagtggtag ctatgaaaac agagttttaca gtgaagaatg      50

<210> 134
<211> 16
<212> PRT
<213> Homo sapiens

<400> 134
Ile Trp Thr Pro Val Val Ala Met Lys Thr Gl u Phe Thr Val Lys Asn
 1          5          10          15

<210> 135
<211> 50
<212> DNA
<213> Homo sapiens

<400> 135
tcaggctctgg ctgtttattg atgaggcacc tgcggaactc gagtcatgca      50

```

## 120161\_426PC\_Sequence\_Listing.TXT

<210> 136  
 <211> 10  
 <212> PRT  
 <213> Homo sapiens

<400> 136  
 Met Arg His Leu Arg Asn Ser Ser His Ala  
 1 5 10

<210> 137  
 <211> 50  
 <212> DNA  
 <213> Homo sapiens

<400> 137  
 gccggcctga caacacaggg cgtggggaga tgcatttcct gagctgaaga

50

<210> 138  
 <211> 310  
 <212> PRT  
 <213> Homo sapiens

<400> 138  
 Met Ile Glu Ala Ala Lys Ala Val Tyr Thr Gln Asp Cys Pro Leu Ala  
 1 5 10 15  
 Ala Ala Lys Ala Ile Gln Gly Leu Arg Ala Val Phe Asp Glu Thr Tyr  
 20 25 30  
 Pro Asp Pro Val Arg Val Val Ser Ile Gly Val Pro Val Ser Glu Leu  
 35 40 45  
 Leu Asp Asp Pro Ser Gly Pro Ala Gly Ser Leu Thr Ser Val Glu Phe  
 50 55 60  
 Cys Gly Gly Thr His Leu Arg Asn Ser Ser His Ala Gly Ala Phe Val  
 65 70 75 80  
 Ile Val Thr Glu Glu Ala Ile Ala Lys Gly Ile Arg Arg Ile Val Ala  
 85 90 95  
 Val Thr Gly Ala Glu Ala Gln Lys Ala Leu Arg Lys Ala Glu Ser Leu  
 100 105 110  
 Lys Lys Cys Leu Ser Val Met Glu Ala Lys Val Lys Ala Gln Thr Ala  
 115 120 125  
 Pro Asn Lys Asp Val Gln Arg Glu Ile Ala Asp Leu Gly Glu Ala Leu  
 130 135 140  
 Ala Thr Ala Val Ile Pro Gln Trp Gln Lys Asp Glu Leu Arg Glu Thr  
 145 150 155 160  
 Leu Lys Ser Leu Lys Lys Val Met Asp Asp Leu Asp Arg Ala Ser Lys  
 165 170 175  
 Ala Asp Val Gln Lys Arg Val Leu Glu Lys Thr Lys Gln Phe Ile Asp  
 180 185 190  
 Ser Asn Pro Asn Gln Pro Leu Val Ile Leu Glu Met Glu Ser Gly Ala  
 195 200 205  
 Ser Ala Lys Ala Leu Asn Glu Ala Leu Lys Leu Phe Lys Met His Ser  
 210 215 220  
 Pro Gln Thr Ser Ala Met Leu Phe Thr Val Asp Asn Glu Ala Gly Lys  
 225 230 235 240  
 Ile Thr Cys Leu Cys Gln Val Pro Gln Asn Ala Ala Asn Arg Gly Leu  
 245 250 255  
 Lys Ala Ser Glu Trp Val Gln Gln Val Ser Gly Leu Met Asp Gly Lys  
 260 265 270  
 Gly Gly Gly Lys Asp Val Ser Ala Gln Ala Thr Gly Lys Asn Val Gly  
 275 280 285  
 Cys Leu Gln Glu Ala Leu Gln Leu Ala Thr Ser Phe Ala Gln Leu Arg  
 290 295 300  
 Leu Gly Asp Val Lys Asn  
 305 310

<210> 139  
 <211> 933

120161\_426PC\_Sequence\_Listing.TXT

<212> DNA

<213> Homo sapiens

<400> 139

```

atgattgagg cagccaaggc cgtctatacc caggattgcc ccctggcagc agcgaaagcc 60
atccagggcc tacgggctgt gtttgatgag acctatcctg accctgtgcg agtcgtctcc 120
attgggggtcc cgggtgtccga gttgctggat gacccctctg ggcctgctgg ctccctgact 180
tctgttgagt tctgtggggg aacgcacctg cggaactcga gtcatgcagg agcttttgtg 240
atcgtgacgg aagaagccat tgccaagggt atccggagga ttgtggctgt cacagggtgcc 300
gaggcccaga aggccctcag gaaagcagag agcttgaaga aatgtctctc tgtcatggaa 360
gccaaagtga aggtcagac tgcaccaaac aaggatgtgc agagggagat cgctgacctt 420
ggagaggccc tggccactgc agtcatcccc cagtggcaga aggatgaatt gcgggagact 480
ctcaaattccc taaagaaggt catggatgac ttggaccgag ccagcaaagc cgaatgtccag 540
aaacgagtgt tagagaagac gaagcagttc atcgacagca accccaacca gcctcttgtc 600
atcctggaga tggagagcgg cgcctcagcc aaggccctga atgaagcctt gaagctcttc 660
aagatgcact cccctcagac ttctgccatg ctcttcacgg tggacaatga ggctggcaag 720
atcacgtgcc tgtgtcaagt tccccagaat gcagccaatc ggggcttaaa agccagcgag 780
tgggtgcagc aggtgtcagg cttgatggac ggtaaagggt gtggcaagga tgtgtctgca 840
caggccacag gcaagaacgt tggctgcctg caggaggcgc tgcagctggc cacttccttc 900
gccagctgc gcctcgggga tgtaaagaac tga 933

```

<210> 140

<211> 278

<212> PRT

<213> Homo sapiens

<400> 140

```

Pro Asp Pro Val Arg Val Val Ser Ile Gly Val Pro Val Ser Glu Leu
 1          5          10          15
Leu Asp Asp Pro Ser Gly Pro Ala Gly Ser Leu Thr Ser Val Glu Phe
 20          25          30
Cys Gly Gly Thr His Leu Arg Asn Ser Ser His Ala Gly Ala Phe Val
 35          40          45
Ile Val Thr Glu Glu Ala Ile Ala Lys Gly Ile Arg Arg Ile Val Ala
 50          55          60
Val Thr Gly Ala Glu Ala Gln Lys Ala Leu Arg Lys Ala Glu Ser Leu
 65          70          75          80
Lys Lys Cys Leu Ser Val Met Glu Ala Lys Val Lys Ala Gln Thr Ala
 85          90          95
Pro Asn Lys Asp Val Gln Arg Glu Ile Ala Asp Leu Gly Glu Ala Leu
100          105          110
Ala Thr Ala Val Ile Pro Gln Trp Gln Lys Asp Glu Leu Arg Glu Thr
115          120          125
Leu Lys Ser Leu Lys Lys Val Met Asp Asp Leu Asp Arg Ala Ser Lys
130          135          140
Ala Asp Val Gln Lys Arg Val Leu Glu Lys Thr Lys Gln Phe Ile Asp
145          150          155          160
Ser Asn Pro Asn Gln Pro Leu Val Ile Leu Glu Met Glu Ser Gly Ala
165          170          175
Ser Ala Lys Ala Leu Asn Glu Ala Leu Lys Leu Phe Lys Met His Ser
180          185          190
Pro Gln Thr Ser Ala Met Leu Phe Thr Val Asp Asn Glu Ala Gly Lys
195          200          205
Ile Thr Cys Leu Cys Gln Val Pro Gln Asn Ala Ala Asn Arg Gly Leu
210          215          220
Lys Ala Ser Glu Trp Val Gln Gln Val Ser Gly Leu Met Asp Gly Lys
225          230          235          240
Gly Gly Gly Lys Asp Val Ser Ala Gln Ala Thr Gly Lys Asn Val Gly
245          250          255
Cys Leu Gln Glu Ala Leu Gln Leu Ala Thr Ser Phe Ala Gln Leu Arg
260          265          270
Leu Gly Asp Val Lys Asn
275

```

<210> 141

<211> 837

<212> DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 141

```

cctgaccctg tgcgagtcgt ctccattggg gtcccgggtgt ccgagttgct ggatgacccc 60
tctgggacctg ctggctccct gacttctgtt gagttctgtg ggggaacgca cctgcggaac 120
tcgagtcatg caggagcttt tgtgatcgtg acggaagaag ccattgccaa gggatatccg 180
aggattgtgg ctgtcacagg tgccgaggcc cagaaggccc tcaggaaagc agagagcttg 240
aagaaatgic tctctgtcat ggaagccaaa gtgaaggctc agactgctcc aaacaaggat 300
gtgcagaggg agatcgctga ccttgagagag gccctggcca ctgcagtcac cccccagtg 360
cagaaggatg aattgcggga gactctcaaa tccctaaaga aggtcatgga tgacttggac 420
cgagccagca aagccgatgt ccagaaacga gtgttagaga agacgaagca gttcatcgac 480
agcaacccca accagcctct tgtcatcctg gagatggaga gcggcgccctc agccaaggcc 540
ctgaatgaag ccttgaagct cttcaagatg cactccccctc agacttctgc catgctcttc 600
acggtggaca atgaggctgg caagatcacg tgcctgtgtc aagttcccca gaatgcagcc 660
aatcggggct taaaagccag cgagtgggtg cagcaggtgt caggcttgat ggacggtaaa 720
ggtggtggca aggatgtgtc tgcacaggcc acaggcaaga acgttggtctg cctgcaggag 780
gcgctgcagc tggccacttc cttcgcccag ctgcgcctcg gggatgtaaa gaactga 837

```

&lt;210&gt; 142

&lt;211&gt; 222

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 142

```

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

```

&lt;210&gt; 143

&lt;211&gt; 669

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 143

```

aagggtatcc ggaggattgt ggctgtcaca ggtgccgagg cccagaaggc cctcaggaaa 60
gcagagagct tgaagaaatg tctctctgtc atggaagcca aagtgaaggc tcagactgct 120
ccaaacaagg atgtgcagag ggagatcgct gaccttggag aggccctggc cactgcagtc 180
atcccccagt ggcagaagga tgaattgcgg gagactctca aatccctaaa gaaggatcatg 240
gatgacttgg accgagccag caaagccgat gtccagaaac gagtgttaga gaagacgaag 300
cagttcatcg acagcaaccc caaccagcct cttgtcatcc tggagatgga gagcggcgcc 360
tcagccaagg ccctgaatga agccttgaag ctcttcaaga tgcactcccc tcagacttct 420
gccatgctct tcacggtgga caatgaggct ggcaagatca cgtgcctgtg tcaagttccc 480

```

120161\_426PC\_Sequence\_Listing.TXT

```
cagaatgcag ccaatcgggg cttaaaagcc agcgagtggg tgcagcaggt gtcaggcttg 540
atggacggta aaggtggtgg caaggatgtg tctgcacagg ccacaggcaa gaacgttggc 600
tgctgcagg aggcgtgca gctggccact tccttcgccc agctgcgcct cggggatgta 660
aagaactga
```

<210> 144  
 <211> 113  
 <212> PRT  
 <213> Homo sapiens

```
<400> 144
Pro Leu Val Ile Leu Glu Met Glu Ser Gly Ala Ser Ala Lys Ala Leu
1 5 10 15
Asn Glu Ala Leu Lys Leu Phe Lys Met His Ser Pro Gln Thr Ser Ala
20 25 30
Met Leu Phe Thr Val Asp Asn Glu Ala Gly Lys Ile Thr Cys Leu Cys
35 40 45
Gln Val Pro Gln Asn Ala Ala Asn Arg Gly Leu Lys Ala Ser Glu Trp
50 55 60
Val Gln Gln Val Ser Gly Leu Met Asp Gly Lys Gly Gly Lys Asp
65 70 75 80
Val Ser Ala Gln Ala Thr Gly Lys Asn Val Gly Cys Leu Gln Glu Ala
85 90 95
Leu Gln Leu Ala Thr Ser Phe Ala Gln Leu Arg Leu Gly Asp Val Lys
100 105 110
Asn
```

<210> 145  
 <211> 342  
 <212> DNA  
 <213> Homo sapiens

```
<400> 145
cctcttgtca tcctggagat ggagagcggc gcctcagcca aggccctgaa tgaagccttg 60
aagctcttca agatgcactc ccctcagact tctgccatgc tcttcacggt ggacaatgag 120
gtggcaaga tcacgtgcct gtgtcaagtt cccagaaatg cagccaatcg gggcttaaaa 180
gccagcgagt ggggtgcagca ggtgtcaggc ttgatggacg gtaaagggtg tggcaaggat 240
gtgtctgcac aggccacagg caagaacgtt ggctgcctgc aggaggcgct gcagctggcc 300
acttccttcg ccagctgcg cctcggggat gtaaagaact ga 342
```

<210> 146  
 <211> 514  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanyl tRNA Synthetase polypeptide with N-terminal  
 6xHis affinity tag

```
<400> 146
Met His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu
1 5 10 15
Gly Leu Asp Ser Thr Arg Gly Leu Glu Val Thr Asp Asp Ser Pro Lys
20 25 30
Tyr Asn Tyr His Leu Asp Ser Ser Gly Ser Tyr Val Phe Glu Asn Thr
35 40 45
Val Ala Thr Val Met Ala Leu Arg Arg Glu Lys Met Phe Val Glu Glu
50 55 60
Val Ser Thr Gly Gln Glu Cys Gly Val Val Leu Asp Lys Thr Cys Phe
65 70 75 80
Tyr Ala Glu Gln Gly Gly Gln Ile Tyr Asp Glu Gly Tyr Leu Val Lys
85 90 95
Val Asp Asp Ser Ser Glu Asp Lys Thr Glu Phe Thr Val Lys Asn Ala
100 105 110
Gln Val Arg Gly Gly Tyr Val Leu His Ile Gly Thr Ile Tyr Gly Asp
115 120 125
```

120161\_426PC\_Sequence\_Listing.TXT

```

Leu 130 Lys Val Gly Asp Gln Val Trp Leu Phe Ile Asp Glu Pro Arg Arg
Arg 145 Pro Ile Met Ser Asn 135 His Thr Ala Thr His 140 Ile Leu Asn Phe Ala
Leu Arg Ser Val Leu 150 Gly Glu Ala Asp Gln Lys Gly Ser Leu Val Ala
Pro Asp Arg Leu 165 Arg Phe Asp Phe Thr 170 Ala Lys Gly Ala Met Ser Thr
Gln Gln Ile 180 Lys Lys Ala Glu Glu 185 Ile Ala Asn Glu Met 190 Ile Glu Ala
Ala Lys Ala Val Tyr Thr Gln 200 Asp Cys Pro Leu Ala 205 Ala Ala Lys Ala
Ile 210 Gln Gly Leu Arg Ala 215 Val Phe Asp Glu Thr Tyr Pro Asp Pro Val
Arg Val Val Ser Ile 225 Gly Val Pro Val Ser Glu Leu Leu Asp Asp Pro
Ser Gly Pro Ala 230 Gly Ser Leu Thr Ser Val Glu Phe Cys Gly Gly Thr
His Leu Arg Asn Ser Ser His 245 Ala Gly Ala Phe Val 250 Ile Val Thr Glu
Glu Ala Ile Ala Lys Gly Ile 260 Arg Arg Ile Val Ala 265 Val Thr Gly Ala
Glu Ala Gln Lys Ala Leu 275 Arg Lys Ala Glu Ser Leu Lys Lys Cys Leu
Ser Val Met Glu Ala 280 Lys Val Lys Ala Gln Thr Ala Pro Asn Lys Asp
Val Gln Arg Glu Ile Ala Asp Leu Gly Glu Ala Leu Ala Thr Ala Val
Ile Pro Gln Trp Gln Lys Asp Glu 295 Leu Arg Glu Thr Leu Lys Ser Leu
Lys Lys Val Met Asp Asp Leu 300 Asp Arg Ala Ser Lys Ala Asp Val Gln
Lys Arg Val Leu Glu Lys Thr Lys Gln Phe Ile Asp Ser Asn Pro Asn
Gln Pro Leu Val Ile Leu Glu Met Glu Ser Gly Ala Ser Ala Lys Ala
Leu Asn Glu Ala Leu Lys Leu Phe Lys Met His Ser Pro Gln Thr Ser
Ala Met Leu Phe Thr Val Asp Asn Glu Ala Gly Lys Ile Thr Cys Leu
Cys Gln Val Pro Gln Asn Ala 315 Ala Asn Arg Gly Leu Lys Ala Ser Glu
Trp Val Gln Gln Val Ser Gly Leu Met Asp Gly Lys Gly Gly Gly Lys
Asp Val Ser Ala Gln Ala Thr Gly Lys Asn Val Gly Cys Leu Gln Glu
Ala Leu Gln Leu Ala Thr Ser Phe Ala Gln Leu Arg Leu Gly Asp Val
Lys Asn

```

<210> 147

<211> 514

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanine tRNA Synthetase polypeptide with C-terminal  
6xHis affinity tag

<400> 147

```

Met Arg Gly Leu Glu Val Thr Asp Asp Ser Pro Lys Tyr Asn Tyr His
1 5 10
Leu Asp Ser Ser Gly Ser Tyr Val Phe Glu Asn Thr Val Ala Thr Val
20 25 30
Met Ala Leu Arg Arg Glu Lys Met Phe Val Glu Glu Val Ser Thr Gly
35 40 45

```



120161\_426PC\_Sequence\_Listing.TXT

Gln	Glu	Cys	Gly	Val	Val	Leu	Asp	Lys	Thr	Cys	Phe	Tyr	Ala	Glu	Gln
50	50					55				60					
Gly	Gly	Gln	Ile	Tyr	Asp	Glu	Gly	Tyr	Leu	Val	Lys	Val	Asp	Asp	Ser
65					70					75					80
Ser	Glu	Asp	Lys	Thr	Glu	Phe	Thr	Val	Lys	Asn	Ala	Gln	Val	Arg	Gly
				85					90					95	
Gly	Tyr	Val	Leu	His	Ile	Gly	Thr	Ile	Tyr	Gly	Asp	Leu	Lys	Val	Gly
			100					105					110		
Asp	Gln	Val	Trp	Leu	Phe	Ile	Asp	Glu	Pro	Arg	Arg	Arg	Pro	Ile	Met
		115					120					125			
Ser	Asn	His	Thr	Ala	Thr	His	Ile	Leu	Asn	Phe	Ala	Leu	Arg	Ser	Val
	130					135				140					
Leu	Gly	Glu	Ala	Asp	Gln	Lys	Gly	Ser	Leu	Val	Ala	Pro	Asp	Arg	Leu
145					150					155					160
Arg	Phe	Asp	Phe	Thr	Ala	Lys	Gly	Ala	Met	Ser	Thr	Gln	Gln	Ile	Lys
				165					170					175	
Lys	Ala	Glu	Glu	Ile	Ala	Asn	Glu	Met	Ile	Glu	Ala	Ala	Lys	Ala	Val
			180					185					190		
Tyr	Thr	Gln	Asp	Cys	Pro	Leu	Ala	Ala	Ala	Lys	Ala	Ile	Gln	Gly	Leu
		195					200					205			
Arg	Ala	Val	Phe	Asp	Glu	Thr	Tyr	Pro	Asp	Pro	Val	Arg	Val	Val	Ser
	210					215				220					
Ile	Gly	Val	Pro	Val	Ser	Glu	Leu	Leu	Asp	Asp	Pro	Ser	Gly	Pro	Ala
225					230					235					240
Gly	Ser	Leu	Thr	Ser	Val	Glu	Phe	Cys	Gly	Gly	Thr	His	Leu	Arg	Asn
				245					250				255		
Ser	Ser	His	Ala	Gly	Ala	Phe	Val	Ile	Val	Thr	Glu	Glu	Ala	Ile	Ala
			260					265					270		
Lys	Gly	Ile	Arg	Arg	Ile	Val	Ala	Val	Thr	Gly	Ala	Glu	Ala	Gln	Lys
		275					280					285			
Ala	Leu	Arg	Lys	Ala	Glu	Ser	Leu	Lys	Lys	Cys	Leu	Ser	Val	Met	Glu
	290					295					300				
Ala	Lys	Val	Lys	Ala	Gln	Thr	Ala	Pro	Asn	Lys	Asp	Val	Gln	Arg	Glu
305					310					315					320
Ile	Ala	Asp	Leu	Gly	Glu	Ala	Leu	Ala	Thr	Ala	Val	Ile	Pro	Gln	Trp
				325					330					335	
Gln	Lys	Asp	Glu	Leu	Arg	Glu	Thr	Leu	Lys	Ser	Leu	Lys	Lys	Val	Met
			340					345					350		
Asp	Asp	Leu	Asp	Arg	Ala	Ser	Lys	Ala	Asp	Val	Gln	Lys	Arg	Val	Leu
		355					360					365			
Glu	Lys	Thr	Lys	Gln	Phe	Ile	Asp	Ser	Asn	Pro	Asn	Gln	Pro	Leu	Val
	370					375					380				
Ile	Leu	Glu	Met	Glu	Ser	Gly	Ala	Ser	Ala	Lys	Ala	Leu	Asn	Glu	Ala
385					390					395					400
Leu	Lys	Leu	Phe	Lys	Met	His	Ser	Pro	Gln	Thr	Ser	Ala	Met	Leu	Phe
				405					410					415	
Thr	Val	Asp	Asn	Glu	Ala	Gly	Lys	Ile	Thr	Cys	Leu	Cys	Gln	Val	Pro
			420					425					430		
Gln	Asn	Ala	Ala	Asn	Arg	Gly	Leu	Lys	Ala	Ser	Glu	Trp	Val	Gln	Gln
	435						440					445			
Val	Ser	Gly	Leu	Met	Asp	Gly	Lys	Gly	Gly	Gly	Lys	Asp	Val	Ser	Ala
	450					455					460				
Gln	Ala	Thr	Gly	Lys	Asn	Val	Gly	Cys	Leu	Gln	Glu	Ala	Leu	Gln	Leu
465					470					475					480
Ala	Thr	Ser	Phe	Ala	Gln	Leu	Arg	Leu	Gly	Asp	Val	Lys	Asn	Gly	Lys
				485					490					495	
Pro	Ile	Pro	Asn	Pro	Leu	Leu	Gly	Leu	Asp	Ser	Thr	His	His	His	His
			500					505					510		
His	His														

<210> 148

<211> 236

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanyl tRNA Synthetase polypeptide with N-terminal  
6xHis affinity tag

<400> 148

```

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu
 1      5      10      15
Gly Leu Asp Ser Thr Gly Ser Ala Glu Ala Gln Lys Ala Leu Arg Lys
 20      25      30
Ala Glu Ser Leu Lys Lys Cys Leu Ser Val Met Glu Ala Lys Val Lys
 35      40      45
Ala Gln Thr Ala Pro Asn Lys Asp Val Gln Arg Glu Ile Ala Asp Leu
 50      55      60
Gly Glu Ala Leu Ala Thr Ala Val Ile Pro Gln Trp Gln Lys Asp Glu
 65      70      75      80
Leu Arg Glu Thr Leu Lys Ser Leu Lys Lys Val Met Asp Asp Leu Asp
 85      90      95
Arg Ala Ser Lys Ala Asp Val Gln Lys Arg Val Leu Glu Lys Thr Lys
100      105      110
Gln Phe Ile Asp Ser Asn Pro Asn Gln Pro Leu Val Ile Leu Glu Met
115      120      125
Glu Ser Gly Ala Ser Ala Lys Ala Leu Asn Glu Ala Leu Lys Leu Phe
130      135      140
Lys Met His Ser Pro Gln Thr Ser Ala Met Leu Phe Thr Val Asp Asn
145      150      155      160
Glu Ala Gly Lys Ile Thr Cys Leu Cys Gln Val Pro Gln Asn Ala Ala
165      170      175
Asn Arg Gly Leu Lys Ala Ser Glu Trp Val Gln Gln Val Ser Gly Leu
180      185      190
Met Asp Gly Lys Gly Gly Gly Lys Asp Val Ser Ala Gln Ala Thr Gly
195      200      205
Lys Asn Val Gly Cys Leu Gln Glu Ala Leu Gln Leu Ala Thr Ser Phe
210      215      220
Ala Gln Leu Arg Leu Gly Asp Val Lys Asn Leu Glu
225      230      235

```

<210> 149

<211> 236

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanyl tRNA Synthetase polypeptide with C-terminal  
6xHis affinity tag

<400> 149

```

Met Gly Ser Ala Glu Ala Gln Lys Ala Leu Arg Lys Ala Glu Ser Leu
 1      5      10      15
Lys Lys Cys Leu Ser Val Met Glu Ala Lys Val Lys Ala Gln Thr Ala
 20      25      30
Pro Asn Lys Asp Val Gln Arg Glu Ile Ala Asp Leu Gly Glu Ala Leu
 35      40      45
Ala Thr Ala Val Ile Pro Gln Trp Gln Lys Asp Glu Leu Arg Glu Thr
 50      55      60
Leu Lys Ser Leu Lys Lys Val Met Asp Asp Leu Asp Arg Ala Ser Lys
 65      70      75      80
Ala Asp Val Gln Lys Arg Val Leu Glu Lys Thr Lys Gln Phe Ile Asp
 85      90      95
Ser Asn Pro Asn Gln Pro Leu Val Ile Leu Glu Met Glu Ser Gly Ala
100      105      110
Ser Ala Lys Ala Leu Asn Glu Ala Leu Lys Leu Phe Lys Met His Ser
115      120      125
Pro Gln Thr Ser Ala Met Leu Phe Thr Val Asp Asn Glu Ala Gly Lys
130      135      140
Ile Thr Cys Leu Cys Gln Val Pro Gln Asn Ala Asn Arg Gly Leu
145      150      155      160
Lys Ala Ser Glu Trp Val Gln Gln Val Ser Gly Leu Met Asp Gly Lys
165      170      175

```

120161\_426PC\_Sequence\_Listing.TXT

Gly	Gly	Gly	Lys	Asp	Val	Ser	Ala	Gln	Ala	Thr	Gly	Lys	Asn	Val	Gly
			180					185					190		
Cys	Leu	Gln	Glu	Ala	Leu	Gln	Leu	Ala	Thr	Ser	Phe	Ala	Gln	Leu	Arg
		195					200					205			
Leu	Gly	Asp	Val	Lys	Asn	Leu	Glu	Gly	Lys	Pro	Ile	Pro	Asn	Pro	Leu
	210					215					220				
Leu	Gly	Leu	Asp	Ser	Thr	His	His	His	His	His	His				
225					230					235					

<210> 150

<211> 402

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanine tRNA Synthetase polypeptide with N-terminal 6xHis affinity tag

<400> 150

Met	His	His	His	His	His	His	Gly	Lys	Pro	Ile	Pro	Asn	Pro	Leu	Leu
1				5				10						15	
Gly	Leu	Asp	Ser	Thr	Gln	Val	Trp	Leu	Phe	Ile	Asp	Glu	Pro	Arg	Arg
			20					25					30		
Arg	Pro	Ile	Met	Ser	Asn	His	Thr	Ala	Thr	His	Ile	Leu	Asn	Phe	Ala
		35					40					45			
Leu	Arg	Ser	Val	Leu	Gly	Glu	Ala	Asp	Gln	Lys	Gly	Ser	Leu	Val	Ala
	50				55						60				
Pro	Asp	Arg	Leu	Arg	Phe	Asp	Phe	Thr	Ala	Lys	Gly	Ala	Met	Ser	Thr
65					70					75				80	
Gln	Gln	Ile	Lys	Lys	Ala	Glu	Glu	Ile	Ala	Asn	Glu	Met	Ile	Glu	Ala
				85					90					95	
Ala	Lys	Ala	Val	Tyr	Thr	Gln	Asp	Cys	Pro	Leu	Ala	Ala	Ala	Lys	Ala
			100					105					110		
Ile	Gln	Gly	Leu	Arg	Ala	Val	Phe	Asp	Glu	Thr	Tyr	Pro	Asp	Pro	Val
		115					120					125			
Arg	Val	Val	Ser	Ile	Gly	Val	Pro	Val	Ser	Glu	Leu	Leu	Asp	Asp	Pro
	130					135					140				
Ser	Gly	Pro	Ala	Gly	Ser	Leu	Thr	Ser	Val	Glu	Phe	Cys	Gly	Gly	Thr
145					150					155				160	
His	Leu	Arg	Asn	Ser	Ser	His	Ala	Gly	Ala	Phe	Val	Ile	Val	Thr	Glu
				165					170					175	
Glu	Ala	Ile	Ala	Lys	Gly	Ile	Arg	Arg	Ile	Val	Ala	Val	Thr	Gly	Ala
			180					185					190		
Glu	Ala	Gln	Lys	Ala	Leu	Arg	Lys	Ala	Glu	Ser	Leu	Lys	Lys	Cys	Leu
		195					200					205			
Ser	Val	Met	Glu	Ala	Lys	Val	Lys	Ala	Gln	Thr	Ala	Pro	Asn	Lys	Asp
	210					215					220				
Val	Gln	Arg	Glu	Ile	Ala	Asp	Leu	Gly	Glu	Ala	Leu	Ala	Thr	Ala	Val
225					230					235					240
Ile	Pro	Gln	Trp	Gln	Lys	Asp	Glu	Leu	Arg	Glu	Thr	Leu	Lys	Ser	Leu
				245					250					255	
Lys	Lys	Val	Met	Asp	Asp	Leu	Asp	Arg	Ala	Ser	Lys	Ala	Asp	Val	Gln
			260					265					270		
Lys	Arg	Val	Leu	Glu	Lys	Thr	Lys	Gln	Phe	Ile	Asp	Ser	Asn	Pro	Asn
		275					280					285			
Gln	Pro	Leu	Val	Ile	Leu	Glu	Met	Glu	Ser	Gly	Ala	Ser	Ala	Lys	Ala
	290					295					300				
Leu	Asn	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Met	His	Ser	Pro	Gln	Thr	Ser
305					310					315				320	
Ala	Met	Leu	Phe	Thr	Val	Asp	Asn	Glu	Ala	Gly	Lys	Ile	Thr	Cys	Leu
				325					330					335	
Cys	Gln	Val	Pro	Gln	Asn	Ala	Ala	Asn	Arg	Gly	Leu	Lys	Ala	Ser	Glu
			340					345					350		
Trp	Val	Gln	Gln	Val	Ser	Gly	Leu	Met	Asp	Gly	Lys	Gly	Gly	Gly	Lys
		355					360					365			
Asp	Val	Ser	Ala	Gln	Ala	Thr	Gly	Lys	Asn	Val	Gly	Cys	Leu	Gln	Glu
370						375					380				

120161\_426PC\_Sequence\_Listing.TXT

Ala Leu Gln Leu Ala Thr Ser Phe Ala Gln Leu Arg Leu Gly Asp Val  
 385 390 395 400  
 Lys Asn

<210> 151  
 <211> 402  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanine tRNA Synthetase polypeptide with C-terminal  
 6xHis affinity tag

<400> 151  
 Met Gln Val Trp Leu Phe Ile Asp Glu Pro Arg Arg Arg Pro Ile Met  
 1 5 10 15  
 Ser Asn His Thr Ala Thr His Ile Leu Asn Phe Ala Leu Arg Ser Val  
 20 25 30  
 Leu Gly Glu Ala Asp Gln Lys Gly Ser Leu Val Ala Pro Asp Arg Leu  
 35 40 45  
 Arg Phe Asp Phe Thr Ala Lys Gly Ala Met Ser Thr Gln Gln Ile Lys  
 50 55 60  
 Lys Ala Glu Glu Ile Ala Asn Glu Met Ile Glu Ala Ala Lys Ala Val  
 65 70 75 80  
 Tyr Thr Gln Asp Cys Pro Leu Ala Ala Lys Ala Ile Gln Gly Leu  
 85 90 95  
 Arg Ala Val Phe Asp Glu Thr Tyr Pro Asp Pro Val Arg Val Val Ser  
 100 105 110  
 Ile Gly Val Pro Val Ser Glu Leu Leu Asp Asp Pro Ser Gly Pro Ala  
 115 120 125  
 Gly Ser Leu Thr Ser Val Glu Phe Cys Gly Gly Thr His Leu Arg Asn  
 130 135 140  
 Ser Ser His Ala Gly Ala Phe Val Ile Val Thr Glu Glu Ala Ile Ala  
 145 150 155 160  
 Lys Gly Ile Arg Arg Ile Val Ala Val Thr Gly Ala Glu Ala Gln Lys  
 165 170 175  
 Ala Leu Arg Lys Ala Glu Ser Leu Lys Lys Cys Leu Ser Val Met Glu  
 180 185 190  
 Ala Lys Val Lys Ala Gln Thr Ala Pro Asn Lys Asp Val Gln Arg Glu  
 195 200 205  
 Ile Ala Asp Leu Gly Glu Ala Leu Ala Thr Ala Val Ile Pro Gln Trp  
 210 215 220  
 Gln Lys Asp Glu Leu Arg Glu Thr Leu Lys Ser Leu Lys Lys Val Met  
 225 230 235 240  
 Asp Asp Leu Asp Arg Ala Ser Lys Ala Asp Val Gln Lys Arg Val Leu  
 245 250 255  
 Glu Lys Thr Lys Gln Phe Ile Asp Ser Asn Pro Asn Gln Pro Leu Val  
 260 265 270  
 Ile Leu Glu Met Glu Ser Gly Ala Ser Ala Lys Ala Leu Asn Glu Ala  
 275 280 285  
 Leu Lys Leu Phe Lys Met His Ser Pro Gln Thr Ser Ala Met Leu Phe  
 290 295 300  
 Thr Val Asp Asn Glu Ala Gly Lys Ile Thr Cys Leu Cys Gln Val Pro  
 305 310 315 320  
 Gln Asn Ala Ala Asn Arg Gly Leu Lys Ala Ser Glu Trp Val Gln Gln  
 325 330 335  
 Val Ser Gly Leu Met Asp Gly Lys Gly Gly Lys Asp Val Ser Ala  
 340 345 350  
 Gln Ala Thr Gly Lys Asn Val Gly Cys Leu Gln Glu Ala Leu Gln Leu  
 355 360 365  
 Ala Thr Ser Phe Ala Gln Leu Arg Leu Gly Asp Val Lys Asn Gly Lys  
 370 375 380  
 Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr His His His His  
 385 390 395 400  
 His His

<210> 152  
 <211> 330  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanyl tRNA Synthetase polypeptide with N-terminal  
 6xHis affinity tag (DNA2.0)

<400> 152  
 Met His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15  
 Gly Leu Asp Ser Thr Ile Glu Ala Ala Lys Ala Val Tyr Thr Gln Asp  
 20 25 30  
 Cys Pro Leu Ala Ala Ala Lys Ala Ile Gln Gly Leu Arg Ala Val Phe  
 35 40 45  
 Asp Glu Thr Tyr Pro Asp Pro Val Arg Val Val Ser Ile Gly Val Pro  
 50 55 60  
 Val Ser Glu Leu Leu Asp Asp Pro Ser Gly Pro Ala Gly Ser Leu Thr  
 65 70 75 80  
 Ser Val Glu Phe Cys Gly Gly Thr His Leu Arg Asn Ser Ser His Ala  
 85 90 95  
 Gly Ala Phe Val Ile Val Thr Glu Glu Ala Ile Ala Lys Gly Ile Arg  
 100 105 110  
 Arg Ile Val Ala Val Thr Gly Ala Glu Ala Gln Lys Ala Leu Arg Lys  
 115 120 125  
 Ala Glu Ser Leu Lys Lys Cys Leu Ser Val Met Glu Ala Lys Val Lys  
 130 135 140  
 Ala Gln Thr Ala Pro Asn Lys Asp Val Gln Arg Glu Ile Ala Asp Leu  
 145 150 155 160  
 Gly Glu Ala Leu Ala Thr Ala Val Ile Pro Gln Trp Gln Lys Asp Glu  
 165 170 175  
 Leu Arg Glu Thr Leu Lys Ser Leu Lys Lys Val Met Asp Asp Leu Asp  
 180 185 190  
 Arg Ala Ser Lys Ala Asp Val Gln Lys Arg Val Leu Glu Lys Thr Lys  
 195 200 205  
 Gln Phe Ile Asp Ser Asn Pro Asn Gln Pro Leu Val Ile Leu Glu Met  
 210 215 220  
 Glu Ser Gly Ala Ser Ala Lys Ala Leu Asn Glu Ala Leu Lys Leu Phe  
 225 230 235 240  
 Lys Met His Ser Pro Gln Thr Ser Ala Met Leu Phe Thr Val Asp Asn  
 245 250 255  
 Glu Ala Gly Lys Ile Thr Cys Leu Cys Gln Val Pro Gln Asn Ala Ala  
 260 265 270  
 Asn Arg Gly Leu Lys Ala Ser Glu Trp Val Gln Gln Val Ser Gly Leu  
 275 280 285  
 Met Asp Gly Lys Gly Gly Gly Lys Asp Val Ser Ala Gln Ala Thr Gly  
 290 295 300  
 Lys Asn Val Gly Cys Leu Gln Glu Ala Leu Gln Leu Ala Thr Ser Phe  
 305 310 315 320  
 Ala Gln Leu Arg Leu Gly Asp Val Lys Asn  
 325 330

<210> 153  
 <211> 330  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanyl tRNA Synthetase polypeptide with C-terminal  
 6xHis affinity tag (DNA2.0)

<400> 153  
 Met Ile Glu Ala Ala Lys Ala Val Tyr Thr Gln Asp Cys Pro Leu Ala  
 1 5 10 15

120161\_426PC\_Sequence\_Listing.TXT

Ala	Ala	Lys	Ala	Ile	Gln	Gly	Leu	Arg	Ala	Val	Phe	Asp	Glu	Thr	Tyr
Pro	Asp	Pro	Val	Arg	Val	Val	Ser	Ile	Gly	Val	Pro	Val	Ser	Glu	Leu
Leu	Asp	Asp	Pro	Ser	Gly	Pro	Ala	Gly	Ser	Leu	Thr	Ser	Val	Glu	Phe
Cys	Gly	Gly	Thr	His	Leu	Arg	Asn	Ser	Ser	His	Ala	Gly	Ala	Phe	Val
Ile	Val	Thr	Glu	Glu	Ala	Ile	Ala	Lys	Gly	Ile	Arg	Arg	Ile	Val	Ala
Val	Thr	Gly	Ala	Glu	Ala	Gln	Lys	Ala	Leu	Arg	Lys	Ala	Glu	Ser	Leu
Lys	Lys	Cys	Leu	Ser	Val	Met	Glu	Ala	Lys	Val	Lys	Ala	Gln	Thr	Ala
Pro	Asn	Lys	Asp	Val	Gln	Arg	Glu	Ile	Ala	Asp	Leu	Gly	Glu	Ala	Leu
Ala	Thr	Ala	Val	Ile	Pro	Gln	Trp	Gln	Lys	Asp	Glu	Leu	Arg	Glu	Thr
Leu	Lys	Ser	Leu	Lys	Lys	Val	Met	Asp	Asp	Leu	Asp	Arg	Ala	Ser	Lys
Ala	Asp	Val	Gln	Lys	Arg	Val	Leu	Glu	Lys	Thr	Lys	Gln	Phe	Ile	Asp
Ser	Asn	Pro	Asn	Gln	Pro	Leu	Val	Ile	Leu	Glu	Met	Glu	Ser	Gly	Ala
Ser	Ala	Lys	Ala	Leu	Asn	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Met	His	Ser
Pro	Gln	Thr	Ser	Ala	Met	Leu	Phe	Thr	Val	Asp	Asn	Glu	Ala	Gly	Lys
Ile	Thr	Cys	Leu	Cys	Gln	Val	Pro	Gln	Asn	Ala	Ala	Asn	Arg	Gly	Leu
Lys	Ala	Ser	Glu	Trp	Val	Gln	Gln	Val	Ser	Gly	Leu	Met	Asp	Gly	Lys
Gly	Gly	Gly	Lys	Asp	Val	Ser	Ala	Gln	Ala	Thr	Gly	Lys	Asn	Val	Gly
Cys	Leu	Gln	Glu	Ala	Leu	Gln	Leu	Ala	Thr	Ser	Phe	Ala	Gln	Leu	Arg
Leu	Gly	Asp	Val	Lys	Asn	Gly	Lys	Pro	Ile	Pro	Asn	Pro	Leu	Leu	Gly
Leu	Asp	Ser	Thr	His	His	His	His	His	His	His	His	His	His	His	His

<210> 154

<211> 299

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanine tRNA Synthetase polypeptide with N-terminal 6xHis affinity tag

<400> 154

Met	His	His	His	His	His	His	Gly	Lys	Pro	Ile	Pro	Asn	Pro	Leu	Leu
Gly	Leu	Asp	Ser	Thr	Pro	Asp	Pro	Val	Arg	Val	Val	Ser	Ile	Gly	Val
Pro	Val	Ser	Glu	Leu	Leu	Asp	Asp	Pro	Ser	Gly	Pro	Ala	Gly	Ser	Leu
Thr	Ser	Val	Glu	Phe	Cys	Gly	Gly	Thr	His	Leu	Arg	Asn	Ser	Ser	His
Ala	Gly	Ala	Phe	Val	Ile	Val	Thr	Glu	Glu	Ala	Ile	Ala	Lys	Gly	Ile
Arg	Arg	Ile	Val	Ala	Val	Thr	Gly	Ala	Glu	Ala	Gln	Lys	Ala	Leu	Arg
Lys	Ala	Glu	Ser	Leu	Lys	Lys	Cys	Leu	Ser	Val	Met	Glu	Ala	Lys	Val
Lys	Ala	Gln	Thr	Ala	Pro	Asn	Lys	Asp	Val	Gln	Arg	Glu	Ile	Ala	Asp

120161\_426PC\_Sequence\_Listing.TXT

```

Leu Gly Glu Ala Leu Ala Thr Ala Val Ile Pro Gln Trp Gln Lys Asp
130 135 140
Glu Leu Arg Glu Thr Leu Lys Ser Leu Lys Lys Val Met Asp Asp Leu
145 150 155 160
Asp Arg Ala Ser Lys Ala Asp Val Gln Lys Arg Val Leu Glu Lys Thr
165 170 175
Lys Gln Phe Ile Asp Ser Asn Pro Asn Gln Pro Leu Val Ile Leu Glu
180 185 190
Met Glu Ser Gly Ala Ser Ala Lys Ala Leu Asn Glu Ala Leu Lys Leu
195 200 205
Phe Lys Met His Ser Pro Gln Thr Ser Ala Met Leu Phe Thr Val Asp
210 215 220
Asn Glu Ala Gly Lys Ile Thr Cys Leu Cys Gln Val Pro Gln Asn Ala
225 230 235 240
Ala Asn Arg Gly Leu Lys Ala Ser Glu Trp Val Gln Gln Val Ser Gly
245 250 255
Leu Met Asp Gly Lys Gly Gly Gly Lys Asp Val Ser Ala Gln Ala Thr
260 265 270
Gly Lys Asn Val Gly Cys Leu Gln Glu Ala Leu Gln Leu Ala Thr Ser
275 280 285
Phe Ala Gln Leu Arg Leu Gly Asp Val Lys Asn
290 295

```

<210> 155

<211> 299

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanyl tRNA Synthetase polypeptide with C-terminal  
6xHis affinity tag

<400> 155

```

Met Pro Asp Pro Val Arg Val Val Ser Ile Gly Val Pro Val Ser Glu
1 5 10 15
Leu Leu Asp Asp Pro Ser Gly Pro Ala Gly Ser Leu Thr Ser Val Glu
20 25 30
Phe Cys Gly Gly Thr His Leu Arg Asn Ser Ser His Ala Gly Ala Phe
35 40 45
Val Ile Val Thr Glu Glu Ala Ile Ala Lys Gly Ile Arg Arg Ile Val
50 55 60
Ala Val Thr Gly Ala Glu Ala Gln Lys Ala Leu Arg Lys Ala Glu Ser
65 70 75 80
Leu Lys Lys Cys Leu Ser Val Met Glu Ala Lys Val Lys Ala Gln Thr
85 90 95
Ala Pro Asn Lys Asp Val Gln Arg Glu Ile Ala Asp Leu Gly Glu Ala
100 105 110
Leu Ala Thr Ala Val Ile Pro Gln Trp Gln Lys Asp Glu Leu Arg Glu
115 120 125
Thr Leu Lys Ser Leu Lys Lys Val Met Asp Asp Leu Asp Arg Ala Ser
130 135 140
Lys Ala Asp Val Gln Lys Arg Val Leu Glu Lys Thr Lys Gln Phe Ile
145 150 155 160
Asp Ser Asn Pro Asn Gln Pro Leu Val Ile Leu Glu Met Glu Ser Gly
165 170 175
Ala Ser Ala Lys Ala Leu Asn Glu Ala Leu Lys Leu Phe Lys Met His
180 185 190
Ser Pro Gln Thr Ser Ala Met Leu Phe Thr Val Asp Asn Glu Ala Gly
195 200 205
Lys Ile Thr Cys Leu Cys Gln Val Pro Gln Asn Ala Ala Asn Arg Gly
210 215 220
Leu Lys Ala Ser Glu Trp Val Gln Gln Val Ser Gly Leu Met Asp Gly
225 230 235 240
Lys Gly Gly Gly Lys Asp Val Ser Ala Gln Ala Thr Gly Lys Asn Val
245 250 255
Gly Cys Leu Gln Glu Ala Leu Gln Leu Ala Thr Ser Phe Ala Gln Leu
260 265 270

```

120161\_426PC\_Sequence\_Listing.TXT

Arg Leu Gly Asp Val Lys Asn Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 275 280 285  
 Gly Leu Asp Ser Thr His His His His His  
 290 295

<210> 156  
 <211> 247  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanyl tRNA Synthetase polypeptide with N-terminal  
 6xHis affinity tag

<400> 156  
 Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15  
 Gly Leu Asp Ser Thr Gly Ser Lys Gly Ile Arg Arg Ile Val Ala Val  
 20 25 30  
 Thr Gly Ala Glu Ala Gln Lys Ala Leu Arg Lys Ala Glu Ser Leu Lys  
 35 40 45  
 Lys Cys Leu Ser Val Met Glu Ala Lys Val Lys Ala Gln Thr Ala Pro  
 50 55 60  
 Asn Lys Asp Val Gln Arg Glu Ile Ala Asp Leu Gly Glu Ala Leu Ala  
 65 70 75 80  
 Thr Ala Val Ile Pro Gln Trp Gln Lys Asp Glu Leu Arg Glu Thr Leu  
 85 90 95  
 Lys Ser Leu Lys Lys Val Met Asp Asp Leu Asp Arg Ala Ser Lys Ala  
 100 105 110  
 Asp Val Gln Lys Arg Val Leu Glu Lys Thr Lys Gln Phe Ile Asp Ser  
 115 120 125  
 Asn Pro Asn Gln Pro Leu Val Ile Leu Glu Met Glu Ser Gly Ala Ser  
 130 135 140  
 Ala Lys Ala Leu Asn Glu Ala Leu Lys Leu Phe Lys Met His Ser Pro  
 145 150 155 160  
 Gln Thr Ser Ala Met Leu Phe Thr Val Asp Asn Glu Ala Gly Lys Ile  
 165 170 175  
 Thr Cys Leu Cys Gln Val Pro Gln Asn Ala Ala Asn Arg Gly Leu Lys  
 180 185 190  
 Ala Ser Glu Trp Val Gln Gln Val Ser Gly Leu Met Asp Gly Lys Gly  
 195 200 205  
 Gly Gly Lys Asp Val Ser Ala Gln Ala Thr Gly Lys Asn Val Gly Cys  
 210 215 220  
 Leu Gln Glu Ala Leu Gln Leu Ala Thr Ser Phe Ala Gln Leu Arg Leu  
 225 230 235 240  
 Gly Asp Val Lys Asn Leu Glu  
 245

<210> 157  
 <211> 247  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanyl tRNA Synthetase polypeptide with C-terminal  
 6xHis affinity tag

<400> 157  
 Met Gly Ser Lys Gly Ile Arg Arg Ile Val Ala Val Thr Gly Ala Glu  
 1 5 10 15  
 Ala Gln Lys Ala Leu Arg Lys Ala Glu Ser Leu Lys Lys Cys Leu Ser  
 20 25 30  
 Val Met Glu Ala Lys Val Lys Ala Gln Thr Ala Pro Asn Lys Asp Val  
 35 40 45  
 Gln Arg Glu Ile Ala Asp Leu Gly Glu Ala Leu Ala Thr Ala Val Ile  
 50 55 60



120161\_426PC\_Sequence\_Listing.TXT

```

Pro 65 Gln Trp Gln Lys Asp 70 Glu Leu Arg Glu Thr 75 Leu Lys Ser Leu Lys 80
Lys Val Met Asp 85 Asp 85 Leu Asp Arg Ala Ser 90 Lys Ala Asp Val Gln Lys 95
Arg Val Leu Glu 100 Lys Thr Lys Gln Phe 105 Ile Asp Ser Asn Pro Asn Gln 110
Pro Leu Val 115 Ile Leu Glu Met Glu 120 Ser Gly Ala Ser Ala Lys Ala Leu 125
Asn Glu 130 Ala Leu Lys Leu Phe 135 Lys Met His Ser Pro Gln Thr Ser Ala 140
Met 145 Leu Phe Thr Val Asp 150 Asn Glu Ala Gly Lys 155 Ile Thr Cys Leu Cys 160
Gln Val Pro Gln Asn 165 Ala Ala Asn Arg Gly 170 Leu Lys Ala Ser Glu Trp 175
Val Gln Gln Val 180 Ser Gly Leu Met Asp 185 Gly Lys Gly Gly Lys Asp 190
Val Ser Ala 195 Gln Ala Thr Gly Lys 200 Asn Val Gly Cys Leu Gln Glu Ala 205
Leu Gln 210 Leu Ala Thr Ser Phe 215 Ala Gln Leu Arg Leu Gly Asp Val Lys 220
Asn 225 Leu Glu Gly Lys Pro 230 Ile Pro Asn Pro Leu 235 Leu Gly Leu Asp Ser 240
Thr His His His His 245 His His

```

<210> 158  
 <211> 134  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanine tRNA Synthetase polypeptide with N-terminal  
 6xHis affinity tag

```

<400> 158
Met 1 His His His 5 His His Gly Lys 10 Pro Ile Pro Asn Pro Leu Leu 15
Gly Leu Asp Ser Thr Pro Leu Val 20 Ile Leu Glu Met Glu Ser Gly Ala 25
Ser Ala Lys 30 Ala Leu Asn Glu Ala Leu Lys Leu Phe Lys Met His Ser 35
Pro Gln Thr Ser Ala Met Leu 40 Phe Thr Val Asp Asn Glu Ala Gly Lys 45
Ile Thr Cys Leu Cys Gln Val Pro Gln Asn Ala Ala Asn Arg Gly Leu 50
Lys Ala Ser Glu Trp Val Gln Gln Val Ser 55 Gly Leu Met Asp Gly Lys 60
Gly Gly Gly Lys 65 Asp Val Ser Ala Gln 70 Ala Thr Gly Lys Asn Val Gly 75
Cys Leu Gln Glu Ala Leu Gln Leu 80 Ala Thr Ser Phe Ala Gln Leu Arg 85
Leu Gly Asp Val Lys Asn 90

```

<210> 159  
 <211> 134  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanine tRNA Synthetase polypeptide with C-terminal  
 6xHis affinity tag

```

<400> 159
Met 1 Pro Leu Val 5 Ile Leu Glu Met Glu 10 Ser Gly Ala Ser Ala Lys Ala 15

```

120161\_426PC\_Sequence\_Listing.TXT

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

<210> 160

<211> 451

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanine tRNA Synthetase polypeptide with N-terminal  
6xHis affinity tag

<400> 160

Met	His	His	His	His	His	His	Gly	Lys	Pro	Ile	Pro	Asn	Pro	Leu	Leu
1				5				10						15	
Gly	Leu	Asp	Ser	Thr	Gly	Ser	Ile	Pro	Gln	Ser	Thr	Ile	Thr	Ile	Trp
			20					25					30		
Thr	Pro	Val	Val	Ala	Met	Lys	Thr	Glu	Phe	Thr	Val	Lys	Asn	Ala	Gln
		35				40						45			
Val	Arg	Gly	Gly	Tyr	Val	Leu	His	Ile	Gly	Thr	Ile	Tyr	Gly	Asp	Leu
	50				55					60					
Lys	Val	Gly	Asp	Gln	Val	Trp	Leu	Phe	Ile	Asp	Glu	Pro	Arg	Arg	Arg
65				70				75						80	
Pro	Ile	Met	Ser	Asn	His	Thr	Ala	Thr	His	Ile	Leu	Asn	Phe	Ala	Leu
			85					90					95		
Arg	Ser	Val	Leu	Gly	Glu	Ala	Asp	Gln	Lys	Gly	Ser	Leu	Val	Ala	Pro
			100					105					110		
Asp	Arg	Leu	Arg	Phe	Asp	Phe	Thr	Ala	Lys	Gly	Ala	Met	Ser	Thr	Gln
		115					120					125			
Gln	Ile	Lys	Lys	Ala	Glu	Glu	Ile	Ala	Asn	Glu	Met	Ile	Glu	Ala	Ala
	130					135					140				
Lys	Ala	Val	Tyr	Thr	Gln	Asp	Cys	Pro	Leu	Ala	Ala	Ala	Lys	Ala	Ile
145				150					155					160	
Gln	Gly	Leu	Arg	Ala	Val	Phe	Asp	Glu	Thr	Tyr	Pro	Asp	Pro	Val	Arg
			165					170						175	
Val	Val	Ser	Ile	Gly	Val	Pro	Val	Ser	Glu	Leu	Leu	Asp	Asp	Pro	Ser
		180						185					190		
Gly	Pro	Ala	Gly	Ser	Leu	Thr	Ser	Val	Glu	Phe	Cys	Gly	Gly	Thr	His
		195					200					205			
Leu	Arg	Asn	Ser	Ser	His	Ala	Gly	Ala	Phe	Val	Ile	Val	Thr	Glu	Glu
	210					215					220				
Ala	Ile	Ala	Lys	Gly	Ile	Arg	Arg	Ile	Val	Ala	Val	Thr	Gly	Ala	Glu
225				230						235				240	
Ala	Gln	Lys	Ala	Leu	Arg	Lys	Ala	Glu	Ser	Leu	Lys	Lys	Cys	Leu	Ser
			245					250					255		
Val	Met	Glu	Ala	Lys	Val	Lys	Ala	Gln	Thr	Ala	Pro	Asn	Lys	Asp	Val
		260						265					270		
Gln	Arg	Glu	Ile	Ala	Asp	Leu	Gly	Glu	Ala	Leu	Ala	Thr	Ala	Val	Ile
		275					280					285			
Pro	Gln	Trp	Gln	Lys	Asp	Glu	Leu	Arg	Glu	Thr	Leu	Lys	Ser	Leu	Lys
	290					295					300				
Lys	Val	Met	Asp	Asp	Leu	Asp	Arg	Ala	Ser	Lys	Ala	Asp	Val	Gln	Lys
305					310					315				320	

120161\_426PC\_Sequence\_Listing.TXT

```

Arg Val Leu Glu Lys Thr Lys Gln Phe Ile Asp Ser Asn Pro Asn Gln
325 330
Pro Leu Val Ile Leu Glu Met Glu Ser Gly Ala Ser Ala Lys Ala Leu
340 345 350
Asn Glu Ala Leu Lys Leu Phe Lys Met His Ser Pro Gln Thr Ser Ala
355 360 365
Met Leu Phe Thr Val Asp Asn Glu Ala Gly Lys Ile Thr Cys Leu Cys
370 375 380
Gln Val Pro Gln Asn Ala Ala Asn Arg Gly Leu Lys Ala Ser Glu Trp
385 390 395 400
Val Gln Gln Val Ser Gly Leu Met Asp Gly Lys Gly Gly Lys Asp
405 410 415
Val Ser Ala Gln Ala Thr Gly Lys Asn Val Gly Cys Leu Gln Glu Ala
420 425 430
Leu Gln Leu Ala Thr Ser Phe Ala Gln Leu Arg Leu Gly Asp Val Lys
435 440 445
Asn Leu Glu
450

```

<210> 161

<211> 451

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanine tRNA Synthetase polypeptide with C-terminal  
6xHis affinity tag

<400> 161

```

Met Gly Ser Ile Pro Gln Ser Thr Ile Thr Ile Trp Thr Pro Val Val
1 5 10 15
Ala Met Lys Thr Glu Phe Thr Val Lys Asn Ala Gln Val Arg Gly Gly
20 25 30
Tyr Val Leu His Ile Gly Thr Ile Tyr Gly Asp Leu Lys Val Gly Asp
35 40 45
Gln Val Trp Leu Phe Ile Asp Glu Pro Arg Arg Arg Pro Ile Met Ser
50 55 60
Asn His Thr Ala Thr His Ile Leu Asn Phe Ala Leu Arg Ser Val Leu
65 70 75 80
Gly Glu Ala Asp Gln Lys Gly Ser Leu Val Ala Pro Asp Arg Leu Arg
85 90 95
Phe Asp Phe Thr Ala Lys Gly Ala Met Ser Thr Gln Gln Ile Lys Lys
100 105 110
Ala Glu Glu Ile Ala Asn Glu Met Ile Glu Ala Ala Lys Ala Val Tyr
115 120 125
Thr Gln Asp Cys Pro Leu Ala Ala Ala Lys Ala Ile Gln Gly Leu Arg
130 135 140
Ala Val Phe Asp Glu Thr Tyr Pro Asp Pro Val Arg Val Val Ser Ile
145 150 155 160
Gly Val Pro Val Ser Glu Leu Leu Asp Asp Pro Ser Gly Pro Ala Gly
165 170 175
Ser Leu Thr Ser Val Glu Phe Cys Gly Gly Thr His Leu Arg Asn Ser
180 185 190
Ser His Ala Gly Ala Phe Val Ile Val Thr Glu Glu Ala Ile Ala Lys
195 200 205
Gly Ile Arg Arg Ile Val Ala Val Thr Gly Ala Glu Ala Gln Lys Ala
210 215 220
Leu Arg Lys Ala Glu Ser Leu Lys Lys Cys Leu Ser Val Met Glu Ala
225 230 235 240
Lys Val Lys Ala Gln Thr Ala Pro Asn Lys Asp Val Gln Arg Glu Ile
245 250 255
Ala Asp Leu Gly Glu Ala Leu Ala Thr Ala Val Ile Pro Gln Trp Gln
260 265 270
Lys Asp Glu Leu Arg Glu Thr Leu Lys Ser Leu Lys Lys Val Met Asp
275 280 285
Asp Leu Asp Arg Ala Ser Lys Ala Asp Val Gln Lys Arg Val Leu Glu
290 295 300

```

120161\_426PC\_Sequence\_Listing.TXT

```

Lys Thr Lys Gln Phe Ile Asp Ser Asn Pro Asn Gln Pro Leu Val Ile
305      310      315
Leu Glu Met Glu Ser Gly Ala Ser Ala Lys Ala Leu Asn Glu Ala Leu
      325      330      335
Lys Leu Phe Lys Met His Ser Pro Gln Thr Ser Ala Met Leu Phe Thr
      340      345      350
Val Asp Asn Glu Ala Gly Lys Ile Thr Cys Leu Cys Gln Val Pro Gln
      355      360      365
Asn Ala Ala Asn Arg Gly Leu Lys Ala Ser Glu Trp Val Gln Gln Val
      370      375      380
Ser Gly Leu Met Asp Gly Lys Gly Gly Gly Lys Asp Val Ser Ala Gln
385      390      395      400
Ala Thr Gly Lys Asn Val Gly Cys Leu Gln Glu Ala Leu Gln Leu Ala
      405      410      415
Thr Ser Phe Ala Gln Leu Arg Leu Gly Asp Val Lys Asn Leu Glu Gly
      420      425      430
Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr His His His
      435      440      445
His His His
450

```

<210> 162

<211> 264

<212> PRT

<213> Artificial Sequence

<220>

<223> Anyl tRNA Synthetase polypeptide with N-terminal  
6xHis affinity tag

<400> 162

```

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu
1      5      10      15
Gly Leu Asp Ser Thr Arg His Leu Arg Asn Ser Ser His Ala Gly Ala
      20      25      30
Phe Val Ile Val Thr Glu Glu Ala Ile Ala Lys Gly Ile Arg Arg Ile
      35      40      45
Val Ala Val Thr Gly Ala Glu Ala Gln Lys Ala Leu Arg Lys Ala Glu
      50      55      60
Ser Leu Lys Lys Cys Leu Ser Val Met Glu Ala Lys Val Lys Ala Gln
65      70      75      80
Thr Ala Pro Asn Lys Asp Val Gln Arg Glu Ile Ala Asp Leu Gly Glu
      85      90      95
Ala Leu Ala Thr Ala Val Ile Pro Gln Trp Gln Lys Asp Glu Leu Arg
      100      105      110
Glu Thr Leu Lys Ser Leu Lys Lys Val Met Asp Asp Leu Asp Arg Ala
      115      120      125
Ser Lys Ala Asp Val Gln Lys Arg Val Leu Glu Lys Thr Lys Gln Phe
130      135      140
Ile Asp Ser Asn Pro Asn Gln Pro Leu Val Ile Leu Glu Met Glu Ser
145      150      155      160
Gly Ala Ser Ala Lys Ala Leu Asn Glu Ala Leu Lys Leu Phe Lys Met
      165      170      175
His Ser Pro Gln Thr Ser Ala Met Leu Phe Thr Val Asp Asn Glu Ala
      180      185      190
Gly Lys Ile Thr Cys Leu Cys Gln Val Pro Gln Asn Ala Ala Asn Arg
      195      200      205
Gly Leu Lys Ala Ser Glu Trp Val Gln Gln Val Ser Gly Leu Met Asp
210      215      220
Gly Lys Gly Gly Gly Lys Asp Val Ser Ala Gln Ala Thr Gly Lys Asn
225      230      235      240
Val Gly Cys Leu Gln Glu Ala Leu Gln Leu Ala Thr Ser Phe Ala Gln
      245      250      255
Leu Arg Leu Gly Asp Val Lys Asn
260

```

120161\_426PC\_Sequence\_Listing.TXT

<210> 163  
 <211> 264  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanine tRNA Synthetase polypeptide with C-terminal  
 6xHis affinity tag

<400> 163  
 Met Arg His Leu Arg Asn Ser Ser His Ala Gly Ala Phe Val Ile Val  
 1 5 10 15  
 Thr Glu Glu Ala Ile Ala Lys Gly Ile Arg Arg Ile Val Ala Val Thr  
 20 25 30  
 Gly Ala Glu Ala Gln Lys Ala Leu Arg Lys Ala Glu Ser Leu Lys Lys  
 35 40 45  
 Cys Leu Ser Val Met Glu Ala Lys Val Lys Ala Gln Thr Ala Pro Asn  
 50 55 60  
 Lys Asp Val Gln Arg Glu Ile Ala Asp Leu Gly Glu Ala Leu Ala Thr  
 65 70 75 80  
 Ala Val Ile Pro Gln Trp Gln Lys Asp Glu Leu Arg Glu Thr Leu Lys  
 85 90 95  
 Ser Leu Lys Lys Val Met Asp Asp Leu Asp Arg Ala Ser Lys Ala Asp  
 100 105 110  
 Val Gln Lys Lys Arg Val Leu Glu Lys Thr Lys Gln Phe Ile Asp Ser Asn  
 115 120 125  
 Pro Asn Gln Pro Leu Val Ile Leu Glu Met Glu Ser Gly Ala Ser Ala  
 130 135 140  
 Lys Ala Leu Asn Glu Ala Leu Lys Leu Phe Lys Met His Ser Pro Gln  
 145 150 155 160  
 Thr Ser Ala Met Leu Phe Thr Val Asp Asn Glu Ala Gly Lys Ile Thr  
 165 170 175  
 Cys Leu Cys Gln Val Pro Gln Asn Ala Ala Asn Arg Gly Leu Lys Ala  
 180 185 190  
 Ser Glu Trp Val Gln Gln Val Ser Gly Leu Met Asp Gly Lys Gly Gly  
 195 200 205  
 Gly Lys Asp Val Ser Ala Gln Ala Thr Gly Lys Asn Val Gly Cys Leu  
 210 215 220  
 Gln Glu Ala Leu Gln Leu Ala Thr Ser Phe Ala Gln Leu Arg Leu Gly  
 225 230 235 240  
 Asp Val Lys Asn Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp  
 245 250 255  
 Ser Thr His His His His His  
 260

<210> 164  
 <211> 334  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanine tRNA Synthetase polypeptide with N-terminal  
 6xHis affinity tag (Genewiz)

<400> 164  
 Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15  
 Gly Leu Asp Ser Thr Gly Ser Ile Glu Ala Ala Lys Ala Val Tyr Thr  
 20 25 30  
 Gln Asp Cys Pro Leu Ala Ala Ala Lys Ala Ile Gln Gly Leu Arg Ala  
 35 40 45  
 Val Phe Asp Glu Thr Tyr Pro Asp Pro Val Arg Val Val Ser Ile Gly  
 50 55 60  
 Val Pro Val Ser Glu Leu Leu Asp Asp Pro Ser Gly Pro Ala Gly Ser  
 65 70 75 80  
 Leu Thr Ser Val Glu Phe Cys Gly Gly Thr His Leu Arg Asn Ser Ser  
 85 90 95

120161\_426PC\_Sequence\_Listing.TXT

```

His Ala Gly Ala Phe Val Ile Val Thr Glu Glu Ala Ile Ala Lys Gly
100 110
Ile Arg Arg Ile Val Ala Val Thr Gly Ala Glu Ala Gln Lys Ala Leu
115 125
Arg Lys Ala Glu Ser Leu Lys Lys Cys Leu Ser Val Met Glu Ala Lys
130 140
Val Lys Ala Gln Thr Ala Pro Asn Lys Asp Val Gln Arg Glu Ile Ala
145 155
Asp Leu Gly Glu Ala Leu Ala Thr Ala Val Ile Pro Gln Trp Gln Lys
165 175
Asp Glu Leu Arg Glu Thr Leu Lys Ser Leu Lys Lys Val Met Asp Asp
180 190
Leu Asp Arg Ala Ser Lys Ala Asp Val Gln Lys Arg Val Leu Glu Lys
195 205
Thr Lys Gln Phe Ile Asp Ser Asn Pro Asn Gln Pro Leu Val Ile Leu
210 220
Glu Met Glu Ser Gly Ala Ser Ala Lys Ala Leu Asn Glu Ala Leu Lys
225 235
Leu Phe Lys Met His Ser Pro Gln Thr Ser Ala Met Leu Phe Thr Val
245 255
Asp Asn Glu Ala Gly Lys Ile Thr Cys Leu Cys Gln Val Pro Gln Asn
260 270
Ala Ala Asn Arg Gly Leu Lys Ala Ser Glu Trp Val Gln Gln Val Ser
275 285
Gly Leu Met Asp Gly Lys Gly Gly Lys Asp Val Ser Ala Gln Ala
290 300
Thr Gly Lys Asn Val Gly Cys Leu Gln Glu Ala Leu Gln Leu Ala Thr
305 315
Ser Phe Ala Gln Leu Arg Leu Gly Asp Val Lys Asn Leu Glu
325 330

```

<210> 165

<211> 334

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanyl tRNA Synthetase polypeptide with C-terminal  
6xHis affinity tag (Genewiz)

<400> 165

```

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

```

120161\_426PC\_Sequence\_Listing.TXT

Gly	Ala	Ser	Ala	Lys	Ala	Leu	Asn	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Met
210	210					215					220				
His	Ser	Pro	Gln	Thr	Ser	Ala	Met	Leu	Phe	Thr	Val	Asp	Asn	Glu	Ala
225					230					235					240
Gly	Lys	Ile	Thr	Cys	Leu	Cys	Gln	Val	Pro	Gln	Asn	Ala	Ala	Asn	Arg
				245					250					255	
Gly	Leu	Lys	Ala	Ser	Glu	Trp	Val	Gln	Val	Ser	Gly	Leu	Met	Asp	
			260					265					270		
Gly	Lys	Gly	Gly	Gly	Lys	Asp	Val	Ser	Ala	Gln	Ala	Thr	Gly	Lys	Asn
		275					280					285			
Val	Gly	Cys	Leu	Gln	Glu	Ala	Leu	Gln	Leu	Ala	Thr	Ser	Phe	Ala	Gln
	290					295					300				
Leu	Arg	Leu	Gly	Asp	Val	Lys	Asn	Leu	Glu	Gly	Lys	Pro	Ile	Pro	Asn
305					310					315					320
Pro	Leu	Leu	Gly	Leu	Asp	Ser	Thr	His	His	His	His	His	His	His	
				325					330						

<210> 166

<211> 234

<212> PRT

<213> Artificial Sequence

<220>

<223> Alanyl tRNA Synthetase polypeptide with C-terminal  
6xHis affinity tag (Version 2)

<400> 166

Met	Lys	Gly	Ile	Arg	Arg	Ile	Val	Ala	Val	Thr	Gly	Ala	Glu	Ala	Gln
1				5					10					15	
Lys	Ala	Leu	Arg	Lys	Ala	Glu	Ser	Leu	Lys	Lys	Cys	Leu	Ser	Val	Met
			20					25					30		
Glu	Ala	Lys	Val	Lys	Ala	Gln	Thr	Ala	Pro	Asn	Lys	Asp	Val	Gln	Arg
		35				40					45				
Glu	Ile	Ala	Asp	Leu	Gly	Glu	Ala	Leu	Ala	Thr	Ala	Val	Ile	Pro	Gln
	50					55				60					
Trp	Gln	Lys	Asp	Glu	Leu	Arg	Glu	Thr	Leu	Lys	Ser	Leu	Lys	Lys	Val
65				70					75					80	
Met	Asp	Asp	Leu	Asp	Arg	Ala	Ser	Lys	Ala	Asp	Val	Gln	Lys	Arg	Val
			85					90						95	
Leu	Glu	Lys	Thr	Lys	Gln	Phe	Ile	Asp	Ser	Asn	Pro	Asn	Gln	Pro	Leu
			100					105					110		
Val	Ile	Leu	Glu	Met	Glu	Ser	Gly	Ala	Ser	Ala	Lys	Ala	Leu	Asn	Glu
		115					120					125			
Ala	Leu	Lys	Leu	Phe	Lys	Met	His	Ser	Pro	Gln	Thr	Ser	Ala	Met	Leu
	130					135					140				
Phe	Thr	Val	Asp	Asn	Glu	Ala	Gly	Lys	Ile	Thr	Cys	Leu	Cys	Gln	Val
145					150					155				160	
Pro	Gln	Asn	Ala	Ala	Asn	Arg	Gly	Leu	Lys	Ala	Ser	Glu	Trp	Val	Gln
			165					170						175	
Gln	Val	Ser	Gly	Leu	Met	Asp	Gly	Lys	Gly	Gly	Gly	Lys	Asp	Val	Ser
			180					185					190		
Ala	Gln	Ala	Thr	Gly	Lys	Asn	Val	Gly	Cys	Leu	Gln	Glu	Ala	Leu	Gln
		195					200					205			
Leu	Ala	Thr	Ser	Phe	Ala	Gln	Leu	Arg	Leu	Gly	Asp	Val	Lys	Asn	Ala
	210					215					220				
Ala	Ala	Leu	Glu	His	His	His	His	His	His	His	His	His	His	His	
225					230										

<210> 167

<211> 1491

<212> DNA

<213> Artificial Sequence

<220>

<223> Codon optimized Alanyl tRNA Synthetase  
polynucleotide

120161\_426PC\_Sequence\_Listing.TXT

<400> 167

```

cgcggcctgg aagtaacgga tgacagccca aaatacaact accatctgga ctctagcggc 60
tcgtacgtct ttgaaaatac cgtcgccacg gttatggctt tgcgccgtga gaagatgttc 120
gtggaagagg tticcaccgg ccaggaatgc ggtgtggtgc tggataagac ctgcttttac 180
gcgagcaggg gtggtcagat ttacgatgag ggttatctgg tcaaagttga cgattctagc 240
gaagataaga cggagttcac cgtcaaaaac gcgcaagtcc gcggtggcta cgtcctgcac 300
attggcacta tctatggtga tctgaaggtt ggtgaccaag ttggctgttt catcgacgag 360
ccgcgtcgtc gtccgatcat gagcaatcat accgcgacgc atatcctgaa ctttgcgctg 420
cgttcgggtcc tgggtgaggg ggatcaaaaa ggctccctgg tggctccgga tcgcctgcgc 480
ttcgacttca ctgccaaagg cgcatgagc acccagcaga ttaagaaagc cgaagagatc 540
gcaaacgaga tgattgaggg tgcaaggcgg gtctataccc aggactgtcc gctggcggca 600
gcgaaggcta ttcaagggtc gcgcgtgttt ttgatgaaa cttatccgga tccagttcgt 660
gttggttagca tcggcgtgcc ggtgagcgaa ctgttgacg atccgagcgg tccggcaggt 720
agcttgacga cgtcgagtt ttgtggtggc acccacctgc gtaacagcag ccacgcgggt 780
gcattcggtt tctgaccga agaggccatt gcaaaaggca tccgtcgcat cgttgcgggt 840
accggtgccg aggcgcagaa ggcactgcgt aaagcggaag gcttgaagaa atgcctgagc 900
gtgatggagg cgaaggtgaa agcacagacg gcaccgaata aggatgtgca acgtgagatt 960
gccgatttgg gtgaggccct ggcgaccgca gttattccgc agtggcaaaa ggacgagctg 1020
cgtgaaaacc tgaaatccct gaaaaagggt atggacgacc tggaccgcgc atccaaagcg 1080
gacgtccaaa agcgtgtcct ggagaaaacc aaacaattca ttgactccaa cccgaaccag 1140
cctctggtca ttctggaaat ggagagcggg gccagcgcaa aggctctgaa tgaggcgctg 1200
aagctgttca agatgcacag cccgcagacg agcgcgatgc tgtttaccgt tgataatgaa 1260
gccggcaaaa tcacgtgtct gtgccaagtc ccgcagaatg cggcgaatcg tggtttgaaa 1320
gccagcgaat ggggtcaaca ggtgtctggt ttgatggacg gtaaggttgg cggcaaagac 1380
gttagcgcac aagctaccgg taagaatgtg ggctgcctgc aagaagcgct gcagctggcg 1440
acgagctttg cacagctgcg tttgggtgat gtgaaaaact aatgactcga g 1491

```

<210> 168

<211> 645

<212> DNA

<213> Artificial Sequence

<220>

<223> Codon optimized Alanyl tRNA Synthetase polynucleotide

<400> 168

```

ggatccgcag aagcacagaa agccctgcgt aaagccgaat ctctgaaaaa atgcctgagc 60
gtgatggaag cgaaagtcaa agcccaaaca gcaccgaata aagacgtcca acgtgaaatt 120
gccgatctgg gtgaggcgct ggctacagcc gttatccac aatggcagaa agatgaactg 180
cgtgaaaccc tgaaaagcct gaaaaaagtg atggatgacc tggatcgtgc ctctaaagca 240
gatgttcaaa aacgtgtcct ggagaaaacg aaacaattca ttgatagcaa cccgaaccaa 300
ccactggtaa tcctggaaat ggaaagcgga gcatacgcaa aagcactgaa cgaagcgctg 360
aaactgttca aaatgcacag cccacagaca tccgccatgc tgttcaccgt tgataatgaa 420
gccggcaaaa tcacctgcct gtgtcaggtt cctcagaacg ctgctaaccg tggactgaaa 480
gcaagcgaat gggttcaaca agtgagtggg ctgatggacg gtaaaggttg tggtaaagat 540
gtgtcagcac aagcgaccgg taaaaatgtt ggggtgcctg aggaagctct gcaactggca 600
acctcatttg cccaactgcg cctgggagat gtgaaaaacc tcgag 645

```

<210> 169

<211> 1155

<212> DNA

<213> Artificial Sequence

<220>

<223> Codon optimized Alanyl tRNA Synthetase polynucleotide

<400> 169

```

caagtctggc tgttcacgaa cgaaccacgc cgccgcccaa ttatgagcaa ccataccgcg 60
acctatattc tgaactttgc cctgcgttct gtgctgggag aagcagatca gaagggcagc 120
ctggttgctc cggatcgtct gcgttttgat ttaccgcga aaggtgcgat gtctacgcaa 180
cagatcaaga aggcggaaga aattgctaac gagatgatcg aggcggcgaa agcggtttac 240
acgcaagatt gcccgctggc tgccgcaaaa gcgattcagg gtctgcgtgc ggtgtttgat 300
gaaacttacc cggaccgggt tcgcgttgtg tccatcggcg tgccggtgag cgagctgttg 360
gacgaccgga cgggtccggc tggcagcctg accagcgctt aattctgtgg tggtaaccac 420
ttgcgcaaca gcagccacgc cgggtgcgtt gttatttgta ccgaagaggc aattgcaaaa 480

```



120161\_426PC\_Sequence\_Listing.TXT

```

ggtatccgtc gtatcgtggc ggtcaccggt gctgaggcac aaaaggcgct gcgtaaggcg 540
gagtcgttga agaaatgttt gagcgttatg gaggcgaaag ttaaagcgca gacggcaccg 600
aataaggatg tccagcgga gatcgcgga ctgggtgagg ccctggccac cgcggtgatt 660
ccgcaatggc agaaagatga gctgcgtgaa acgctgaaga gcttgaaaaa ggtcatggac 720
gacctggacc gtgcgagcaa ggcagacgta cagaagcgcg ttctggagaa aactaaacag 780
tttatcgaca gcaacccgaa ccaaccgctg gtcatctctg aaatggaaa cggcgcaagc 840
gctaaggcac tgaatgaggc gctgaaactg ttcaaaatgc acagccctca gacgtctgca 900
atgctgttca ccgtcgataa tgaagcgggt aagatcacgt gcctgtgtca ggtgccgcaa 960
aacgccgca atcgtggtct gaaagcgctc gagtgggtcc aacaggtgag cggctctgatg 1020
gatggcaagg gtggcggtaa agatgtttcc gccaggcca ccggcaaaaa tgttggttgc 1080
ctgcaagagg cactgcaact ggcaaccagc ttgcccagc tgcgtttggg cgacgtcaag 1140
aattaatgac tcgag                                     1155

```

<210> 170

<211> 939

<212> DNA

<213> Artificial Sequence

<220>

<223> Codon optimized Alanyl tRNA Synthetase  
polynucleotide (DNA2.0)

<400> 170

```

atcgaagccg caaaagccgt ctacaccag gactgcccgc tggcggctgc caaggccatc 60
caaggtttgc gtgcagcttt tgacgaaacg tatcctgac cgggtgcgct cgtgagcatc 120
ggtgtgccag tgtctgagct gctggacgat ccgtcgggtc cggcaggctc tctgaccagc 180
gttgaattct gcggcggtag ccatctgcgt aatagcagcc acgccggtgc gttcgttatc 240
gttaccgaag aggcgattgc gaaaggcatt cgtcgtatcg ttgcagtcac tgggtcggag 300
gcgcaaaagg cgttgcgtaa agcagaaagc ctgaagaaat gtctgagcgt aatggaagct 360
aaagttaagg cccaaacggc accgaacaaa gacgtgcagc gcgagattgc agacctgggc 420
gaggcgctgg cgacggcagt tattccgcag tggcagaaag acgaactgcg cgaaacgctg 480
aaaagcttga aaaaggttat ggatgacttg gaccgtgcaa gcaaggctga tgtccaaaag 540
cgcgctcctg agaaaaccaa gcagttcatt gatagcaatc cgaaccaacc gctggtcatt 600
ctggagatgg agagcggtag gagcgcgaag gcgctgaacg aggctctgaa gttgtttaag 660
atgcactccc cgcaaacgag cgcaatgctg ttaccggtg ataatgaagc gggtaagatc 720
acctgtctgt gccagggtgc gcaaaatgcg gctaaccgtg gcctgaaagc gtccgagttg 780
gtccagcagg ttccgggtct gatggatggc aaaggcgggt gtaaagacgt gagcgcgcag 840
gccaccggca aaaatgtggg ttgtctgcaa gaggccctgc agctggcgac cagcttcgcg 900
caactgcgtc tgggtgatgt gaagaactaa tgactcgag                                     939

```

<210> 171

<211> 846

<212> DNA

<213> Artificial Sequence

<220>

<223> Codon optimized Alanyl tRNA Synthetase  
polynucleotide

<400> 171

```

ccagaccag tccgcgtagt aagcatcggc gtcccagtaa gcgagctgct ggacgatccg 60
tctggccctg ccggctccct gaccagctg gagttttgc gtggcaccca tctgcgcaat 120
tccagccacg cgggtgcttt tgtgatctg accgaagaag ctattgcaaa gggatttcgt 180
cgtatcgtgg cggtagccgg tgcaagact cagaaggccc tgcgtaaagc cgagtcgctg 240
aagaaatgtt tgagcgttat ggaggcaaag gttaaagcgc aaaccgcacc gaacaaggac 300
gttcagcgtg aaattgcgga cctgggtgag gcgctggcga cggcagtgat cccgcaatgg 360
cagaaagatg agctgcgcga aaccttgaag agcttgaaga aagttatgga tgatttggac 420
cgcgccagca aagctgatgt tcaaaagcgt gtcttggaag agacaaaca attcattgac 480
agcaaccgga atcagccgct ggttatctg gagatggaga gcggcgcatc cgcaaggcg 540
ctgaacgaag cactgaaact gtttaagat cacagcccgc aaactagcgc gatgctgttc 600
acggttgaca atgagcggg taaaatcacg tgacctgtcc aagtgccgca gaacgcggca 660
aatcgtggtc tgaagccag cgaatgggtt cagcaggtca cgggcctgat ggatggtaaa 720
ggtggtggca aggacgtcag cgcgcaagcg accggtgaag atgtgggctg tctgcaagag 780
gcgctgcagc tggcgacgtc tttcgcccag ctgcgtctgg gtgatgtcaa aaactaatga 840
ctcgag                                     846

```

<210> 172

<211> 678

120161\_426PC\_Sequence\_Listing.TXT

<212> DNA  
<213> Artificial Sequence

<220>  
<223> Codon optimized Alanyl tRNA Synthetase  
polynucleotide

```
<400> 172
ggatccaaag gcatccgccg tattgtagca gtaaccggtg ccgaggctca aaaagcgctg 60
cgaaagctg aaagcctgaa aaaatgcctg tcggtgatgg aggcaaaagt gaaagcacag 120
acagctccta acaaagatgt tcaacgtgag atcgctgacc tgggagaagc actggctacc 180
gccgtcattc ctacgtggca aaaagacgag ctgcgcgaga cactgaaatc actgaaaaaa 240
gtcatggacg atctggaccg tgctagtaaa gccgatgttc aaaaacgtgt cctggagaaa 300
accaaacaat tcatcgactc caaccctaata cagccgctgg ttatcctgga gatggagagt 360
ggcgctagtg ccaaagcgct gaacgaggct ctgaaactgt tcaaaatgca ctcccctcaa 420
acctcagcaa tgctgttcac agtggaacaac gaggctggca aaatcacatg cctgtgccaa 480
gtgccacaaa atgccgcaa tcgtggtctg aaagcgctcag aatgggtaca acaagtgagc 540
ggactgatgg atggcaaagg gggcggaaaa gacgttagtg cccaagcaac tgggaaaaaac 600
gtggggtgtc tgcaagaggc tctgcagctg gctacgagtt tcgccaact gcgtctggga 660
gatgtgaaaa atctcgag                                     678
```

<210> 173  
<211> 351  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Codon optimized Alanyl tRNA Synthetase  
polynucleotide

```
<400> 173
ccactggtca ttttgaaat ggagagcggg gcaagcgcaa aagcgctgaa cgaagctctg 60
aagctgttca agatgcactc tccgcaaacg agcgcgatgc tgtttaccgt cgataatgag 120
gccggtaaaa tcacctgcct gtgtcaggtt ccgcagaatg cggccaaccg tggtttgaag 180
gcttccgagt gggttcagca agtgagcggg ctgatggatg gcaaaggtgg tggcaaagac 240
gtgagcgcgc aggtacggg caaaaacgtg ggttgccctg aagaagcgct gcaattggcg 300
accagcttcg cacagctgcg cctgggcgac gtttaagaatt aatgactcga g 351
```

<210> 174  
<211> 1290  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Codon optimized Alanyl tRNA Synthetase  
polynucleotide

```
<400> 174
ggatccatcc ctacagagcac cattacgatt tggaccccgg ttgttgccat gaaaacggag 60
tttaccgtga aaaatgccca gggtcgtgga ggttatgttc tgcacattgg gacgatctat 120
ggcgacctga aagtgtgtga ccaagtgtgg ctgttcacatg atgagcctcg ccgccgtcct 180
atcatgagta accataccgc caccatatt ctgaatttcg ccctgcgttc agtactggga 240
gaagcagacc aaaaaggcag cctggtagcc cctgatcgcc tgcgctttga ttttacggcc 300
aaaggagcca tgtctacaca gcagatcaaa aaagccgaag agatcgaaa tgagatgac 360
gaagccgcca aagcagttta taccaggat tgtccgctgg cagcagcaaa agccattcaa 420
ggcctgcgtg cgggtgttga cgaaacttat ccggaccag tccgtgttgt gagtatcgg 480
gtccctgtta gcgaactgct ggatgaccca tctggtcctg ccggttact gacaagcgtg 540
gaattttgtg gtggaaccca cctgcgtaat agcagtcatt ctggggcctt tgttattgtg 600
accgaagagg ccatgtctaa aggaattcgc cgtattgttg ccgttacagg tgctgaagct 660
cagaaagcac tgcgcaaagc tgaatcgctg aaaaaatgcc tgagcgtgat ggaagcgaaa 720
gtgaaagctc aaaccgccc gaataaagat gttcaacgtg agattgccga tctgggagag 780
gcgctggcta cagctgttat tcctcagtg cagaaagatg agctgcgtga aaccctgaaa 840
agtctgaaaa aagtgtatga tgacctggac cgtgttcaa aagcagatgt tcaaaaacgt 900
gtgctggaga aaacgaaaca gtttatcgac agcaaccgga atcaacctct ggttattctg 960
gagatggaaa gcggtgctag tgctaaagcc ctgaatgaag ccctgaaact gttcaaaatg 1020
cacagccctc agacatcagc catgctgttt acgttgata acgaagcagg caaaatcacc 1080
tgtctgtgtc aggtccctca aaacgctgct aatcgtggcc tgaaagcaag cgaatgggta 1140
cagcaggttt ctggactgat ggatggtaaa ggtggtggca aagacgttag tgcccaagcg 1200
```

120161\_426PC\_Sequence\_Listing.TXT

acgggtaaaa atgttggctg tctgcaggaa gctctgcagc tggccacctc atttgcctca 1260  
ctgcgtctgg gagatgtgaa aaacctcgag 1290

<210> 175  
<211> 741  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Codon optimized Alanyl tRNA Synthetase  
polynucleotide

<400> 175  
cgatcatctga gaaactctag ccacgccggc gccttcgtca tcgttactga agaagcgatc 60  
gcgaaaaggta tccgtcgcat tgtagcggg accggtgcgg aagcgcaaaa agcgctgcgt 120  
aaagcggaga gcctgaagaa atgcttgagc gttatggagg cgaaagtga ggctcagacc 180  
gcaccgaaca aagatgtcca acgtgaaatt gcggacctgg gtgaagcgct ggccaccgca 240  
gttatcccg c agtggcaaaa ggacgagctg cgtgaaacct tgaagagcct gaaaaagggt 300  
atggacgatc tggaccgtgc ctcgaaagcg gacgtgcaaa aacgcgttct ggagaaaacc 360  
aagcagttta ttgattccaa cccgaatcaa ccgctggtta tcctggagat ggaatccggt 420  
gcgagcgcaa aggccctgaa tgaggccctg aagctgttca aaatgcacag cccgcagacg 480  
agcgcaatgc tgtttacggt ggacaatgag gctggtaaga ttacgtgtct gtgtcaggtg 540  
cctcagaatg cagcgaaccg cggctctgaag gcgagcgagt gggccaaca ggtgagcggc 600  
ctgatggatg gcaaagggtg cggcaaggat gttagcgcac aagcaaccgg caaaaatgtc 660  
ggttgccctgc aagaggctct gcagttggca acgtctttcg ctcaattgcg cttgggtgat 720  
gtcaagaact aatgactcga g 741

<210> 176  
<211> 939  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Codon optimized Alanyl tRNA Synthetase  
polynucleotide (Genewiz)

<400> 176  
ggatccattg aagccgccaa agccgtgtat acacaagatt gtcctctggc tgctgctaaa 60  
gcaattcagg gactgcgtgc cgtgtttgat gaaacctatc cggaccgggt tcgtgtttgtg 120  
agtattgggt tcccggttag tgaactgctg gatgaccctt ctggtccagc gggttctctg 180  
acatctgtgg agttctgtgg tggtaacacac ctgcgtaata gcagtcatgc cgggtgccttt 240  
gtgattgtta ccgaagaggc cattgccaaa gggattcgcc gtatcgttgc tgttactgga 300  
gctgaagcac aaaaagcact gcgcaaagcc gaatctctga aaaaatgcct gtccgtgatg 360  
gaagcaaaaag tgaagctca aaccgccccg aataaagacg ttcaacgtga aatcgctgat 420  
ctgggcgaag cactggctac cgctgtcatt cctcagtggc aaaaagatga actgcgcgaa 480  
acgctgaaaa gtctgaaaaa agtgatggat gacctggatc gtgctcaaaa agccgatgtc 540  
caaaaacgtg tgctggagaa aacaaaacag tttatcgata gcaaccctaa tcaacctctg 600  
gtgatccctg aaatggaaag tggagcgtcg gctaaagctc tgaatgaagc actgaaactg 660  
ttcaaaatgc atagccctca gacatccgcc atgctgttca ccgtggataa tgaagctggc 720  
aaaatcacct gtctgtgcca agttcctcaa aatgccgcca atcgtgggtc gaaagcaagc 780  
gaatgggttc aacaagtgtc tggctctgat gatggtaaag ggggtggtaa agatgtttct 840  
gccagggcaa ccggtaaaaa cgtgggttgt ctgcaagaag cactgcaact ggcaacctca 900  
tttgcccaac tgcgtctggg tgatgtgaaa aacctcgag 939

<210> 177  
<211> 274  
<212> PRT  
<213> Homo sapiens

<400> 177  
Arg Gly Leu Glu Val Thr Asp Asp Ser Pro Lys Tyr Asn Tyr His Leu  
1 5 10 15  
Asp Ser Ser Gly Ser Tyr Val Phe Glu Asn Thr Val Ala Thr Val Met  
20 25 30  
Ala Leu Arg Arg Glu Lys Met Phe Val Glu Glu Val Ser Thr Gly Gln  
35 40 45  
Glu Cys Gly Val Val Leu Asp Lys Thr Cys Phe Tyr Ala Glu Gln Gly  
50 55 60

120161\_426PC\_Sequence\_Listing.TXT

Gly 65	Gln	Ile	Tyr	Asp	Glu 70	Gly	Tyr	Leu	Val	Lys 75	Val	Asp	Asp	Ser	Ser 80
Glu	Asp	Lys	Thr	Glu 85	Phe	Thr	Val	Lys	Asn 90	Ala	Gln	Val	Arg	Gly 95	Gly
Tyr	Val	Leu	His 100	Ile	Gly	Thr	Ile	Tyr 105	Gly	Asp	Leu	Lys	Val 110	Gly	Asp
Gln	Val	Trp 115	Leu	Phe	Ile	Asp	Glu 120	Pro	Arg	Arg	Arg	Pro 125	Ile	Met	Ser
Asn	His 130	Thr	Ala	Thr	His 135	Ile	Leu	Asn	Phe	Ala	Leu 140	Arg	Ser	Val	Leu
Gly 145	Glu	Ala	Asp	Gln	Lys 150	Gly	Ser	Leu	Val	Ala 155	Pro	Asp	Arg	Leu	Arg 160
Phe	Asp	Phe	Thr	Ala 165	Lys	Gly	Ala	Met	Ser 170	Thr	Gln	Gln	Ile	Lys 175	Lys
Ala	Glu	Glu	Ile 180	Ala	Asn	Glu	Met	Ile 185	Glu	Ala	Ala	Lys	Ala 190	Val	Tyr
Thr	Gln	Asp 195	Cys	Pro	Leu	Ala	Ala 200	Ala	Lys	Ala	Ile	Gln 205	Gly	Leu	Arg
Ala	Val	Phe 210	Asp	Glu	Thr	Tyr 215	Pro	Asp	Pro	Val	Arg 220	Val	Val	Ser	Ile
Gly 225	Val	Pro	Val	Ser	Glu 230	Leu	Leu	Asp	Asp	Pro 235	Ser	Gly	Pro	Ala	Gly 240
Ser	Leu	Thr	Ser	Val 245	Glu	Phe	Cys	Gly	Gly 250	Thr	His	Leu	Arg	Asn 255	Ser
Ser	His	Ala	Gly 260	Ala	Phe	Val	Ile	Val 265	Thr	Glu	Glu	Ala	Ile 270	Ala	Lys
Gly	Ile														

<210> 178  
 <211> 822  
 <212> DNA  
 <213> Homo sapiens

<400> 178  
 cggggtctgg aggtcacaga tgattcccca aagtacaatt accattttgga ctccagtggg 60  
 agctatgtat ttgagaacac agtggctacg gtgatggctc tgcgcaggga gaagatgttc 120  
 gtggaagagg tgtccacagg ccaggagtgt ggagtgggtc tggacaagac ctgtttctat 180  
 gctgagcaag gaggccagat ctatgacgaa ggctacctgg tgaaggtgga tgacagcagt 240  
 gaagataaaa cagagtttac agtgaagaat gctcagggtc gaggagggtg tgtgctacac 300  
 attggaacca tctacggtga cctgaaagtg ggggatcagg tctggctgtt tattgatgag 360  
 ccccgacgaa gacccatcat gagcaaccac acagctacgc acattctgaa cttcgccctg 420  
 cgctcagtgc ttggggaagc tgaccagaaa ggctcattgg ttgctcctga ccgcctcaga 480  
 ttgacttta ctgccaaggg agccatgtcc acccaacaga tcaagaaggc tgaagagatt 540  
 gctaatgaga tgattgaggc agccaaggcc gtctataccc aggattgccc cctggcagca 600  
 gcgaaagcca tccagggcct acgggctgtg ttgatgaga cctatcctga ccctgtgcga 660  
 gtcgtctcca ttgggggtccc ggtgtccgag ttgctggatg acccctctgg gcctgtggc 720  
 tccctgactt ctgttgagtt ctgtggggga acgcacgtgc ggaactcgag tcatgcagga 780  
 gcttttgtga tcgtgacgga agaagccatt gccaaaggga tc 822

<210> 179  
 <211> 299  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Alanine tRNA Synthetase polypeptide with N-terminal 6xHis affinity tag

<400> 179  
 Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15  
 Gly Leu Asp Ser Thr Gly Ser Arg Gly Leu Glu Val Thr Asp Asp Ser  
 20 25 30  
 Pro Lys Tyr Asn Tyr His Leu Asp Ser Ser Gly Ser Tyr Val Phe Glu  
 35 40 45  
 Asn Thr Val Ala Thr Val Met Ala Leu Arg Arg Glu Lys Met Phe Val

	50					55					60				
Glu 65	Glu Val	Ser Thr	Gly 70	Gln Gly	Glu Cys	Gly Val	Val 75	Leu Asp	Lys Thr						
Cys	Phe Tyr	Ala 85	Glu Gln	Gly Gly	Gln Ile	Tyr 90	Tyr Asp	Glu Gly	Tyr 95	Leu Phe	Thr Thr	Ile Tyr			
Val	Lys Val	Asp 100	Asp Ser	Ser Ser	Glu Asp	Lys 105	Thr Leu	His Ile	Gly 110	Thr Thr	Ile Tyr				
Asn	Ala Gln	Val Arg	Gly Gly	Tyr 120	Val Val	Trp Leu	Phe 140	His Ile	Asp Glu	Pro Pro					
Gly	Asp 130	Leu Lys	Val Gly	Asp 135	Ser Asn	His Thr	Ala 155	Thr His	Ile Leu	Asn 160					
Arg 145	Arg Arg	Pro Ile	Met 150	Ser Ser	Asn His	Thr Ala	170	Asp Gln	Lys Gly	Ser 175	Leu Asn				
Phe	Ala Leu	Arg Ser	165	Val Leu	Gly Glu	Ala 170	Asp Gln	Lys Gly	Ser 175	Leu Asn					
Val	Ala Pro	Asp 180	Arg Arg	Leu Arg	Phe Asp	185	Phe Thr	Ala Lys	Gly 190	Ala Met					
Ser	Thr Gln	Gln Ile	Lys Lys	Ala 200	Glu Glu	Ile Ala	Asn 205	Glu Met	Ile Ile						
Glu	Ala 210	Ala Lys	Ala Val	Tyr 215	Arg Ala	Val Phe	Asp 235	Glu Thr	Tyr Pro	Asp 240					
Lys 225	Ala Ile	Gln Gly	Leu 230	Arg Arg	Ala Val	Phe Asp	235	Glu Thr	Tyr Pro	Asp 240					
Pro	Val Arg	Val Val	245	Ser Ile	Gly Val	Pro 250	Val Ser	Glu Leu	Leu 255	Asp Gly					
Asp	Pro Ser	Gly 260	Pro Ala	Gly Ser	Leu 265	Thr Ser	Val Ser	Glu Phe	Cys Gly						
Gly	Thr His	Leu Arg	Asn Ser	Ser 280	His Ala	Gly Ala	Phe 285	Val Ile	Val Val						
Thr	Glu 290	Ala Ile	Ala Lys	295	Gly Ile	Leu Glu									

<220>  
<223> Al anyl tRNA Synthetase pol ypepti de wi th C-termi nal  
6xHi s affi ni ty tag

Page 85

Gly 195	Leu 210	Arg 215	Ala 220	Val 225	Phe 230	Asp 235	Asp 240	Glu 245	Thr 250	Tyr 255	Pro 260	Asp 265	Pro 270	Val 275	Arg 280	Val 285
Val 225	Ser 230	Ile 235	Gly 240	Val 245	Pro 250	Thr 255	Ser 260	Val 265	Glu 270	Phe 275	Leu 280	Cys 285	Gly 290	Gly 295	Thr 300	His 305
Arg 285	Asn 290	Ser 295	Ser 300	His 305	Ala 310	Gly 315	Ala 320	Phe 325	Val 330	Ile 335	Val 340	Thr 345	Glu 350	Glu 355	Ala 360	Leu 365
Ile 355	Ala 360	Lys 365	Gly 370	Ile 375	Leu 380	Glu 385	Gly 390	Lys 395	Pro 400	Ile 405	Pro 410	Asn 415	Pro 420	Leu 425	Leu 430	Val 435
Gly 425	Leu 430	Asp 435	Ser 440	Thr 445	His 450	His 455	His 460	His 465	His 470	His 475	His 480	His 485	His 490	His 495	His 500	His 505

<220>  
<223> Codon optimized A1 anyl tRNA Synthetase  
pol ynucl eoti de

<210> 182  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<210> 183  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<210> 184  
<211> 5  
<212> PRT  
<213> Arti fi ci al Sequence

Page 86

<223> Linker

<400> 184

Ala Ala Ala Leu Glu  
1 5

<210> 185

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligonucleotide

<400> 185

gcggccgcac tcgagcacca ccaccaccac cactga

36

<210> 186

<211> 56

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligonucleotide

<400> 186

cctgtgacag ccacaatcct ccggatacc ttcatatgta tatctccttc ttaaag

56