

(19) AUSTRALIAN PATENT OFFICE

(54) Title
Acid alpha-glucosidase and fragments thereof

(51)⁶ International Patent Classification(s)
C12N 9/24 (2006.01) 9/26
A61K 38/47 (2006.01) 20060101ALI2006072
C12N 9/26 (2006.01) ^{2BMEP} **C12N**
C12N 15/62 (2006.01) 15/62
C12P 21/04 (2006.01) 20060101ALI2005100
A61K 38/00 (2006.01) ^{8BMEP} **C12P**
C12N 9/24 21/04
20060101AFI2005100 20060101ALI2005111
^{8BMEP} **A61K** ^{0BMEP} **A61K**
38/47 38/00
20060101ALI2007072 20060101ALN200707
^{1BMEP} **C12N** ^{21BMEP}
PCT/US2005/004286

(21) Application No: 2005212435 (22) Application Date: 2005 .02 .10

(87) WIPO No: W005/078077

(30) Priority Data

(31) Number	(32) Date	(33) Country
60/543,812	2004 .02 .10	US

(43) Publication Date : 2005 .08 .25

(71) Applicant(s)
Zystor Therapeutics, Inc.

(72) Inventor(s)
Maga, John; Lebowitz, Jonathan

(74) Agent/Attorney
Pizzeys, Level 2, Woden Plaza Offices Woden Town Square Woden, Canberra, ACT, 2606

(56) Related Art
WO 2003/102583 A

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

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
25 August 2005 (25.08.2005)

PCT

(10) International Publication Number
WO 2005/078077 A3

- (51) International Patent Classification⁷: C12N 9/24,
A61K 38/47, C12N 15/62
- (21) International Application Number: PCT/US2005/004286
- (22) International Filing Date: 10 February 2005 (10.02.2005)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/543,812 10 February 2004 (10.02.2004) US
- (71) Applicant (for all designated States except US): ZYSTOR
THERAPEUTICS, INC. [US/US]; 10437 Innovation
Drive, Suite 100, Milwaukee, WI 53226-4838 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): LEBOWITZ,
Jonathan [US/US]; 13 Lynnbrook Road, Frontenac, MO
63131 (US). MAGA, John [US/US]; 5939 N. Bay Ridge
Ave., Whitefish Bay, WI 53217 (US).
- (74) Agents: CHEN, Fangli et al.; Kirkpatrick & lockhart
Nicholson Graham LLP, 75 State Street, Boston, MA
02109-1808 (US).
- (81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
- (84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IT, LI, LU, MC, NL, PL, PT, RO,
SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN,
GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— with international search report
— before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments
- (88) Date of publication of the international search report:
20 October 2005
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2005/078077 A3

(54) Title: ACID ALPHA-GLUCOSIDASE AND FRAGMENTS THEREOF

(57) Abstract: Targeted acid alpha-glucosidase therapeutics that localize to the lysosome are provided. The targeted therapeutics include a therapeutic agent, GAA, and a targeting moiety that binds a receptor on an exterior surface of the cell, permitting proper subcellular localization of the targeted therapeutic upon internalization of the receptor. Nucleic acids, cells, and methods relating to the practice of the invention are also provided.

ACID ALPHA-GLUCOSIDASE AND FRAGMENTS THEREOF**REFERENCE TO RELATED APPLICATIONS**

[0001] This application claims the benefit of U.S. Serial No. 60/543,812, filed February 10, 2004, the contents of which are incorporated by reference.

5

BACKGROUND

[0002] Acid alpha-glucosidase (GAA) is a lysosomal enzyme that hydrolyzes the alpha 1-4 linkage in maltose and other linear oligosaccharides, including the outer branches of glycogen, thereby breaking down excess glycogen in the lysosome (Hirschhorn *et al.* (2001) in The Metabolic and Molecular Basis of Inherited Disease, Scriver, *et al.*, eds. (2001), McGraw-Hill: New York, p. 3389-3420). Like other mammalian lysosomal enzymes, GAA is synthesized in the cytosol and traverses the ER where it is glycosylated with N-linked, high mannose type carbohydrate. In the golgi, the high mannose carbohydrate is modified on lysosomal proteins by the addition of mannose-6-phosphate (M6P) which targets these proteins to the lysosome. The M6P-modified proteins are delivered to the lysosome via interaction with either of two M6P receptors. The most favorable form of modification is when two M6Ps are added to a high mannose carbohydrate.

[0003] Insufficient GAA activity in the lysosome results in Pompe disease, a disease also known as acid maltase deficiency (AMD), glycogen storage disease type II (GSDII), glycogenosis type II, or GAA deficiency. The diminished enzymatic activity occurs due to a variety of missense and nonsense mutations in the gene encoding GAA. Consequently, glycogen accumulates in the lysosomes of all cells in patients with Pompe disease. In particular, glycogen accumulation is most pronounced in lysosomes of cardiac and skeletal muscle, liver, and other tissues. Accumulated glycogen ultimately impairs muscle function. In the most severe form of Pompe disease, death occurs before two years of age due to cardio-respiratory failure.

[0004] Presently, there is no approved treatment available to cure or slow the progress of Pompe disease. Enzyme replacement therapeutics currently in clinical trials require that

administered recombinant GAA be taken up by the cells in muscle and liver tissues and be transported to the lysosomes in those cells in an M6P-dependent fashion. However, recombinant GAA produced in engineered CHO cells and in the milk of transgenic rabbits, two sources of enzymes used in recent Pompe enzyme replacement therapy trials, contains extremely little M6P (Van Hove et al. (1996) Proc Natl Acad Sci U S A, 93(1):65-70; and U.S. Patent No. 6,537,785). Therefore, M6P-dependent delivery of recombinant GAA to lysosomes is not efficient, requiring high dosages and frequent infusions. Accordingly, there remains a need for new, simpler, more efficient, and more cost-effective methods for targeting therapeutic GAA enzymes to patient lysosomes.

10

SUMMARY OF THE INVENTION

[0005] The present invention permits M6P-independent targeting of human GAA or GAA-like enzymes to patient lysosomes by using a peptide tag-based targeting strategy. As a result, the present invention provides efficient delivery of GAA or GAA-like enzymes into target cells.

15

[0006] The invention relates, in part, to the discovery that GAA can be expressed recombinantly using a plurality of open reading frames encoding polypeptides representing different portions of the GAA protein. When provided together, the resulting polypeptides can cooperate to provide the desired enzymatic activity.

20

[0007] Accordingly, the present invention in one aspect relates to a nucleic acid sequence (such as a DNA sequence) encoding an open reading frame of a polypeptide including an amino acid sequence at least 50% identical (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or 100% identical) to amino acid residues 70-790 of human GAA or a fragment thereof. The open reading frame does not include an amino acid sequence at least 50% identical (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or 100% identical) to amino acid residues 880-952 of human GAA.

25

[0008] In another aspect, the invention relates to a nucleic acid sequence (such as a DNA sequence) encoding an open reading frame of a polypeptide including an amino acid sequence at least 50% identical (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or 100% identical) to amino acid residues 880-952 of human GAA or a fragment thereof. The open reading frame does not include an amino acid sequence at least 50% identical (e.g. at least

2005212435 23 Nov 2006

1 60%, at least 70%, at least 80%, at least 90%, at least 95%, or 100% identical) to amino acid
2 residues 70-790 of human GAA.

3 [0009] The invention also relates to cells containing one or both such nucleic acid
4 sequences.

5 [0010] In one embodiment, a nucleic acid of the invention also encodes a peptide tag fused
6 to the GAA polypeptide. A preferred peptide tag is a ligand for an extracellular receptor. In
7 some embodiments, a peptide tag is a targeting domain that binds an extracellular domain of a
8 receptor on the surface of a target cell and, upon internalization of the receptor, permits
9 localization of the polypeptide in a human lysosome. In one embodiment, the targeting domain
10 includes a urokinase-type plasminogen receptor moiety capable of binding the cation-
11 independent mannose-6-phosphate receptor. In another embodiment, the targeting domain
12 incorporates one or more amino acid sequences of IGF-II (*e.g.* at least amino acids 48-55; at
13 least amino acids 8-28 and 41-61; or at least amino acids 8-87) or a sequence variant thereof
14 (*e.g.* R68A) or truncated form thereof (*e.g.* C-terminally truncated from position 62) that binds
15 the cation-independent mannose-6-phosphate receptor. In one embodiment, a peptide tag is
16 fused directly to the N- or C-terminus of the GAA polypeptide. In another embodiment, a
17 peptide tag is fused to the N- or C-terminus of the GAA polypeptide by a spacer. In one specific
18 embodiment, a peptide tag is fused to the GAA polypeptide by a spacer of 10-25 amino acids.
19 In another specific embodiment, a peptide tag is fused to the GAA polypeptide by a spacer
20 including glycine residues. In another specific embodiment, a peptide tag is fused to the GAA
21 polypeptide by a spacer including a helical structure. In another specific embodiment, a peptide
22 tag is fused to the GAA polypeptide by a spacer at least 50% identical to the sequence
23 GGGTVGDDDDK (SEQ ID NO: 1).

24 [0011] The invention also relates to polypeptides encoded by the nucleic acids of the
25 invention and to pharmaceutical preparations incorporating those polypeptides.

26 [0012] The invention also relates, in part, to an appreciation of particular positions of GAA
27 to which a peptide tag can be fused. Accordingly, in one aspect the invention relates to a
28 targeted therapeutic including a peptide tag fused to amino acid 68, 69, 70, 71, 72, 779, 787,
29 789, 790, 791, 792, 793, or 796 of human GAA or a portion thereof. The targeted therapeutic
30 can include, for example, amino acid residues 70-952 of human GAA, or a smaller portion, such

1 as amino acid residues 70-790. In one embodiment, a peptide tag is fused to amino acid 70, or
2 to an amino acid within one or two positions of amino acid 70. In some embodiments, the
3 peptide tag is a ligand for an extracellular receptor. For example, some peptide tags are
4 targeting domains that bind an extracellular domain of a receptor on the surface of a target cell
5 and, upon internalization of the receptor, permit localization of the therapeutic agent to a human
6 lysosome. In one embodiment, the targeting domain includes a urokinase-type plasminogen
7 receptor moiety capable of binding the cation-independent mannose-6-phosphate receptor. In
8 another embodiment, the targeting domain incorporates one or more amino acid sequences of
9 IGF-II (*e.g.* at least amino acids 48-55; at least amino acids 8-28 and 41-61; or at least amino
10 acids 8-87) or a sequence variant thereof (*e.g.* R68A) or truncated form thereof (*e.g.* C-
11 terminally truncated from position 62) that binds the cation-independent mannose-6-phosphate
12 receptor. The peptide tag is fused to the GAA polypeptide directly or by a spacer. In one
13 specific embodiment, a peptide tag is fused to the GAA polypeptide by a spacer of 10-25 amino
14 acids. In another specific embodiment, a peptide tag is fused to the GAA polypeptide by a
15 spacer including glycine residues. In another specific embodiment, a peptide tag is fused to the
16 GAA polypeptide by a spacer including a helical structure. In another specific embodiment, a
17 peptide tag is fused to the GAA polypeptide by a spacer at least 50% identical to the sequence
18 GGGTVGDDDDK (SEQ ID NO: 1).

9 BRIEF DESCRIPTION OF THE DRAWINGS

- 0 [0013] Figures 1-1 to 1-14 depict an amino acid sequence alignment of selected members of
21 family 31 of glycoside hydrolyases.
- 22 [0014] Figure 2 is a schematic depiction of the GAA protein.
- 23 [0015] Figure 3 depicts exemplary strategies for creating a peptide-tagged GAA.
- 24 [0016] Figure 4 depicts an exemplary uptake experiment using wild-type GAA and SS-
25 GAA Δ 1-69.
- 26 [0017] Figure 5 depicts an exemplary uptake experiment using SS-GAA Δ 1-69 and SS-
27 GILT Δ 2-7-GAA Δ 1-69.
- 28 [0018] Figure 6 depicts an exemplary Western blot analysis of 1-87-IGF-II-tagged GAA
29 proteins: the left panel was probed with an anti-GAA antibody; the right panel was probed with

an anti-IGF-II antibody. Lane 1: pCEP-GILT1-87-GAA56-952; lane 2: pCEP-GILT1-87-R68A-GAA56-952-1; lane 3: pCEP-GILT1-87-R68A-ΔGS-GAA56-952-1; lane 4: pCEP-GILTA2-7-spcr1-GAA70-952-1; lane 5: pCEP-GAA; lane 6: pCEP-GILT-GAA29-952.

[0019] Figure 7 depicts an exemplary Western blot analysis comparing proteolysis of wild-type GAA with GAA-791A.sc.

[0020] Figure 8 depicts an exemplary Western analysis of wild-type GAA and GAA constructs with a GILT tag engineered with a downstream Factor X protease site, GAA787GILTxa, GAA779GILTxa, and GAA796GILTxa. It also depicts a GAA C-terminal processing model.

10

DETAILED DESCRIPTION OF THE INVENTION

[0021] This invention provides a means of producing GAA that is more effectively targeted to the lysosomes of mammalian cells, for example, human cardiac and skeletal muscle cells. GAA is a member of family 31 of glycoside hydrolyases (Figures 1-1 to 1-14). Human GAA is synthesized as a 110 kDal precursor (Wisselaar *et al.* (1993) J. Biol. Chem. 268(3): 2223-31). The mature form of the enzyme is a mixture of monomers of 70 and 76 kDal (Wisselaar *et al.* (1993) J. Biol. Chem. 268(3): 2223-31). The precursor enzyme has seven potential glycosylation sites and four of these are retained in the mature enzyme (Wisselaar *et al.* (1993) J. Biol. Chem. 268(3): 2223-31). The proteolytic cleavage events which produce the mature enzyme occur in late endosomes or in the lysosome (Wisselaar *et al.* (1993) J. Biol. Chem. 268(3): 2223-31).

[0022] The C-terminal 160 amino acids are absent from the mature 70 and 76 kDal species. However, certain Pompe alleles resulting in the complete loss of GAA activity map to this region, for example Val949Asp (Becker *et al.* (1998) J. Hum. Genet. 62:991). The phenotype of this mutant indicates that the C-terminal portion of the protein, although not part of the 70 or 76 kDal species, plays an important role in the function of the protein. It has recently been reported that the C-terminal portion of the protein, although cleaved from the rest of the protein during processing, remains associated with the major species (Moreland *et al.* (Nov. 1, 2004) J. Biol. Chem., Manuscript 404008200). Accordingly, the C-terminal residues could play a direct role in the catalytic activity of the protein. Alternatively, the C-terminal residues may be involved in promoting proper folding of the N-terminal portions of the protein.

30

[0023] This latter possibility is supported by the behavior of certain alleles of sucrase-isomaltase, a related protein. This family includes the sucrase-isomaltase (SI) protein which contains the two distinct but homologous glycoside hydrolyase catalytic domains in tandem on a single polypeptide. Each of these is similar to the entire GAA polypeptide in size and the two domains share 36 and 39% identity with GAA. SI is expressed in intestinal brush border cells and is localized to the apical membrane of these polarized cells with the catalytic domains facing the gut lumen due to an amino-terminal trans membrane domain. Once arriving at the apical membrane, the sucrase domain is cleaved from the amino-proximal isomaltase domain by trypsin while the isomaltase domain remains membrane associated. Recent studies indicate that the sucrase domain is required for proper folding and subsequent transport of the isomaltase domain; sucrase is said to be an intramolecular chaperone required for the folding of the isomaltase domain (Jacob *et al.* (2002) J. Biol. Chem. 277:32141).

[0024] Analysis of the expression of a number of engineered GAA cassettes has enabled the identification of two regions that, although cleaved from the mature polypeptide, are nevertheless required for secretion of functional protein from mammalian cells. Figure 2 summarizes the organization of GAA as we now understand it. The precursor polypeptide possesses a signal sequence and an adjacent putative trans-membrane domain, a trefoil domain (PFAM PF00088) which is a cysteine-rich domain of about 45 amino acids containing 3 disulfide linkages and thought to be involved in protein-protein or protein carbohydrate interactions (Thim (1989) FEBS Lett. 250:85), the domain defined by the mature 70/76 kDal polypeptide, and the C-terminal domain. A mutation in the trefoil domain of SI has an impact on the apical membrane sorting pattern of SI (Spodsberg (2001) J. Biol. Chem. 276:23506). Data presented in examples 1 and 2 indicate that both the trefoil domain and the C-terminal domain are required for the production of functional GAA. It is possible that the C-terminal domain interacts with the trefoil domain during protein folding perhaps facilitating appropriate disulfide bond formation in the trefoil domain.

[0025] In one embodiment of the invention, DNA sequence encoding a peptide tag is fused in frame to the 3' terminus of a GAA cassette that encodes the entire GAA polypeptide with the exception of the C-terminal domain. This cassette is co-expressed in mammalian cells that also express the C-terminal domain of GAA as a separate polypeptide.

[0026] The C-terminal domain then functions *in trans* in conjunction with the 70/76 kDal species to generate active GAA. The boundary between the catalytic domain and the C-terminal domain appears to be at about amino acid residuc 791, based on its presence in a short region of less than 18 amino acids that is absent from most members of the family 31 hydrolyases and which contains 4 consecutive proline residues in GAA. Indeed, it has now been reported that the C-terminal domain associated with the mature species begins at amino acid residue 792 (Moreland *et al.* (Nov. 1, 2004) J. Biol. Chem., Manuscript 404008200).

[0027] Co-expression can be achieved by driving expression of both polypeptides from one plasmid construct introduced into mammalian cells to produce a stable cell line. Expression can be driven by two promoters on such a plasmid or by one promoter driving expression of a bicistronic construct in which the two cassettes are separated by an IRES element. Alternatively, cell lines expressing both proteins can be constructed sequentially with separate plasmids employing distinct selecttable markers.

[0028] The peptide tag used in these fusions can be derived from IGF-II to target the CI-MPR. Alternatively, peptide tags that preferentially bind to receptors on the surface of myotubes can be employed. Such peptides have been described (Samoylova *et al.* (1999) Muscle and Nerve 22:460; US Patent No. 6,329,501). Other cell surface receptors, such as the Fc receptor, the LDL receptor, or the transferrin receptor are also appropriate targets and can promote targeting of GAA.

[0029] In another embodiment of the invention, the cassette encoding the peptide tag is inserted into the native GAA coding sequence at the junction of the mature 70/76 kDal polypeptide and the C-terminal domain, for example at position 791. This creates a single chimeric polypeptide. Because the peptide tag may be unable to bind to its cognate receptor in this configuration, a protease cleavage site may be inserted just downstream of the peptide tag. Once the protein is produced in correct folded form, the C-terminal domain can be cleaved by protease treatment.

[0030] It may be desirable to employ a protease cleavage site that is acted upon by a protease normally found in human serum. In this way, the tagged GAA can be introduced into the blood stream in a prodrug form and become activated for uptake by the serum resident

protease. This might improve the distribution of the enzyme. As before, the peptide tag could be the GILT tag or a muscle-specific tag.

5 [0031] In another embodiment of the invention, the tag is fused at the N-terminus of GAA in such a way as to retain enzymatic activity (Example 3). In the case of N-terminal fusions, it is possible to affect the level of secretion of the enzyme by substituting a heterologous signal peptide for the native GAA signal peptide.

10 [0032] The GAA signal peptide is not cleaved in the ER thereby causing GAA to be membrane bound in the ER (Tsuiji *et al.* (1987) Biochem. Int. 15(5):945-952). In some cell types, the enzyme can be found bound to the plasma membrane with retention of the membrane topology of the ER presumably due to the failure to cleave the signal peptide (Hirschhorn *et al.*, in The Metabolic and Molecular Basis of Inherited Disease, Valle, ed., 2001, McGraw-Hill: New York, pp. 3389-3420). Sequence analysis suggests the presence of a trans-membrane domain adjacent to the signal peptide, which presumably enables the enzyme to remain membrane attached under certain conditions.

15 [0033] It is possible that membrane association of GAA via its signal peptide is an important contributory factor in correct lysosomal targeting of the enzyme. This could happen in two ways: First, the membrane association could directly steer the protein to the lysosome. Second, the membrane association could increase the residence time of the GAA in the Golgi thereby increasing the level of mannose-6-phosphate added to the protein. This would have the net effect of increasing the proportion of the enzyme that is sorted to the lysosome. In either case, 20 if this membrane association were eliminated, then more of the produced enzyme would be secreted and if the latter model were correct, the secreted enzyme would have less mannose-6-phosphate.

25 [0034] Disruption of the membrane association of GAA can be accomplished by replacing the GAA signal peptide and adjacent sequence with an alternate signal peptide for GAA. In the context of GILT tagging of GAA, the chimeric gene contains the IGF-II tag including its signal peptide fused to the N-terminus of GAA at the native signal peptide cleavage site or at appropriate downstream sites. Such a chimeric fusion will direct the production of a recombinant GAA enzyme that is secreted at high levels and that contains a high affinity ligand 30 for the M6P/IGF-II receptor.

Subcellular targeting domains

[0035] The present invention permits targeting of a therapeutic agent to a lysosome using a protein, or an analog of a protein, that specifically binds a cellular receptor for that protein. The exterior of the cell surface is topologically equivalent to endosomal, lysosomal, golgi, and endoplasmic reticulum compartments. Thus, endocytosis of a molecule through interaction with an appropriate receptor(s) permits transport of the molecule to any of these compartments without crossing a membrane. Should a genetic deficiency result in a deficit of a particular enzyme activity in any of these compartments, delivery of a therapeutic protein can be achieved by tagging it with a ligand for the appropriate receptor(s).

[0036] Multiple pathways directing receptor-bound proteins from the plasma membrane to the golgi and/or endoplasmic reticulum have been characterized. Thus, by using a targeting portion from, for example, SV40, cholera toxin, or the plant toxin ricin, each of which coopt one or more of these subcellular trafficking pathways, a therapeutic can be targeted to the desired location within the cell. In each case, uptake is initiated by binding of the material to the exterior of the cell. For example, SV40 binds to MHC class I receptors, cholera toxin binds to GM1 ganglioside molecules and ricin binds to glycolipids and glycoproteins with terminal galactose on the surface of cells. Following this initial step the molecules reach the ER by a variety of pathways. For example, SV40 undergoes caveolar endocytosis and reaches the ER in a two step process that bypasses the golgi whereas cholera toxin undergoes caveolar endocytosis but traverses the golgi before reaching the ER.

[0037] If a targeting moiety related to cholera toxin or ricin is used, it is important that the toxicity of cholera toxin or ricin be avoided. Both cholera toxin and ricin are heteromeric proteins, and the cell surface binding domain and the catalytic activities responsible for toxicity reside on separate polypeptides. Thus, a targeting moiety can be constructed that includes the receptor-binding polypeptide, but not the polypeptide responsible for toxicity. For example, in the case of ricin, the B subunit possesses the galactose binding activity responsible for internalization of the protein, and can be fused to a therapeutic protein. If the further presence of the A subunit improves subcellular localization, a mutant version (mutein) of the A chain that is properly folded but catalytically inert can be provided with the B subunit-therapeutic agent fusion protein.

1 [0038] Proteins delivered to the golgi can be transported to the endoplasmic reticulum (ER)
2 via the KDEL receptor, which retrieves ER-targeted proteins that have escaped to the golgi.
3 Thus, inclusion of a KDEL (SEQ ID NO: 2) motif at the terminus of a targeting domain that
4 directs a therapeutic protein to the golgi permits subsequent localization to the ER. For
5 example, a targeting moiety (e.g. an antibody, or a peptide identified by high-throughput
6 screening such as phage display, yeast two hybrid, chip-based assays, and solution-based assays)
7 that binds the cation-independent M6P receptor both at or about pH 7.4 and at or about pH 5.5
8 permits targeting of a therapeutic agent to the golgi; further addition of a KDEL motif permits
9 targeting to the ER.

10 *Lysosomal targeting moieties*

11 [0039] The invention permits targeting of a therapeutic agent to a lysosome. Targeting may
12 occur, for example, through binding of a plasma membrane receptor that later passes through a
13 lysosome. Alternatively, targeting may occur through binding of a plasma receptor that later
14 passes through a late endosome; the therapeutic agent can then travel from the late endosome to
15 a lysosome. A preferred lysosomal targeting mechanism involves binding to the cation-
16 independent M6P receptor.

17 Cation-independent M6P receptor

18 [0040] The cation-independent M6P receptor is a 275 kDa single chain transmembrane
19 glycoprotein expressed ubiquitously in mammalian tissues. It is one of two mammalian
20 receptors that bind M6P: the second is referred to as the cation-dependent M6P receptor. The
21 cation-dependent M6P receptor requires divalent cations for M6P binding; the cation-
22 independent M6P receptor does not. These receptors play an important role in the trafficking of
23 lysosomal enzymes through recognition of the M6P moiety on high mannose carbohydrate on
24 lysosomal enzymes. The extracellular domain of the cation-independent M6P receptor contains
25 15 homologous domains ("repeats") that bind a diverse group of ligands at discrete locations on
26 the receptor.

27 [0041] The cation-independent M6P receptor contains two binding sites for M6P: one
28 located in repeats 1-3 and the other located in repeats 7-9. The receptor binds monovalent M6P
29 ligands with a dissociation constant in the μM range while binding divalent M6P ligands with a
30 dissociation constant in the nM range, probably due to receptor oligomerization. Uptake of

IGF-II by the receptor is enhanced by concomitant binding of multivalent M6P ligands such as lysosomal enzymes to the receptor.

[0042] The cation-independent M6P receptor also contains binding sites for at least three distinct ligands that can be used as targeting moieties. The cation-independent M6P receptor
5 binds IGF-II with a dissociation constant of about 14 nM at or about pH 7.4, primarily through interactions with repeat 11. Consistent with its function in targeting IGF-II to the lysosome, the dissociation constant is increased approximately 100-fold at or about pH 5.5 promoting dissociation of IGF-II in acidic late endosomes. The receptor is capable of binding high molecular weight O-glycosylated IGF-II forms.

10 [0043] An additional useful ligand for the cation-independent M6P receptor is retinoic acid. Retinoic acid binds to the receptor with a dissociation constant of 2.5 nM. Affinity photolabeling of the cation-independent M6P receptor with retinoic acid does not interfere with IGF-II or M6P binding to the receptor, indicating that retinoic acid binds to a distinct site on the receptor. Binding of retinoic acid to the receptor alters the intracellular distribution of the
15 receptor with a greater accumulation of the receptor in cytoplasmic vesicles and also enhances uptake of M6P modified β -glucuronidase. Retinoic acid has a photoactivatable moiety that can be used to link it to a therapeutic agent without interfering with its ability to bind to the cation-independent M6P receptor.

[0044] The cation-independent M6P receptor also binds the urokinase-type plasminogen
20 receptor (uPAR) with a dissociation constant of 9 μ M. uPAR is a GPI-anchored receptor on the surface of most cell types where it functions as an adhesion molecule and in the proteolytic activation of plasminogen and TGF- β . Binding of uPAR to the CI-M6P receptor targets it to the lysosome, thereby modulating its activity. Thus, fusing the extracellular domain of uPAR, or a portion thereof competent to bind the cation-independent M6P receptor, to a therapeutic agent
25 permits targeting of the agent to a lysosome.

IGF-II

[0045] In a preferred embodiment, the lysosomal targeting portion is a protein, peptide, or other moiety that binds the cation independent M6P/IGF-II receptor in a mannose-6-phosphate-independent manner. Advantageously, this embodiment mimics the normal biological

mechanism for uptake of LSD proteins, yet does so in a manner independent of mannose-6-phosphate.

[0046] For example, by fusing DNA encoding the mature IGF-II polypeptide to the 3' end of LSD gene cassettes, fusion proteins are created that can be taken up by a variety of cell types and transported to the lysosome. Alternatively, DNA encoding a precursor IGF-II polypeptide can be fused to the 3' end of an LSD gene cassette; the precursor includes a carboxyterminal portion that is cleaved in mammalian cells to yield the mature IGF-II polypeptide, but the IGF-II signal peptide is preferably omitted (or moved to the 5' end of the LSD gene cassette). This method has numerous advantages over methods involving glycosylation including simplicity and cost effectiveness, because once the protein is isolated, no further modifications need be made.

[0047] IGF-II is preferably targeted specifically to the M6P receptor. Particularly useful are mutations in the IGF-II polypeptide that result in a protein that binds the M6P receptor with high affinity while no longer binding the other two receptors with appreciable affinity. IGF-II can also be modified to minimize binding to serum IGF-binding proteins (Baxter (2000) Am. J. Physiol. Endocrinol. Metab. 278(6):967-76) to avoid sequestration of IGF-II/GILT constructs. A number of studies have localized residues in IGF-1 and IGF-II necessary for binding to IGF-binding proteins. Constructs with mutations at these residues can be screened for retention of high affinity binding to the M6P/IGF-II receptor and for reduced affinity for IGF-binding proteins. For example, replacing Phe 26 of IGF-II with Ser is reported to reduce affinity of IGF-II for IGFBP-1 and -6 with no effect on binding to the M6P/IGF-II receptor (Bach *et al.* (1993) J. Biol. Chem. 268(13):9246-54). Other substitutions, such as Ser for Phe 19 and Lys for Glu 9, can also be advantageous. The analogous mutations, separately or in combination, in a region of IGF-I that is highly conserved with IGF-II result in large decreases in IGF-BP binding (Magee *et al.* (1999) Biochemistry 38(48):15863-70).

[0048] An alternate approach is to identify minimal regions of IGF-II that can bind with high affinity to the M6P/IGF-II receptor. The residues that have been implicated in IGF-II binding to the M6P/IGF-II receptor mostly cluster on one face of IGF-II (Terasawa *et al.* (1994) EMBO J. 13(23):5590-7). Although IGF-II tertiary structure is normally maintained by three intramolecular disulfide bonds, a peptide incorporating the amino acid sequence on the

M6P/IGF-II receptor binding surface of IGF-II can be designed to fold properly and have binding activity. Such a minimal binding peptide is a highly preferred targeting portion. Designed peptides based on the region around amino acids 48-55 can be tested for binding to the M6P/IGF-II receptor. Alternatively, a random library of peptides can be screened for the ability to bind the M6P/IGF-II receptor either via a yeast two hybrid assay, or via a phage display type assay.

Blood-brain barrier

[0049] One challenge in therapy for lysosomal storage diseases is that many of these diseases have significant neurological involvement. Therapeutic enzymes administered into the blood stream generally do not cross the blood brain barrier and therefore cannot relieve neurological symptoms associated with the diseases. IGF-II, however, has been reported to promote transport across the blood-brain barrier via transcytosis (Bickel *et al.* (2001) Adv. Drug Deliv. Rev. 46(1-3):247-79). Thus, appropriately designed GILT constructs should be capable of crossing the blood brain barrier, affording for the first time a means of treating neurological symptoms associated with lysosomal storage diseases. The constructs can be tested using GUS minus mice as described in Example 12. Further details regarding design, construction and testing of targeted therapeutics that can reach neuronal tissue from blood are disclosed in U.S. Serial No. 60/329,650, filed October 16, 2001, and in U.S. Serial No. 10/136,639, filed April 30, 2002.

Structure of IGF-II

[0050] NMR structures of IGF-II have been solved by two groups (Terasawa *et al.* (1994) EMBO J. 13(23):5590-7; Torres *et al.* (1995) J. Mol. Biol. 248(2):385-401) (see, e.g., Protein Data Bank record 1IGL). The general features of the IGF-II structure are similar to IGF-I and insulin. The A and B domains of IGF-II correspond to the A and B chains of insulin. Secondary structural features include an alpha helix from residues 11-21 of the B region connected by a reverse turn in residues 22-25 to a short beta strand in residues 26-28. Residues 25-27 appear to form a small antiparallel beta sheet; residues 59-61 and residues 26-28 may also participate in intermolecular beta-sheet formation. In the A domain of IGF-II, alpha helices spanning residues 42-49 and 53-59 are arranged in an antiparallel configuration perpendicular to the B-domain helix. Hydrophobic clusters formed by two of the three disulfide bridges and

conserved hydrophobic residues stabilize these secondary structure features. The N and C termini remain poorly defined as is the region between residues 31-40.

[0051] IGF-II binds to the IGF-II/M6P and IGF-I receptors with relatively high affinity and binds with lower affinity to the insulin receptor. IGF-II also interacts with a number of serum IGF-BPs.

Binding to the IGF-II/M6P receptor

[0052] Substitution of IGF-II residues 48-50 (Phe Arg Ser) with the corresponding residues from insulin, (Thr Ser Ile), or substitution of residues 54-55 (Ala Leu) with the corresponding residues from IGF-I (Arg Arg) result in diminished binding to the IGF-II/ M6P receptor but retention of binding to the IGF-I and insulin receptors (Sakano *et al.* (1991) J. Biol. Chem. 266(31):20626-35).

[0053] IGF-I and IGF-II share identical sequences and structures in the region of residues 48-50 yet have a 1000-fold difference in affinity for the IGF-II receptor. The NMR structure reveals a structural difference between IGF-I and IGF-II in the region of IGF-II residues 53-58 (IGF-I residues 54-59): the alpha-helix is better defined in IGF-II than in IGF-I and, unlike IGF-I, there is no bend in the backbone around residues 53 and 54 (Torres *et al.* (1995) J. Mol. Biol. 248(2):385-401). This structural difference correlates with the substitution of Ala 54 and Leu 55 in IGF-II with Arg 55 and Arg 56 in IGF-I. It is possible either that binding to the IGF-II receptor is disrupted directly by the presence of charged residues in this region or that changes in the structure engendered by the charged residues yield the changes in binding for the IGF-II receptor. In any case, substitution of uncharged residues for the two Arg residues in IGF-I resulted in higher affinities for the IGF-II receptor (Cacciari *et al.* (1987) Pediatrician 14(3):146-53). Thus the presence of positively charged residues in these positions correlates with loss of binding to the IGF-II receptor.

[0054] IGF-II binds to repeat 11 of the cation-independent M6P receptor. Indeed, a minireceptor in which only repeat 11 is fused to the transmembrane and cytoplasmic domains of the cation-independent M6P receptor is capable of binding IGF-II (with an affinity approximately one tenth the affinity of the full length receptor) and mediating internalization of IGF-II and its delivery to lysosomes (Grimme *et al.* (2000) J. Biol. Chem. 275(43):33697-33703). The structure of domain 11 of the M6P receptor is known (Protein Data Base entries

IGP0 and IGP3; Brown *et al.* (2002) EMBO J. 21(5):1054-1062). The putative IGF-II binding site is a hydrophobic pocket believed to interact with hydrophobic amino acids of IGF-II; candidate amino acids of IGF-II include leucine 8, phenylalanine 48, alanine 54, and leucine 55. Although repeat 11 is sufficient for IGF-II binding, constructs including larger portions of the cation-independent M6P receptor (*e.g.* repeats 10-13, or 1-15) generally bind IGF-II with greater affinity and with increased pH dependence (see, for example, Linnell *et al.* (2001) J. Biol. Chem. 276(26):23986-23991).

Binding to the IGF-I receptor

[0055] Substitution of IGF-II residues Tyr 27 with Leu, Leu 43 with Val or Ser 26 with Phe diminishes the affinity of IGF-II for the IGF-I receptor by 94-, 56-, and 4-fold respectively (Torres *et al.* (1995) J. Mol. Biol. 248(2):385-401). Deletion of residues 1-7 of human IGF-II resulted in a 30-fold decrease in affinity for the human IGF-I receptor and a concomitant 12 fold increase in affinity for the rat IGF-II receptor (Hashimoto *et al.* (1995) J. Biol. Chem. 270(30):18013-8). The NMR structure of IGF-II shows that Thr 7 is located near residues 48 Phe and 50 Ser as well as near the 9 Cys-47 Cys disulfide bridge. It is thought that interaction of Thr 7 with these residues can stabilize the flexible N-terminal hexapeptide required for IGF-I receptor binding (Terasawa *et al.* (1994) EMBO J. 13(23):5590-7). At the same time this interaction can modulate binding to the IGF-II receptor. Truncation of the C-terminus of IGF-II (residues 62-67) also appear to lower the affinity of IGF-II for the IGF-I receptor by 5 fold (Roth *et al.* (1991) Biochem. Biophys. Res. Commun. 181(2):907-14).

Deletion mutants of IGF-II

[0056] The binding surfaces for the IGF-I and cation-independent M6P receptors are on separate faces of IGF-II. Based on structural and mutational data, functional cation-independent M6P binding domains can be constructed that are substantially smaller than human IGF-II. For example, the amino terminal amino acids 1-7 and/or the carboxy terminal residues 62-67 can be deleted or replaced. Additionally, amino acids 29-40 can likely be eliminated or replaced without altering the folding of the remainder of the polypeptide or binding to the cation-independent M6P receptor. Thus, a targeting moiety including amino acids 8-28 and 41-61 can be constructed. These stretches of amino acids could perhaps be joined directly or separated by a linker. Alternatively, amino acids 8-28 and 41-61 can be provided on separate polypeptide chains. Comparable domains of insulin, which is homologous to IGF-II and has a tertiary

structure closely related to the structure of IGF-II, have sufficient structural information to permit proper refolding into the appropriate tertiary structure, even when present in separate polypeptide chains (Wang *et al.* (1991) Trends Biochem. Sci. 279-281). Thus, for example, amino acids 8-28, or a conservative substitution variant thereof, could be fused to a therapeutic agent; the resulting fusion protein could be admixed with amino acids 41-61, or a conservative substitution variant thereof, and administered to a patient.

[0057] In order to facilitate proper presentation and folding of the IGF-II tag, longer portions of IGF-II proteins can be used. For example, an IGF-II tag including amino acid residues 1-67, 1-87, or the entire precursor form can be used.

10 Binding to IGF Binding proteins

[0058] IGF-II and related constructs can be modified to diminish their affinity for IGFBPs, thereby increasing the bioavailability of the tagged proteins.

[0059] Substitution of IGF-II residue phenylalanine 26 with serine reduces binding to IGFBPs 1-5 by 5-75 fold (Bach *et al.* (1993) J. Biol. Chem. 268(13):9246-54). Replacement of IGF-II residues 48-50 with threonine-serine-isoleucine reduces binding by more than 100 fold to most of the IGFBPs (Bach *et al.* (1993) J. Biol. Chem. 268(13):9246-54); these residues are, however, also important for binding to the cation-independent mannose-6-phosphate receptor. The Y27L substitution that disrupts binding to the IGF-I receptor interferes with formation of the ternary complex with IGFBP3 and acid labile subunit (Hashimoto *et al.* (1997) J. Biol. Chem. 272(44):27936-42); this ternary complex accounts for most of the IGF-II in the circulation (Yu *et al.* (1999) J. Clin. Lab Anal. 13(4):166-72). Deletion of the first six residues of IGF-II also interferes with IGFBP binding (Luthi *et al.* (1992) Eur. J. Biochem. 205(2):483-90).

[0060] Studies on IGF-I interaction with IGFBPs revealed additionally that substitution of serine for phenylalanine 16 did not effect secondary structure but decreased IGFBP binding by between 40 and 300 fold (Magee *et al.* (1999) Biochemistry 38(48):15863-70). Changing glutamate 9 to lysine also resulted in a significant decrease in IGFBP binding. Furthermore, the double mutant lysine 9/ serine 16 exhibited the lowest affinity for IGFBPs. Although these mutations have not previously been tested in IGF-II, the conservation of sequence between this

region of IGF-I and IGF-II suggests that a similar effect will be observed when the analogous mutations are made in IGF-II (glutamate 12 lysine/ phenylalanine 19 serine).

IGF-II homologs

- 5 [0061] The amino acid sequence of human IGF-II, or a portion thereof affecting binding to the cation-independent M6P receptor, may be used as a reference sequence to determine whether a candidate sequence possesses sufficient amino acid similarity to have a reasonable expectation of success in the methods of the present invention. Preferably, variant sequences are at least 70% similar or 60% identical, more preferably at least 75% similar or 65% identical, and most preferably 80% similar or 70% identical to human IGF-II.
- 10 [0062] To determine whether a candidate peptide region has the requisite percentage similarity or identity to human IGF-II, the candidate amino acid sequence and human IGF-II are first aligned using the dynamic programming algorithm described in Smith and Waterman (1981) *J. Mol. Biol.* 147:195-197, in combination with the BLOSUM62 substitution matrix described in Figure 2 of Henikoff and Henikoff (1992) *PNAS* 89:10915-10919. For the present invention, an appropriate value for the gap insertion penalty is -12, and an appropriate value for the gap extension penalty is -4. Computer programs performing alignments using the algorithm of Smith-Waterman and the BLOSUM62 matrix, such as the GCG program suite (Oxford Molecular Group, Oxford, England), are commercially available and widely used by those skilled in the art.
- 15 20 [0063] Once the alignment between the candidate and reference sequence is made, a percent similarity score may be calculated. The individual amino acids of each sequence are compared sequentially according to their similarity to each other. If the value in the BLOSUM62 matrix corresponding to the two aligned amino acids is zero or a negative number, the pairwise similarity score is zero; otherwise the pairwise similarity score is 1.0. The raw similarity score is the sum of the pairwise similarity scores of the aligned amino acids. The raw score is then normalized by dividing it by the number of amino acids in the smaller of the candidate or reference sequences. The normalized raw score is the percent similarity. Alternatively, to calculate a percent identity, the aligned amino acids of each sequence are again compared sequentially. If the amino acids are non-identical, the pairwise identity score is zero; otherwise the pairwise identity score is 1.0. The raw identity score is the sum of the identical aligned
- 25 30

amino acids. The raw score is then normalized by dividing it by the number of amino acids in the smaller of the candidate or reference sequences. The normalized raw score is the percent identity. Insertions and deletions are ignored for the purposes of calculating percent similarity and identity. Accordingly, gap penalties are not used in this calculation, although they are used
5 in the initial alignment.

IGF-II structural analogs

[0064] The known structures of human IGF-II and the cation-independent M6P receptors permit the design of IGF-II analogs and other cation-independent M6P receptor binding proteins using computer-assisted design principles such as those discussed in U.S. Patent Nos. 6,226,603
10 and 6,273,598. For example, the known atomic coordinates of IGF-II can be provided to a computer equipped with a conventional computer modeling program, such as INSIGHTII, DISCOVER, or DELPHI, commercially available from Biosym, Technologies Inc., or QUANTA, or CHARMM, commercially available from Molecular Simulations, Inc. These and other software programs allow analysis of molecular structures and simulations that predict the
15 effect of molecular changes on structure and on intermolecular interactions. For example, the software can be used to identify modified analogs with the ability to form additional intermolecular hydrogen or ionic bonds, improving the affinity of the analog for the target receptor.

[0065] The software also permits the design of peptides and organic molecules with
20 structural and chemical features that mimic the same features displayed on at least part of the surface of the cation-independent M6P receptor binding face of IGF-II. Because a major contribution to the receptor binding surface is the spatial arrangement of chemically interactive moieties present within the sidechains of amino acids which together define the receptor binding surface, a preferred embodiment of the present invention relates to designing and producing a
25 synthetic organic molecule having a framework that carries chemically interactive moieties in a spatial relationship that mimics the spatial relationship of the chemical moieties disposed on the amino acid sidechains which constitute the cation-independent M6P receptor binding face of IGF-II. Preferred chemical moieties, include but are not limited to, the chemical moieties defined by the amino acid side chains of amino acids constituting the cation-independent M6P
30 receptor binding face of IGF-II. It is understood, therefore, that the receptor binding surface of

1 the IGF-II analog need not comprise amino acid residues but the chemical moieties disposed
2 thereon.

3 [0066] For example, upon identification of relevant chemical groups, the skilled artisan
4 using a conventional computer program can design a small molecule having the receptor
5 interactive chemical moieties disposed upon a suitable carrier framework. Useful computer
6 programs are described in, for example, Dixon (1992) *Tibtech* 10: 357-363; Tschinke et al.
7 (1993) *J. Med. Chem* 36: 3863-3870; and Eisen et al. (1994) *Proteins: Structure, Function, and*
8 *Genetics* 19: 199-221, the disclosures of which are incorporated herein by reference.

9 [0067] One particular computer program entitled "CAVEAT" searches a database, for
0 example, the Cambridge Structural Database, for structures which have desired spatial
1 orientations of chemical moieties (Bartlett et al. (1989) in "Molecular Recognition: Chemical
2 and Biological Problems" (Roberts, S. M., ed) pp 182-196). The CAVEAT program has been
3 used to design analogs of tendamistat, a 74 residue inhibitor of α -amylase, based on the
4 orientation of selected amino acid side chains in the three-dimensional structure of tendamistat
5 (Bartlett et al. (1989) *supra*).

6 [0068] Alternatively, upon identification of a series of analogs which mimic the cation-
7 independent M6P receptor binding activity of IGF-II, the skilled artisan may use a variety of
8 computer programs which assist the skilled artisan to develop quantitative structure activity
9 relationships (QSAR) and further to assist in the de novo design of additional morphogen
10 analogs. Other useful computer programs are described in, for example, Connolly-Martin (1991)
11 *Methods in Enzymology* 203:587-613; Dixon (1992) *supra*; and Waszkowycz et al. (1994) *J.*
12 *Med. Chem.* 37: 3994-4002.

23 *Fusion junctions*

24 [0069] Where GAA is expressed as a fusion protein with a peptide tag or targeting domain,
25 the peptide tag can be fused directly to the GAA polypeptide or can be separated from the GAA
26 polypeptide by a linker. An amino acid linker incorporates an amino acid sequence other than
27 that appearing at that position in the natural protein and is generally designed to be flexible or to
28 interpose a structure, such as an α -helix, between the two protein moieties. A linker can be
29 relatively short, such as the sequence Gly-Ala-Pro or Gly-Gly-Gly-Gly-Gly-Pro (SEQ ID NO:
30 3), or can be longer, such as, for example, 10-25 amino acids in length. For example, flexible

1 repeating linkers of 3-4 copies of the sequence (GGGGS (SEQ ID NO: 4) and α -helical
2 repeating linkers of 2-5 copies of the sequence (EAAAK (SEQ ID NO: 5)) have been described
3 (Arai *et al.* (2004) Proteins:Structure, Function and Bioinformatics 57:829-838). The use of
4 another linker, GGGGTVGDDDDK (SEQ ID NO:1), in the context of an IGF-II fusion protein
5 has also been reported (DiFalco *et al.* (1997) Biochem. J. 326:407-413). Linkers incorporating
6 an α -helical portion of a human serum protein can be used to minimize immunogenicity of the
7 linker region.

8 [0070] The site of a fusion junction should be selected with care to promote proper folding
9 and activity of both fusion partners and to prevent premature separation of a peptide tag from a
0 GAA polypeptide. Figure 3 illustrates four exemplary strategies for creating a GILT-tagged
1 GAA, based on the model for the organization of GAA protein as illustrated in Figure 2.

- 2 1. Fusion of the tag at the amino terminus.
- 3 2. Insertion of the tag between the trefoil domain and the mature region.
- 4 3. Insertion of the tag between the mature region and the C-terminal domain.
- 5 4. Fusion of the tag to the C-terminus of a truncated GAA and co-expressing the C-terminal
6 domain.

7 For example, a targeting domain can be fused, directly or by a spacer, to amino acid 70 of GAA,
8 a position permitting expression of the protein, catalytic activity of the GAA moiety, and proper
9 targeting by the targeting moiety as described in Example 4. Alternatively, a targeting domain
10 can be fused at or near the cleavage site separating the C-terminal domain of GAA from the
11 mature polypeptide. This permits synthesis of a GAA protein with an internal targeting domain,
12 which optionally can be cleaved to liberate the mature polypeptide or the C-terminal domain
13 from the targeting domain, depending on placement of cleavage sites. Alternatively, the mature
14 polypeptide can be synthesized as a fusion protein at about position 791 without incorporating
15 C-terminal sequences in the open reading frame of the expression construct.

16 [0071] In order to facilitate folding of the GILT tag, GAA amino acid residues adjacent to
17 the fusion junction can be modified. For example, since it is possible that GAA cystine residues
18 may interfere with proper folding of the GILT tag, the terminal GAA cystine 952 can be deleted
19 or substituted with serine to accommodate a C-terminal GILT tag. The GILT tag can also be
20 fused immediately preceding the final Cys952. The penultimate cys938 can be changed to
21 proline in conjunction with a mutation of the final Cys952 to serine.
22
23
24
25
26
27
28
29
30
31

[0072] Alternatively, a tag can be chemically coupled to a GAA polypeptide.

Targeting moiety affinities

[0073] Preferred targeting moieties bind to their target receptors with a submicromolar dissociation constant. Generally speaking, lower dissociation constants (*e.g.* less than 10^{-7} M, less than 10^{-8} M, or less than 10^{-9} M) are increasingly preferred. Determination of dissociation constants is preferably determined by surface plasmon resonance as described in Linnell *et al.* (2001) *J. Biol. Chem.* 276(26):23986-23991. A soluble form of the extracellular domain of the target receptor (*e.g.* repeats 1-15 of the cation-independent M6P receptor) is generated and immobilized to a chip through an avidin-biotin interaction. The targeting moiety is passed over the chip, and kinetic and equilibrium constants are detected and calculated by measuring changes in mass associated with the chip surface.

Computation of sequence similarity

[0074] In order to produce variants of the disclosed sequences that may also serve as catalytic domain, chaperone domain or subcellular targeting domain, any one or more of the naturally-occurring alpha-glucosidases or subcellular targeting domain, such as, for example, IGF-II, disclosed herein may be used as a reference sequence to determine whether a candidate sequence possesses sufficient amino acid similarity to have a reasonable expectation of success in the methods of the present invention. For example, variant sequences of a catalytic domain are at least 50% similar or 30% identical, preferably at least 55% similar or 35% identical, more preferably at least 60% similar or 40% identical, more preferably at least 65% similar or 45% identical, more preferably at least 70% similar or 50% identical, more preferably at least 75% similar or 55% identical, more preferably at least 80% similar or 60% identical, more preferably at least 85% similar or 65% identical, more preferably at least 90% similar or 70% identical, more preferably at least 95% similar or 75% identical, and most preferably 80% identical, 85% identical, 90% identical, or 95% identical to one of the disclosed, naturally-occurring catalytic domain of acid alpha-glucosidase. Variant sequences of a chaperone domain are at least 40% similar or 20% identical, preferably at least 45% similar or 25% identical, more preferably at least 50% similar or 30% identical, more preferably at least 55% similar or 35% identical, more preferably at least 60% similar or 40% identical, more preferably at least 65% similar or 45% identical, more preferably at least 70% similar or 50% identical, more preferably at least 75% similar or 55% identical, more preferably at least 80% similar or 60% identical, more preferably

at least 85% similar or 65% identical, more preferably at least 90% similar or 70% identical, more preferably at least 95% similar or 75% identical, and most preferably 80% identical, 85% identical, 90% identical, or 95% identical to one of the disclosed, naturally-occurring chaperone domain of acid alpha-glucosidase. Variant sequences of a targeting domain are at least 70% similar or 60% identical, more preferably at least 75% similar or 65% identical, more preferably 80% similar or 70% identical, more preferably 85% similar or 75% identical, more preferably 90% similar or 80% identical, more preferably 95% similar or 85% identical, and most preferably, 90% identical, or 95% identical to one of the disclosed, naturally-occurring targeting domain.

10 [0075] To determine whether a candidate peptide region has the requisite percentage similarity or identity to a reference polypeptide or peptide oligomer, the candidate amino acid sequence and the reference amino acid sequence are first aligned using the dynamic programming algorithm described in Smith and Waterman (1981), *J. Mol. Biol.* 147:195-197, in combination with the BLOSUM62 substitution matrix described in Figure 2 of Henikoff and
15 Henikoff (1992), "Amino acid substitution matrices from protein blocks", *PNAS* (1992 Nov), 89:10915-10919. For the present invention, an appropriate value for the gap insertion penalty is -12, and an appropriate value for the gap extension penalty is -4. Computer programs performing alignments using the algorithm of Smith-Waterman and the BLOSUM62 matrix, such as the GCG program suite (Oxford Molecular Group, Oxford, England), are commercially
20 available and widely used by those skilled in the art.

[0076] Once the alignment between the candidate and reference sequence is made, a percent similarity score may be calculated. The individual amino acids of each sequence are compared sequentially according to their similarity to each other. If the value in the BLOSUM62 matrix corresponding to the two aligned amino acids is zero or a negative number, the pairwise
25 similarity score is zero; otherwise the pairwise similarity score is 1.0. The raw similarity score is the sum of the pairwise similarity scores of the aligned amino acids. The raw score is then normalized by dividing it by the number of amino acids in the smaller of the candidate or reference sequences. The normalized raw score is the percent similarity. Alternatively, to calculate a percent identity, the aligned amino acids of each sequence are again compared
30 sequentially. If the amino acids are non-identical, the pairwise identity score is zero; otherwise the pairwise identity score is 1.0. The raw identity score is the sum of the identical aligned

amino acids. The raw score is then normalized by dividing it by the number of amino acids in the smaller of the candidate or reference sequences. The normalized raw score is the percent identity. Insertions and deletions are ignored for the purposes of calculating percent similarity and identity. Accordingly, gap penalties are not used in this calculation, although they are used
5 in the initial alignment.

Administration

[0077] The targeted therapeutics produced according to the present invention can be administered to a mammalian host by any route. Thus, as appropriate, administration can be oral or parenteral, including intravenous and intraperitoneal routes of administration. In
10 addition, administration can be by periodic injections of a bolus of the therapeutic or can be made more continuous by intravenous or intraperitoneal administration from a reservoir which is external (e.g., an i.v. bag). In certain embodiments, the therapeutics of the instant invention can be pharmaceutical-grade. That is, certain embodiments comply with standards of purity and quality control required for administration to humans. Veterinary applications are also within
15 the intended meaning as used herein.

[0078] The formulations, both for veterinary and for human medical use, of the therapeutics according to the present invention typically include such therapeutics in association with a pharmaceutically acceptable carrier therefor and optionally other ingredient(s). The carrier(s) can be "acceptable" in the sense of being compatible with the other ingredients of the
20 formulations and not deleterious to the recipient thereof. Pharmaceutically acceptable carriers, in this regard, are intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is known in the art. Except insofar as any conventional
25 media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds (identified according to the invention and/or known in the art) also can be incorporated into the compositions. The formulations can conveniently be presented in dosage unit form and can be prepared by any of the methods well known in the art of pharmacy/microbiology. In general, some formulations are prepared by
30 bringing the therapeutic into association with a liquid carrier or a finely divided solid carrier or both, and then, if necessary, shaping the product into the desired formulation.

[0079] A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include oral or parenteral, e.g., intravenous, intradermal, inhalation, transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide.

[0080] Useful solutions for oral or parenteral administration can be prepared by any of the methods well known in the pharmaceutical art, described, for example, in Remington's Pharmaceutical Sciences, (Gennaro, A., ed.), Mack Pub., 1990. Formulations for parenteral administration also can include glycocholate for buccal administration, methoxysalicylate for rectal administration, or cutric acid for vaginal administration. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic. Suppositories for rectal administration also can be prepared by mixing the drug with a non-irritating excipient such as cocoa butter, other glycerides, or other compositions that are solid at room temperature and liquid at body temperatures. Formulations also can include, for example, polyalkylene glycols such as polyethylene glycol, oils of vegetable origin, hydrogenated naphthalenes, and the like. Formulations for direct administration can include glycerol and other compositions of high viscosity. Other potentially useful parenteral carriers for these therapeutics include ethylene-vinyl acetate copolymer particles, osmotic pumps, implantable infusion systems, and liposomes. Formulations for inhalation administration can contain as excipients, for example, lactose, or can be aqueous solutions containing, for example, polyoxyethylene-9-lauryl ether, glycocholate and deoxycholate, or oily solutions for administration in the form of nasal drops, or as a gel to be applied intranasally. Retention enemas also can be used for rectal delivery.

[0081] Formulations of the present invention suitable for oral administration can be in the form of discrete units such as capsules, gelatin capsules, sachets, tablets, troches, or lozenges,

- each containing a predetermined amount of the drug; in the form of a powder or granules; in the form of a solution or a suspension in an aqueous liquid or non-aqueous liquid; or in the form of an oil-in-water emulsion or a water-in-oil emulsion. The therapeutic can also be administered in the form of a bolus, electuary or paste. A tablet can be made by compressing or moulding the
- 5 drug optionally with one or more accessory ingredients. Compressed tablets can be prepared by compressing, in a suitable machine, the drug in a free-flowing form such as a powder or granules, optionally mixed by a binder, lubricant, inert diluent, surface active or dispersing agent. Molded tablets can be made by molding, in a suitable machine, a mixture of the powdered drug and suitable carrier moistened with an inert liquid diluent.
- 10 **[0082]** Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients. Oral compositions prepared using a fluid carrier for use as a mouthwash include the compound in the fluid carrier and are applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part
- 15 of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose; a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or
- 20 saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.
- [0083]** Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor ELTM (BASF,
- 25 Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition can be sterile and can be fluid to the extent that easy syringability exists. It can be stable under the conditions of manufacture and storage and can 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 (for example, glycerol, propylene glycol, and
- 30 liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can 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. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, and sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[0084] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a 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, methods of preparation include vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0085] Formulations suitable for intra-articular administration can be in the form of a sterile aqueous preparation of the therapeutic which can be in microcrystalline form, for example, in the form of an aqueous microcrystalline suspension. Liposomal formulations or biodegradable polymer systems can also be used to present the therapeutic for both intra-articular and ophthalmic administration.

[0086] Formulations suitable for topical administration, including eye treatment, include liquid or semi-liquid preparations such as liniments, lotions, gels, applicants, oil-in-water or water-in-oil emulsions such as creams, ointments or pastes; or solutions or suspensions such as drops. Formulations for topical administration to the skin surface can be prepared by dispersing the therapeutic with a dermatologically acceptable carrier such as a lotion, cream, ointment or soap. In some embodiments, useful are carriers capable of forming a film or layer over the skin to localize application and inhibit removal. Where adhesion to a tissue surface is desired the composition can include the therapeutic dispersed in a fibrinogen-thrombin composition or other bioadhesive. The therapeutic then can be painted, sprayed or otherwise applied to the desired tissue surface. For topical administration to internal tissue surfaces, the agent can be dispersed

in a liquid tissue adhesive or other substance known to enhance adsorption to a tissue surface. For example, hydroxypropylcellulose or fibrinogen/thrombin solutions can be used to advantage. Alternatively, tissue-coating solutions, such as pectin-containing formulations can be used.

- 5 [0087] For inhalation treatments, such as for asthma, inhalation of powder (self-propelling or spray formulations) dispensed with a spray can, a nebulizer, or an atomizer can be used. Such formulations can be in the form of a finely comminuted powder for pulmonary administration from a powder inhalation device or self-propelling powder-dispensing formulations. In the case of self-propelling solution and spray formulations, the effect can be
10 achieved either by choice of a valve having the desired spray characteristics (i.e., being capable of producing a spray having the desired particle size) or by incorporating the active ingredient as a suspended powder in controlled particle size. For administration by inhalation, the therapeutics also can be delivered in the form of an aerosol spray from a pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.
15 Nasal drops also can be used.

- [0088] Systemic administration also can be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants generally are known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and folic acid
20 derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the therapeutics typically are formulated into ointments, salves, gels, or creams as generally known in the art.

- [0089] In one embodiment, the therapeutics are prepared with carriers that will protect against rapid elimination from the body, such as a controlled release formulation, including
25 implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials also can be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions can also be used as
30 pharmaceutically acceptable carriers. These can be prepared according to methods known to

those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811. Microsomes and microparticles also can be used.

[0090] Oral or parenteral compositions can be formulated in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form refers to physically discrete units
5 suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active
10 compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

[0091] Generally, the therapeutics identified according to the invention can be formulated for parenteral or oral administration to humans or other mammals, for example, in therapeutically effective amounts, e.g., amounts which provide appropriate concentrations of the drug to target tissue for a time sufficient to induce the desired effect. Additionally, the
15 therapeutics of the present invention can be administered alone or in combination with other molecules known to have a beneficial effect on the particular disease or indication of interest. By way of example only, useful cofactors include symptom-alleviating cofactors, including antiseptics, antibiotics, antiviral and antifungal agents and analgesics and anesthetics.

[0092] The effective concentration of the therapeutics identified according to the invention
20 that is to be delivered in a therapeutic composition will vary depending upon a number of factors, including the final desired dosage of the drug to be administered and the route of administration. The preferred dosage to be administered also is likely to depend on such variables as the type and extent of disease or indication to be treated, the overall health status of the particular patient, the relative biological efficacy of the therapeutic delivered, the
25 formulation of the therapeutic, the presence and types of excipients in the formulation, and the route of administration. In some embodiments, the therapeutics of this invention can be provided to an individual using typical dose units deduced from the earlier-described mammalian studies using non-human primates and rodents. As described above, a dosage unit
30 refers to a unitary, i.e. a single dose which is capable of being administered to a patient, and which can be readily handled and packed, remaining as a physically and biologically stable unit

1 dose comprising either the therapeutic as such or a mixture of it with solid or liquid
2 pharmaceutical diluents or carriers.

3 [0093] In certain embodiments, organisms are engineered to produce the therapeutics
4 identified according to the invention. These organisms can release the therapeutic for harvesting
5 or can be introduced directly to a patient. In another series of embodiments, cells can be utilized
6 to serve as a carrier of the therapeutics identified according to the invention.

7 [0094] Therapeutics of the invention also include the "prodrug" derivatives. The term
8 prodrug refers to a pharmacologically inactive (or partially inactive) derivative of a parent
9 molecule that requires biotransformation, either spontaneous or enzymatic, within the organism
0 to release or activate the active component. Prodrugs are variations or derivatives of the
1 therapeutics of the invention which have groups cleavable under metabolic conditions. Prodrugs
2 become the therapeutics of the invention which are pharmaceutically active *in vivo*, when they
3 undergo solvolysis under physiological conditions or undergo enzymatic degradation. Prodrug
4 of this invention can be called single, double, triple, and so on, depending on the number of
5 biotransformation steps required to release or activate the active drug component within the
6 organism, and indicating the number of functionalities present in a precursor-type form.
7 Prodrug forms often offer advantages of solubility, tissue compatibility, or delayed release in the
8 mammalian organism (see, Bundgard, Design of Prodrugs, pp. 7-9, 21-24, Elsevier, Amsterdam
9 1985 and Silverman, The Organic Chemistry of Drug Design and Drug Action, pp. 352-401,
20 Academic Press, San Diego, Calif., 1992). Moreover, the prodrug derivatives according to this
21 invention can be combined with other features to enhance bioavailability.

22 **Examples**

23 Example 1: *Trans* expression of GAA.

24 [0095] The following primers were used to generate a gene cassette containing the human
25 IGF-II signal sequence fused to human GAA residues 791-952 (the C-terminal domain).
26 GAA41: GGAATTCAGGCGCGCCGGCAGCTCCCCGTGAGCCAGCC (SEQ ID NO: 6)
27 GAA 27: GCTCTAGACTAACACCAGCTGACGAGAAACTGC (SEQ ID NO: 7)
28 GAA41 and GAA27 were used to amplify the C-terminal domain of GAA by PCR. The
29 amplified fragment contains an Asc I site at the 5' terminus. The SS N-tag encoding the IGF-II
30 signal sequence (residues 1-25) with an AscI site at the 3' end was then fused at the Asc I site to

2005212435 23 Nov 2006

1 the GAA C-terminal domain and the cassette was cloned in pCEP4 to generate plasmid pCEP-
2 SS-GAA-791-952. The SS N-tag nucleic acid sequence is shown as below.

3 DNA sequence of the SS N-tag (SEQ ID NO: 8):
4 gaattcACACCAATGGGAATCCCAATGGGGAAGTCGATGCTGGTGTCTTCTCACCTTCTT
5 GGCCTTCGCCTCGTGCTGCATTGCTGCTggcgcgcgcg

6 [0096] The following additional plasmids were generated similarly: pCEP-GAA Δ 817-952
7 that lacks C-terminal GAA residues 817-952 and pCEP-GAA Δ 817-952-GILT Δ 1-7 that is
8 similar to pCEP-GAA Δ 817-952 except for the addition of a C-terminal GILTΔ1-7 tag. GAA Δ
9 817-952 was generated by introducing a stop codon after amino acid residue 816. To facilitate
0 the cloning process, the stop codon was followed by 3' end XbaI restriction site and 5' end
1 contains an EcoRI restriction site. DNA and amino acid sequences of GAA Δ 817-952 are
2 shown below.

3 DNA sequence of GAA Δ 817-952 (SEQ ID NO: 9).
4 gaattcCAAACCATGGGAGTGAGGCACCCGCCCTGCTCCCACCGGCTCCTGGCCGTCTG
5 CGCCCTCGTGTCTTGGCAACCGCTGCACTCCTGGGGCACATCCTACTCCATGATTT
6 CCTGTGGTTCCCCGAGAGCTGAGTGGCTCCTCCCAGTCTGGAGGAGACTCACCC
7 AGCTACCAGCAGGGAGCCAGCAGACCAGGGCCCCGGGATGCCAGGCACACCCC
8 GGCCGTCCCAGAGCAGTGCCACACAGTGCAGCTCCCCCAACAGCCGCTTCGA
9 TTGCGCCCCTGACAAGGCCATCACCCAGGAACAGTGCAGGCCCCGCGGCTGTGCT
10 ACATCCCTGCAAAGCAGGGGCTGCAGGGAGCCCAGATGGGGCAGCCCTGGTGTCTC
11 TTCCCACCCAGTACCCAGCTACAAGCTGGAGAACCCTGAGCTCCTCTGAAATGGG
12 CTACACGGCCACCCTGACCCGTACCACCCACCTTCTTCCCAAGGACATCCTGAC
13 CCTGCGGCTGGACGTGATGATGGAGACTGAGAACCGCCTCCACTTCACGATCAAAG
14 ATCCAGCTAACAGGCGCTACGAGGTGCCCTTGGAGACCCCGCGTGTCCACAGCCGG
15 GCACCGTCCCCTCTACAGCGTGGAGTTCTCAGAGGCCCTTCGGGGTGATCGTG
16 CACCGGCAGCTGGACGGCCGCTGCTGCTGAACACGACGGTGGCGCCCCTGTTCTT
17 TGCGGACCAGTTCCTCAGCTGTCCACCTCGCTGCCCTCGCAGTATATCACAGGCCT
18 CGCCGAGCACCTCAGTCCCCTGATGCTCAGCACCAGCTGGACCAGGATCACCCCTGT
19 GGAACCGGGACCTTGCGCCACGCCCGGTGCGAACCTTACGGGTCTACCCTTCT
20 ACCTGGCGCTGGAGGACGGCGGGTGGCACACGGGGTGTCTGCTAAACAGCAAT
21 GCCATGGATGTGGTCTGCAGCCGAGCCCTGCCCTTAGCTGGAGGTCGACAGGTGG

2005212435 23 Nov 2006

31

1 GATCCTGGATGTCTACATCTTCCTGGGCCAGAGCCCAAGAGCGTGGTGCAGCAGT
2 ACCTGGACGTTGTGGGATACCCGTTTCATGCCGCCATACTGGGGCCTGGGCTTCCACC
3 TGTGCCGCTGGGGCTACTCCTCCACCCTATCACCCGCCAGGTGGTGGAGAACATG
4 ACCAGGGCCCACTTCCCCCTGGACGTCCAATGGAACGACCTGGACTACATGGACTC
5 CCGGAGGGACTTCACGTTCAACAAGGATGGCTTCCGGGACTTCCCGGCCATGGTGC
6 AGGAGCTGCACCAGGGCGGGCGCTACATGATGATCGTGGATCCTGCCATCAGC
7 AGCTCGGGCCCTGCCGGGAGCTACAGGCCCTACGACGAGGGTCTGCCGAGGGGG
8 TTTTCATACCAACGAGACCGGCCAGCCGCTGATTGGGAAGGTATGGCCCGGTCC
9 ACTGCCTTCCCGACTTACCAACCCACAGCCCTGGCCTGGTGGGAGGACATGGT
10 GGCTGAGTTCATGACCAGGTGCCCTTCGACGGCATGTGGATTGACATGAACGAGC
11 CTTCCAATTTCATCAGGGGCTCTGAGGACGGCTGCCCAACAATGAGCTGGAGAAC
12 CCACCCTACGTGCCTGGGGTGGTGGGGGGACCCCTCCAGGCGCAACCATCTGTGC
13 CTCCAGCCACCAGTTCTCTCCACACACTACAACCTGCACAACCTTACGGCCTGAC
14 CGAAGCCATCGCCTCCACAGGGCGCTGGTGAAGGCTCGGGGGACACGCCCATTTG
15 TGATCTCCCGCTCGACCTTTGCTGGCCACGGCCGATACGCCGGCCACTGGACGGGG
16 GACGTGTGGAGCTCCTGGGAGCAGCTCGCCTCCTCCGTGCCAGAAATCCTGCAGTT
17 AACCTGCTGGGGTGCCTCTGGTTCGGGGCCGACGTCTGCCGCTTCTGGGCAACAC
18 CTCAGAGGAGCTGTGTGTGCGCTGGACCCAGCTGGGGCCCTTCTACCCCTTCATGCG
19 GAACCACAACAGCCTGCTCAGTCTGCCCCAGGAGCCGTACAGCTTCAGCGAGCCGG
20 CCCAGCAGGCCATGAGGAAGGCCCTCACCCCTGCGCTACGCACTCCTCCCCACCTCT
21 ACACGCTGTTCCACCAGGCCACGTCGCGGGGAGACCGTGGCCCGGCCCTCTTC
22 CTGGAGTTCCCAAGGACTCTAGCACCTGGACTGTGGACCACCAGCTCCTGTGGGG
23 GGAGGCCCTGCTCATACCCCAAGTGTCCAGGCCGGGAAGGCCGAAGTACTGGCT
24 ACTTCCCCTTGGGCACATGGTACGACCTGCAGACGGTGCCAATAGAGGCCCTTGGC
25 AGCCTCCACCCCCACCTGCAGCTCCCGTGAGCCAGCCATCCACAGCGAGGGGCA
26 GTGGGTGACGCTGCCGGCCCCCTGGACACCATCAACGTCTAGtctaga

27 Amino acid sequence of GAA Δ 817-952 (SEQ ID NO: 10).
28 MGVRHPPCSHRLLAVALVSLATAALLGHILLHDFLLVPRELSGSSPVLEETHPAHQGG
29 ASRPGPRDAQAHGRPRAVPTQCDVPPNSRFDCAKAITQEQCEARGCCYIPAKQGLQ
30 GAQMGQPWCFFPPSYPSYKLENLSSSEMGYTATLRTTPTFFPKDILTLRLDVMETEN
31 RLHFTIKDPANRRYEVPLETPRVHSRAPSPLYSEVFSEEPFGVIVHRQLDGRVLLNTTVA
32 PLFFADQLQLSTSLPSQYITGLAEHLSPLMLSTSWTRITLWNRDLAPTPGANLYGSHPF

YLALEDGGS AHGVFLLSNAMDVVLQSPALSWRSTGGILDVYIFLGPEPKSVVQQYL
 DVVGYPFMPYWG LGFHL CRWGYSSAITRQVVENMTRAHFPLDVQWNDLDYMSR
 R DFTFNKDGFRDFPAMVQELHQGGRYMMIVDPAISSSGPAGSYRYPYDEGLRRGVFITN
 ETGQPLIGKVWPGSTAFPDTNPTALAWWEDMVAEFHDQVPFDGMWIDMNEPSNFIR
 5 GSEDGCPNNELENPPYVPGVVGTLQAATICASSHQFLSTHYNLHNL YGLTEAIA SHRA
 LVKARGTRPFVISRSTFAGHG RYAGHW TGDVWSSWEQLASSVPEILQFNLLGVPIVGA
 DVC GFLGNTSEELCVRWTQLGAFYPFMRNHNSLLSLPQEPYSFSEPAQQAMRKALTLR
 YALLPHLYTLFHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPVLQAGK
 AEVTGYFPLGTWYDIQTVPIEALGSLPPPPAAPREPAIHSEGQWVWTL PAPTIN V.

- 10 [0097] To determine if the GAA C-terminal region functions when expressed in *trans*,
 pCEP-SS-GAA-791-952 was transfected into HEK293 cells alone as well as in combination
 with either plasmid pCEP-GAA Δ 817-952 or with pCEP-GAA Δ 817-952-GILT Δ 1-7. As
 controls, pCEP-GAA Δ 817-952 and pCEP-GAA Δ 817-952-GILT Δ 1-7 were also transfected
 into HEK293 cells alone. Standard transfection methods were used for the experiments. For
 15 single plasmid transfections, 1 μg of plasmid DNA was used. For co-transfections, 0.5 μg of
 each plasmid were used. 1 μg of total DNA was mixed with 96 μL of HEK293 growth media
 lacking serum and 4 μL FuGene6 (Roche) as directed by the manufacturer. 50 μL of the
 mixture were added to each duplicate well of HEK293 cells growing in 12-well plates in 1 mL
 Dulbecco's Modified Eagles Media supplemented with 1.5 g/L sodium bicarbonate, 10% heat-
 20 inactivated FBS, and 4 mM L-glutamine. Cells were incubated 2-3 days at 37°C in 5% CO₂.

- [0098] Growth media were collected and assayed to determine GAA activities as described
 (Reuser, A.J., et al. (1978) Am. J. Hum. Genet. 30:132-143). No GAA activity was detected in
 the media collected from HEK293 cells transfected with single plasmids. By contrast, GAA
 activities were present in the growth media collected from HEK293 cells co-transfected with
 25 pCEP-SS-GAA-791-952 and either pCEP-GAAΔ817-952 or pCEP-GAAΔ817-952-GILTΔ1-7
 (Table 1).

- [0099] Therefore, the two C-terminal deletion constructs pCEP-GAAΔ817-952 and pCEP-
 GAAΔ817-952-GILTΔ1-7 only express functional proteins when co-expressed with the C-
 terminal domain plasmid, pCEP-SS-GAA-791-952. This experiment demonstrated that the C-
 30 terminal GAA region cooperates with the mature, N-terminal region when coexpressed *in trans*.

Table 1. Transient co-transfection of GAA C-terminal and N-terminal domains.

Plasmid 1	Plasmid 2	GAA activity (nmol/hr-ml)
pCEP-SS-GAA-791-952	None	0
pCEP-GAAΔ817-952	None	0
pCEP-GAAΔ817-952-GILTA1-7	None	0
PCEP-SS-GAA-791-952	pCEP-GAAΔ817-952	14
PCEP-SS-GAA-791-952	pCEP-GAAΔ817-952-GILTA1-7	3

Example 2. Region required for efficient GAA trans-expression.

- [0100] In the transient co-transfection experiment described in Example 1, the GAA region including amino acid residues 792-817 is present in both halves of the trans-expression constructs. In order to determine if the overlap of this region is necessary for efficient trans-expression, a pair of constructs, pCEP-GAAΔ791-952-GILTA1-7 and PCEP-SS-GAA-791-952, were designed with no overlap and the GILT tag was fused at position 791. As indicated in Table 2, transient co-transfection experiments demonstrated that the presence of amino acid residues 792-817 within the C-terminal domain is required for efficient GAA trans-expression.

Table 2. Amino acid residues 792-817 is required for efficient GAA trans-expression.

Plasmid 1	Plasmid 2	GAA activity (nmol/hr-ml)
PCEP-GAA		58
PCEP-SS-GAA-791-952		1
PCEP-GAAΔ817-952-GILTA1-7		1
pCEP-GAAΔ791-952-GILTA1-7		1
pCEP-SS-GAA-817-952		1
PCEP-SS-GAA-791-952	pCEP-GAAΔ791-952-GILTA1-7	2
PCEP-SS-GAA-791-952	PCEP-GAAΔ817-952-GILTA1-7	16
pCEP-SS-GAA-817-952	pCEP-GAAΔ791-952-GILTA1-7	1
pCEP-SS-GAA-817-952	PCEP-GAAΔ817-952-GILTA1-7	1

Example 3. Construction of a GAA protein with an internal GILT tag.

23 Nov 2006

2005212435

1 [0101] PCR was used to first generate an insertion of the nucleotide sequence
2 GGCGCGCCG (SEQ ID NO: 11) after nucleotide 2370 of the complete human GAA sequence.
3 This insertion forms an *AscI* restriction site preceding Ala791. The GILT tag was PCR-
4 amplified with the following DNA oligos:

5 IGF7: gctctagaggcgccCTCGGACTTGGCGGGGTAGC (SEQ ID NO: 12)
6 IGF8: ggaattcaggcgccgGCTTACCGCCCCAGTGAGAC (SEQ ID NO: 13)
7 The amplified GILT tag contains an *AscI* restriction site at each terminus. This GILT tag was
8 digested with *AscI* and inserted into the *AscI* site preceding GAA Ala791 as described above.
9 DNA sequencing confirmed the in-frame orientation of the GILT insertion. This GAA cassette
10 containing an internal GILT tag preceding Ala791 was expressed in vector pCEP4 in a plasmid
11 named pCEP-GAA-IRGILT-4. pCEP-GAA-IRGILT-4 was found to contain a PCR-generated
12 mutation T1712C within the GAA coding sequence. This construct produced functional GAA
13 protein.

4 Example 4. GAA deletion constructs with N-terminal GILT tag.

5 [0102] A set of five tags suitable for N-terminal GAA expression (N-tags) were generated
6 by PCR amplification using primers indicated in Table 3. The GILT N-tag contains the native
7 IGF-II signal sequence and complete GILT epitope. The SS N-tag contains only the IGF-II
8 signal sequence.

9 [0103] For example, the GILTA1-7 N-tag contains the IGF-II signal sequence and GILT
10 epitope residues 8-67. It was generated with three PCR reactions: (1) PCR amplification from
11 human IGF-II DNA template using primers IGF1 and IGF4; (2) PCR amplification from human
12 IGF-II DNA template using primers IGF2 and IGF7; and (3) PCR amplification from the
13 products of the first two PCR reactions using primers IGF1 and IGF7.

14 [0104] The GILTA2-7 N-tag contains the IGF-II signal sequence, and GILT epitope residue
15 1 followed by residues 8-67. It was generated with three PCR reactions: (1) PCR amplification
16 from human IGF-II DNA template using primers IGF1 and IGF5; (2) PCR amplification from
17 human IGF-II DNA template using primers IGF3 and IGF7; and (3) PCR amplification from the
18 products of the first two PCR reactions using primers IGF1 and IGF7.

19 [0105] The SSGAA-GILT N-tag contains the GAA signal sequence within residues 1-69,
20 followed by the complete GILT epitope. It was generated with three PCR reactions: (1) PCR
21
22
23
24
25
26
27
28
29
30

2005212435 23 Nov 2006

- 1 amplification from human GAA DNA template using primers GAA13 and GI1; (2) PCR
 - 2 amplification from human IGF-II template using primers GI2 and IGF7; and (3) PCR
 - 3 amplification from the products of the first two PCR reactions using primers GAA13 and IGF7.
- 4 **[0106]** Each N-tag contains a 5' EcoRI restriction site and 3' AscI and XbaI sites. The AscI
- 5 site was used to fuse each tag to the GAA N-terminal deletion constructs described below.

Table 3. N-terminal tag constructs.

N-Tag Name	DNA Primers	PCR Template
GILT	IGF1: GGAATTCACACCAATGGGAATCCCAATGG (SEQ ID NO: 14) IGF7: GCTCTAGAGGCGCGCCCTCGGACTTGGCGGGGTAGC (SEQ ID NO: 12)	Human IGF-II
SS	IGF1: GGAATTCACACCAATGGGAATCCCAATGG (SEQ ID NO: 14) IGF6: GCTCTAGAGGCGCGCCAGCAGCAATGCAGCAGCAGG (SEQ ID NO: 15)	Human IGF-II
GILTΔ1-7	IGF1: GGAATTCACACCAATGGGAATCCCAATGG (SEQ ID NO: 14) IGF4: ACCAGTCCCCGCCGCACAGAGCAATGCAGCAGGAGGCG (SEQ ID NO: 16) IGF2: TCGCCTCGTGCTGCATTGCTCTGTGCGGCGGGGAGCTGG (SEQ ID NO: 17) IGF7: GCTCTAGAGGCGCGCCCTCGGACTTGGCGGGGTAGC (SEQ ID NO: 12)	Human IGF-II Human IGF-II

GILTA2-7	<p>IGF1: GGAATTCACACCAATGGGAATCCCAATGG (SEQ ID NO: 14)</p> <p>IGF5: ACCAGCTCCCCGCCGCACAGAGCAGCAATGCAGCACGAGG (SEQ ID NO: 18)</p> <p>IGF3: CCTCGTGCTGCATTGCTGCTCTGTGCGGGGGAGCTGG (SEQ ID NO: 19)</p> <p>IGF7: GCTCTAGAGGCGCGCCCTCGGACTTGGCGGGGGTAGC (SEQ ID NO: 12)</p>	Human IGF-II
SSGAA-GILT	<p>GAA13: GGAATTCCAACCATGGGAGTGAGGCACCCGCC (SEQ ID NO: 20)</p> <p>GII: GGGTCTCACTGGGGCGGTATGCCTGGGCATCCCGGGGC (SEQ ID NO: 21)</p> <p>GI2: GGCCCCGGGATGCCAGGCATACCGCCCCAGTGAGACC (SEQ ID NO: 22)</p> <p>IGF7: GCTCTAGAGGCGCGCCCTCGGACTTGGCGGGGGTAGC (SEQ ID NO: 12)</p>	Human GAA Human IGF-II

1
2 **[0107]** Portions of the N-terminal human GAA DNA sequence were deleted and replaced
3 with an *AscI* restriction site using PCR techniques. 5' DNA oligos used to define the site of
4 deletion are listed below (Table 4). 5' oligos were paired with various 3' oligos within the GAA
5 coding sequence, and the resulting DNA fragments were subsequently fused to the complete C-
6 terminal GAA coding sequence. The N-terminal *AscI* sites were fused to one of the five N-
7 terminal tags (N-tags) listed above to complete the expression cassettes (Table 4).

2005212435 23 Nov 2006

Table 4. GAA N-terminal deletion constructs.

Sequences complementary to GAA coding sequence are in upper case. EcoRI and Ascl restriction sites are in lower case.

Deletion Name	5' DNA Oligos
GAAΔ1-24	GAA32: ggaattcaggcgccgGCACTCCTGGGGCACATCC (SEQ ID NO:23)
GAAΔ1-28	GAA28: ggaattcaggcgccgCACATCCTACTCCATGATTTC (SEQ ID NO:24)
GAAΔ1-55	GAA29: ggaattcaggcgccgCACCAGCAGGGAGCCAGCAG (SEQ ID NO:25)
GAAΔ1-69	GAA30: ggaattcaggcgccgGCACACCCCGCCGTCCCAG (SEQ ID NO:26)
GAAΔ1-80	GAA39: ggaattcaggcgccgCAGTGCACGTCCCaCCCAAC (SEQ ID NO:27)
GAAΔ1-122	GAA33: ggaattcaggcgccgGGGCAGCCCTGGTGCTTCTTC (SEQ ID NO:28)
GAAΔ1-203	GAA34: ggaattcaggcgccgGCACCGTCCCCACTCTACAG (SEQ ID NO:29)

[0108] The expression cassettes listed in Table 4 contain an N-terminal tag (N-tag) fused to an N-terminal GAA deletion at a mutual Ascl site. The cassettes were cloned into the multiple cloning site of expression vector pCEP4 and transfected into HEK293 cells using the FuGene6 transfection reagent (Roche). Media from transient expression were collected 2-3 days post transfection and assayed for secreted GAA activity using a standard enzymatic assay (Reuser, A.J., et al. (1978) Am. J. Hum. Genet. 30:132-143).

Table 5. Relative transient expression of N-tagged GAA constructs.

Plasmid Name			Relative Transient Expression
Vector	N-tag	GAA Deletion	
PCEP	GILT	GAAAΔ1-24	+
PCEP	SS	GAAAΔ1-24	++
PCEP	GILTA1-7	GAAAΔ1-24	-
PCEP	GILTA2-7	GAAAΔ1-24	-
PCEP	SSGAA-GILT	GAAAΔ1-24	-
PCEP	GILT	GAAAΔ1-28	++
PCEP	SS	GAAAΔ1-28	++
PCEP	GILTA1-7	GAAAΔ1-28	+
PCEP	GILTA2-7	GAAAΔ1-28	++
PCEP	SSGAA-GILT	GAAAΔ1-28	++
PCEP	GILT	GAAAΔ1-55	++
PCEP	SS	GAAAΔ1-55	++++
PCEP	GILTA1-7	GAAAΔ1-55	++
PCEP	GILTA2-7	GAAAΔ1-55	++
PCEP	SSGAA-GILT	GAAAΔ1-55	++
PCEP	GILT	GAAAΔ1-69	++
PCEP	SS	GAAAΔ1-69	++++
PCEP	GILTA2-7	GAAAΔ1-69	++
PCEP	SSGAA-GILT	GAAAΔ1-69	+
PCEP	GILT	GAAAΔ1-80	++
PCEP	SS	GAAAΔ1-80	++++
PCEP	GILTA1-7	GAAAΔ1-80	++
PCEP	GILTA2-7	GAAAΔ1-80	++
PCEP	SSGAA-GILT	GAAAΔ1-80	+
PCEP	GILT	GAAAΔ1-122	-
PCEP	SS	GAAAΔ1-122	-
PCEP	GILTA1-7	GAAAΔ1-122	-
PCEP	GILTA2-7	GAAAΔ1-122	-

PCEP	SSGAA-GILT	GAAΔ1-122	-
PCEP	GILT	GAAΔ1-203	-
PCEP	SS	GAAΔ1-203	-
PCEP	GILTΔ1-7	GAAΔ1-203	-
PCEP	GILTΔ2-7	GAAΔ1-203	-
PCEP	SSGAA-GILT	GAAΔ1-203	-
PCEP		GAA	++

[0109] As these data indicated, the N-terminal portion of GAA including residues 1-80 is dispensable for transient expression, but deletions that disrupt or eliminate the trefoil domain do not produce functional protein.

- 5 [0110] Furthermore, as indicated in Table 6, the secretion of GAA can be improved by appropriately positioning a heterologous signal peptide, in this case the IGF-II signal peptide. Positioning the IGF-II signal peptide at either residue 56 or 70 of GAA gave a three fold increase in GAA secretion compared to native GAA, while positioning the IGF-II signal peptide at position 29 did not. This may be due to the retention of a putative trans-membrane domain adjacent to the GAA signal peptide.
- 10

Table 6. Changing GAA signal sequence positions affects GAA secretion

Plasmid	Transient GAA Activity (nmol/hr-mL)	
	Experiment 1	Experiment 2
pCEP-GAA	121	111
PCEP-SS-GAAΔ1-28	NA	89
pCEP-SS-GAAΔ1-55	402	NA
pCEP-SS-GAAΔ1-69	325	NA

- [0111] However, replacement of the native GAA signal peptide and trans-membrane domain with a heterologous signal peptide lowers the level of mannose-6-phosphate dependant cellular uptake associated with the protein (Figure 4). Uptake experiment was described in U.S. Patent Application Nos. 20040005309 and 20040006008, the contents of which are hereby incorporated by reference. As illustrated in Figure 4, pCEP-SS-GAAΔ1-69 had one third the amount of uptake into Pompe fibroblasts as did wild-type pCEP-GAA.
- 15

2005212435 23 Nov 2006

1 [0112] In contrast, as illustrated in Figure 5, by comparing uptake of pCEP-SS-GAAΔ1-69
2 with a construct with GILT tag, pCEP-SS-GILTΔ2-7-GAAΔ1-69, it was evident that the GILT
3 tag promotes specific uptake that can be competed by IGF-II. Thus, placement of the peptide
4 tag at position 70 not only permits efficient expression of the fusion protein and GAA activity,
5 but also provides a peptide tag that is properly folded and accessible, permitting receptor-
6 mediated uptake into target cells.

7 Example 5. Constructs with GILT1-87 tag and variants.

8 [0113] In order to increase the likelihood of proper folding of an N-terminal GILT tag, a
9 longer version of the GILT tag was generated that spans from IGF-II residues 1-87. The
0 additional IGF-II sequence should still allow receptor binding, and also provide a more native
1 folding environment for the core of the tag. The GILT1-87 tag was fused to positions 56 and 70
2 of GAA, resulting in GILT1-87-GAA56-952 and GILT1-87-GAA70-952, respectively. The
3 DNA and amino acid sequences GILT1-87-GAA56-952 are shown below.

4 DNA sequence of GILT1-87-GAA56-952 (SEQ ID NO: 30):
5 ggtaccACACCAATGGGAATCCCAATGGGGAAGTCGATGCTGGTGCTTCTCACCTTCTT
6 GGCCTTCGCCTCGTGCATTGCTGCTTACCGCCCCAGTGAGACCCTGTGCGGCGG
7 GGAGCTGGTGGACACCCTCCAGTTCGTCTGTGGGACC GCGCTTCTACTTCAGCA
8 GGCCCGCAAGCCGTGTGAGCCGTGCGAGCCGTGGCATCGTTGAGGAGTGCTGTTTC
9 CGCAGCTGTGACCTGGCCCTCCTGGAGACGTA CTGTGCTACCCCGCCAAGTCCGA
10 GAGGGACGTGTCGACCCCTCCGACCGTGTCCGGACA ACTTCCCCAGATAACCCG
11 TGGGCggcgcccgcgCACCAGCAGGGAGCCAGCAGACCAGGGCCCCGGGATGCCAGGC
12 ACACCCCGGCCGTCCCAGAGCAGTGCCACACAGTGCGACGTCCCCCAACAGCC
13 GCTTCGATTGCGCCCTGACAAGCCATCACCCAGGAACAGTGCGAGGCCGCGGC
14 TGCTGCTACATCCCTGCAAAGCAGGGGCTGCAGGGAGCCAGATGGGGCAGCCCTG
15 GTGCTTCTTCCCACCCAGCTACCCAGCTACAAGCTGGAGAACCTGAGCTCCTCTGA
16 AATGGGCTACACGGCCACCCTGACCCGTACACCCCACTTCTTCCCCAAGGACAT
17 CCTGACCCTGCGGCTGGACGTGATGATGGAGACTGAGAACCGCCTCCACTTCACGA
18 TCAAAGATCCAGCTAACAGGCGCTACGAGGTGCCCTTGAGACCCCGGTGTCCAC
19 AGCCGGGCACCGTCCCCACTCTACAGCGTGGAGTTCTCIGAGGAGCCCTTCGGGGTG
20 ATCGTGACCCGGCAGCTGGACGGCCGCTGCTGTGAAACACGACGGTGGCGCCCT
21 GTTCTTTGCGGACCAGTTCCTTCAGCTGTCCACCTCGCTGCCCTCGCAGTATATCAC

AGGCCTCGCCGAGCACCTCAGTCCCTGATGCTCAGCACCAAGCTGGACCAGGATCA
CCCTGTGGAACCGGGACCTTGCGCCACGCCCGGTGCGAACTCTACGGGTCTCAC
CCTTTCTACCTGGCGCTGGAGGACGGCGGTGCGCACACGGGTGTTCTGCTAAA
CAGCAATGCCATGGATGTGGTCTGCAGCCGAGCCCTGCCCTTAGCTGGAGGTGGA
5 CAGGTGGGATCCTGGATGTCTACATCTTCTGGGCCAGAGCCCAAGAGCGTGGTG
CAGCAGTACCTGGACGTTGTGGGATACCCGTTTCATGCCGCCATACTGGGGCCTGGG
CTTCCACTGTGCCGCTGGGGCTACTCCTCCACCGCTATCA.CCCGCCAGTGGTGGGA
GAACATGACCAGGGCCACTTCCCCCTGGACGTCCAATGGAACGACCTGGACTACA
TGGACTCCCGGAGGGACTTCACGTTCAACAAGGATGGCTTCCGGGACTTCCCGGCC
10 ATGGTGCAGGAGCTGCACCAGGGCGGCCGCGCTACATGATGATCGTGGATCCTGC
CATCAGCAGCTCGGGCCCTGCCGGGAGCTACAGGCCCTACGACGAGGGTCTGCGGA
GGGGGGTTTTTCATCACCACGAGACCGGCCAGCCGCTGATTTGGGAAGGTATGGCCC
GGGTCCACTGCCTTCCCGACTTCACCAACCCACAGCCCTGGCCTGGTGGGAGGA
CATGGTGGCTGAGTTCATGACCAGGTGCCCTTCGACGGCATGTGGATTGACATGA
15 ACGAGCCTTCAACTTCATCAGGGGCTCTGAGGACGGCTGCCCAACAATGAGCTG
GAGAACCACCCCTACGTGCCGTTGGGGTGGTGGGGGGACCC.TCCAGGCGGCAACCAT
CTGTGCCTCCAGCCACCAGTTTCTCTCCACACACTACAACC.TGCACAACCTCTACGG
CCTGACCGAAGCCATCGCCTCCACAGGGCGCTGGTGAAG.GCTCGGGGGACACGCC
CATTTGTGATCTCCCGCTCGACCTTTGCTGGCCACGGCCGATACGCCGGCCACTGGA
20 CGGGGGACGTGTGGAGCTCCTGGGAGCAGCTCGCCTCCTCCGTGCCAGAAATCCTG
CAGTTTAACTGTGGGGTGCCTCTGGTCCGGGCCGACGTCTGCCGCTTCTCTGGGC
AACACCTCAGAGGAGCTGTGTGTGCGCTGGACCCAGCTGGGGGCTTCTACCCCTT
CATGCGGAACCACAACAGCCTGCTCAGTCTGCCCCAGGAG.CCGTACAGCTTACGG
AGCCGGCCAGCAGGCCATGAGGAAGGCCCTCACCTGCGCTACGCACTCCTCCCC
25 CACCTTACACGCTGTTCCACCAGGCCACGTCGCGGGGGAGACCGTGGCCCGGCC
CCTTCTCTGGAGTTCCCAAGGACTCTAGCACCTGGACTGTGGACCACCAGTCTCT
GTGGGGGGAGGCCCTGCTCATCCCCAGTGTCCAGGCCGGGAAGGCCGAAGTGA
CTGGCTACTTCCCTTGGGCACATGGTACGACCTGCAGACGGTGCCTAATAGAGGCC
CTTGGCAGCCTCCACCCCCACCTGCAGCTCCCCGTGAGCCAGCCATCCACAGCGA
30 GGGGCAGTGGGTGACGCTGCCGGCCCCCTGGACACCATCAACGTCCACCTCCGGG
CTGGGTACATCATCCCCTGAGGGCCCTGGCCTCACAACCACAGAGTCCCGCCAG
CAGCCCATGGCCCTGGCTGTGGCCCTGACCAAGGGTGGAGAGGCCCGAGGGGAGCT

2005212435 23 Nov 2006

1 GTTCTGGGACGATGGAGAGAGCCTGGAAGTGCTGGAGCGAGGGGCCTACACACAG
2 GTCATCTTCTGGCCAGGAATAACACGATCGTGAATGAGCTGGTACGTGTGACCAG
3 TGAGGGAGCTGGCCTGCAGCTGCAGAAGGTGACTGTCCTGGGCGTGGCCACGGCGC
4 CCCAGCAGGTCCTCTCCAACGGTGTCCCTGTCTCCAACCTCACCTACAGCCCCGACA
5 CCAAGGTCCTGGACATCTGTGTCTCGCTGTTGATGGGAGAGCAGTTTCTCGTCAGCT
6 GGTGTTAGtctaga

7 Amino acid sequence of GILT1-87-GAA56-952 (SEQ ID NO: 31):

8 MGIPMGKSMVLVLLTFLAFASCCIAAYRPSETLCGGELVDTLQFVCGDRGFYFSRPASRV
9 SRRSRGIVEECCFRSCDLALLETYCATPAKSERDVSTPPTVLPDNFPRYPVGGAPHQQGA
0 SRPGPRDAQAHPGRPRAVPTQCDVPPNSRFDCAPDKAITQEQCEARGCCYIPAKQLQG
1 AQMGQPWCFPPSPYSYKLENLSSSEMGYTATLRTTPTFFPKDILTLRLDVMETENR
2 LHFTIKDPANRRYEVPLETPRVHSRAPSPLYSVEFSEEPFGVIVHRQLDGRVLLNTTVAP
3 LFFADQFLQLSTSLPSQYITGLAEHLSPLMLSTSWTRITLWNRDLAPTGANLYGSHPFY
4 LALEDGGSAGHVFLNSNAMDVVLQSPALSWRSTGGILDVYIFLGPEPKSVVQQYLD
5 VVGYPFMPYWGFGFHLCRWGYSSAITRQVVENMTRAHFPLDVQWNDLDYMDSTRR
6 DFTFNKDGFRDFPAMVQELHQGGRRYMMIVDPAISSSGPAGSYRYPDEGLRRGVFITNE
7 TGQPLIGKVWPGSTAFPDTNPTALAWWEDMVAEFHDQVPFDGMWIDMNEPSNFIRGS
8 EDGCPNNELENPYPVPGVVGGLQAATICASSHQFLSTHYNLHNLVGLTEAIAASHRALV
9 KARGTRPFVISRSTFAGHGRYAGHWTDVWSSWEQLASSVPEILQFNLLGVPLVGADV
0 CGFLGNTSEELCVRWTQLGAFYPMRNHNSLLSLQEPYSFSEPAQQAMRKALTLRYA
1 LLPHLYTLFHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPVLQAGKAE
2 VTGYFPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGQWVTLAPLDTINVHLRAGY
3 IIPLQGPGLTTESRQQPMALAVALKGGEARGELFWDDGESLEVLERGAYTQVIFLAR
4 NNTIVNELVRVTSEGAGLQLQKVTVLGVATAPQQVLSNGVPPVSNFTYSPDTKVLDCV
5 LLMGEQFLVSWC.

26 DNA sequence of GILT1-87-GAA70-952 (SEQ ID NO: 32):

27 ggtaccACACCAATGGGAATCCAATGGGGAAGTCGATGCTGGTCTTCTCACCTTCTT
28 GGCCTTCGCCTCGTGTGCATTGCTGCTTACCGCCCCAGTGAGACCCTGTGCGGCGG
29 GGAGCTGGTGGACACCCTCCAGTTCGTCTGTGGGGACCGCGGCTTCTACTTCAGCA
30 GGCCCGCAAGCCGTGTGAGCCGTCGAGCCGTGGCATCGTTGAGGAGTGCTGTTTC
31 CGCAGCTGTGACCTGGCCCTCCTGGAGACGTAAGTGTGCTACCCCGCCAAGTCCGA

GAGGGACGTGTCGACCCCTCCGACCGTGCTTCCGGACAACITCCCCAGATAACCCG
TGGGCggggccgGCACACCCCGGCCGTCCCAGAGCAGTGCCACACAGTGCGA.CGTC
CCCCCAACAGCCGCITCGATTGCGCCCCTGACAAGGCCATCACCCAGGAAC.AGTG
CGAGGCCCGCGGCTGCTGCTACATCCCTGCAAAGCAGGGGCTGCAGGGAGCC.CAGA
5 TGGGGCAGCCCTGGTGTCTTCTCCACCCAGCTACCCAGCTACAAGCTGGAGAAC
CTGAGCTCCTCTGAAATGGGCTACACGGCCACCCTGACCCGTACCACCCCCACCTTC
TTCCCAAGGACATCTGACCCTGCGGCTGGACGTGATGATGGAGACTGAGA.ACCG
CCTCCACTTACGATCAAAGATCCAGCTAACAGGCGCTACGAGGTGCCCTTGGAGA
CCCCGCGTGTCCACAGCCGGGACCGTCCCCACTTACAGCGTGGAGTTCTC.GAGG
10 AGCCCTTCGGGGTGATCGTGCACCGGCAGCTGGACGGCCGCGTGTCTGTAAC.ACG
ACGGTGGCGCCCCIGTTCITTTGCGGACCAGITTCCTTCAGCTGTCCACCTCGCTGCCC
TCGCAGTATATCACAGGCCTCGCCGAGCACCTCAGTCCCCTGATGCTCAGCAC.CAG
CTGGACCAGGATCACCTGTGGAACCGGGACCTTGCGCCACGCCCGGTGCG.AACC
TCTACGGGTCTCACCTTTCTACCTGGCGCTGGAGGACGGCGGGTGGGCACACGGG
15 GTGTTCTGCTAAACAGCAATGCCATGGATGTGGTCTGCAGCCGAGCCCTGCCCTT
AGCTGGAGGTCGACAGGTGGGATCCTGGATGTCTACATCTTCTGGGCCCAG.AGCC
CAAGAGCGTGGTGCAGCAGTACCTGGACGTTGTGGGATACCCGTTTCATGCCGCCAT
ACTGGGGCTGGGCTTCCACCTGTGCCGCTGGGGCTACTCCTCCACCGCTATC.ACCC
GCCAGGTGGTGGAGAACATGACCAGGGCCCACTTCCCCCTGGACGTCCAATGGAAC
20 GACCTGGACTACATGGACTCCCGGAGGGACTTACGTTCAACAAGGATGGCTTCCG
GGACTTCCCGCCATGGTGCAGGAGCTGCACCAGGGCGGCCGCGCTACATG.ATGA
TCGTGGATCCTGCCATCAGCAGCTCGGGCCCTGCGGGAGCTACAGGCCCTACGAC
GAGGGTCTGCGGAGGGGGGTTTTTCATCACCACGAGACCGGCCAGCCGCTGA.TTGG
GAAGGTATGGCCCGGTCCACTGCCTTCCCCGACTTCAACAACCCACAGCCC.TGG
25 CCTGGTGGGAGGACATGGTGGCTGAGTTCATGACCAGGTGCCCTTCGACGGC.ATG
TGGATTGACATGAACGAGCCTTCCAACTTCATCAGGGGCTCTGAGGACGGCTGCCC
CAACAATGAGCTGGAGAACCCACCCTACGTGCCTGGGGTGGTTGGGGGACCCTCC
AGGCGGCAACCATCTGTGCCTCCAGCCACCAGTTTCTCTCCACACACTACAAC.CTGC
ACAACCTTACGGCCTGACCGAAGCCATCGCCTCCACAGGGCGTGGTGAAGGCT
30 CGGGGGACACGCCATTTGTGATCTCCCGCTCGACCTTGCTGGCCACGGCCG.ATAC
GCCGGCCACTGGACGGGGACGTGTGGAGCTCCTGGGAGCAGCTCGCCTCCTCCGT
GCCAGAAATCCTGCAGTTTAACTGCTGGGGGTGCCTCTGGTCGGGGCCGACGTCT

2005212435 23 Nov 2006

1 GCGGCTTCCTGGGCAACACCTCAGAGGAGCTGTGTGTGCGCTGGACCCAGCTGGGG
2 GCCTTCTACCCCTTCATGCGGAACCACAACAGCCTGCTCAGTCTGCCCCAGGAGCCG
3 TACAGCTTCAGCGAGCCGGCCAGCAGGCCATGAGGAAGGCCCTCACCTGCGCTA
4 CGCACTCCTCCCCACCTCTACACGCTGTTCCACCAGGCCACGTGCGGGGGGAGA
5 CCGTGGCCCCGGCCCTCTCCTGGAGTTCCCCAAGGACTCTAGCACCTGGACTGTGG
6 ACCACCAGCTCCTGTGGGGGGAGGCCCTGCTCATACCCCAGTGTCCAGGCCGGG
7 AAGGCCGAAGTACTGGCTACTTCCCCTTGGGCACATGGTACGACCTGCAGACGGT
8 GCCAATAGAGGCCCTTGGCAGCCTCCCACCCACCTGCAGCTCCCCGTGAGCCAG
9 CCATCCACAGCGAGGGGCAGTGGGTGACGCTGCCGGCCCCCTGGACACCATCAAC
0 GTCCACCTCCGGGCTGGGTACATCATCCCCCTGCAGGGCCCTGGCCTCACAACCAC
1 AGAGTCCCGCCAGCAGCCATGGCCCTGGCTGTGGCCCTGACCAAGGGTGGAGAGG
2 CCCGAGGGGAGCTGTTCTGGGACGATGGAGAGAGCCTGGAAGTGTGGAGCGAGG
3 GGCCTACACACAGGTCATCTCCTGGCCAGGAATAACACGATCGTGAATGAGCTGG
4 TACGTGTGACCAGTGAGGGAGCTGGCCTGCAGCTGCAGAAGGTGACTGTCCTGGGC
5 GTGGCCACGGCGCCCCAGCAGGTCCTCTCCAACGGTGTCCCTGTCTCCAACCTACC
6 TACAGCCCCGACACCAAGGTCCTGGACATCTGTGTCTCGCTGTTGATGGGAGAGCA
7 GTTCTCGTCAGCTGGTGTTA Gtctaga

8 Amino acid sequence of GILT1-87-GAA70-952 (SEQ ID NO: 33):
9 MGIPMGKSMVLVLLTFLAFASCCIAAYRPSETLCGGELVDTLQFVCGDRGFYFSRPASRV
10 SRRSRGIVEECCFRSCDLALLETYCATPAKSERDVSTPPTVLPDNFPRYPVGGAPAHPR
11 PRAVPTQCDVPPNSRFDCAIDKAITQEQCEARGCCYIPAKQGLQGAQMGQPWCFPPS
12 YPSYKLENLSSSEMGYTATLTRITPTFFPKDILTRLVDVMMETENRLHFTIKDPANRRYE
13 VPLETPRVHSRAPSPLYSVEFSEEPFGVIVHRQLDGRVLLNTTVAPLFFADQLQLSTSLP
14 SQYITGLAEHLSPLMLSTSWTRITLWNRDLAPTPGANLYGSHPFYLALEDGGS AHGVFL
15 LNSNAMDVVLQPSALSWRSTGGILDVYIFLGPEPKSVVQQYLDVVGYPFMPYWG LG
16 FHLCRWGYSSAITRQVVENMTRAHFPLDVQWNDLDYMSRRDFTFNKDGFRDFPAM
17 VQELHQGGRRYMMIVDPAISSSGPAGSYRKYDEGLRRGVFITNETGQPLIGKVPWGSTA
18 FPDFTNPTALAWWEDMVAEFHDQVPFDGMWIDMNEPSNFIRGSEDGCPNNELENPPYV
19 PGVVGTLQAATICASSHQFLSTHYNLHNLGLTEAIAASHRALVKARGTRPFVISRSTFA
20 GHGRYAGHWTGDVWSSWEQLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRW
21 TQLGAFYFPMRNHNSLLSLPQEPYSFSEPAQQAMRKALTRYALLPHLYTLFHQA HVA
22 GETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPVLQAGKAEVTGYFPLGTWYDLQT

23 Nov 2006

2005212435

1 VPIEALGSLPPPPAAPREPAIHSEGQWVTLPAPLDTINVHLRAGYIIPLQGPGLTTESRQ
2 QPMALAVALTGGGEARGELFWDDGESLEVLERGAYTQVIFLARNNTIVNELVRVTSEG
3 AGLQLQKVTVLGVATAPQQVLSNGVPSNFTYSPDTKVLDCVSLLMGEQFLVSWC.

4 [0114] The 5' Asp718 site was cloned into the Asp718 site of pCEP4, and the 3' Xba site was
5 blunted with Klenow and cloned into the HindIII site of pCEP4, resulting in pCEP-GILT1-87-
6 GAA56-952 and pCEP-GILT1-87-GAA70-952, respectively. The constructs also contain Gly-
7 Ala-Pro linker sequence (AclI site). These constructs express proteins with GAA enzymatic
8 activity.

9 [0115] In addition, the modification R68A was introduced to the GILT1-87 tag to remove a
10 potential proteolysis site within the GILT tag (GILT1-87-R68A). The DNA (SEQ ID NO: 34)
11 and amino acid (SEQ ID NO: 35) sequences of GILT1-87-R68A are shown below (the mutated
12 sequences are underlined).

13 DNA sequence of GILT1-87-R68A (SEQ ID NO: 34)
14 GGTACCACACCAATGGGAATCCCAATGGGGAAGTCGATGCTGGTCTTCTCACCTT
15 CTTGGCCTTCGCCTCGTGCTGCATTGCTGCTTACCGCCCCAGTGAGACCCTGTGCGG
16 CGGGGAGCTGGTGGACACCCTCCAGTTCGTCTGTGGGGACCGCGGCTTCTACTTCA
17 GCAGGCCCGCAAGCCGTGTGAGCCGTCGCAGCCGTGGCATCGTTGAGGAGTGCTGT
18 TTCCGCAGCTGTGACCTGGCCCTCCTGGAGACGTACTGTGCTACCCCGCCAAGTCC
19 GAGGCGGACGTGTCGACCCCTCCGACCGTGCTCCGGACAACCTCCCCAGATACCC
20 CGTGGGCGGCGCGCCG

21 Amino acid sequence of GILT1-87-R68A (SEQ ID NO: 35)
22 MGIPMGKSMVLVLLTFLAFASCCIAAYRPSETLGGELVDTLQFVCGDRGFYFSRPASRV
23 SRRSRGIVEECCFRSCDLALLETYCATPAKSEADVSTPPTVLPDNFPRYPVGGAP

24 Fusion of this tag to GAA positions 56 and 70 resulted in pCEP-GILT1-87-R68A-GAA56-952
25 and pCEP-GILT1-87-R68A-GAA70-952. These constructs expressed proteins with GAA
26 enzymatic activity.

27 [0116] In addition, point mutations were also introduced to substitute three Ser/Thr residues
28 within the GILT1-87 tag to remove glycosylation sites (Δ GS) (GILT1-87- Δ GS). The

29

2005212435 23 Nov 2006

1 DNA (SEQ ID NO: 36) and amino acid (SEQ ID NO: 37) sequences of GILT1-87- Δ GS are
2 shown below (the mutated sequences are underlined).

3 DNA sequence of GILT1-87- Δ GS (SEQ ID NO: 36).
4 GGTACCACACCAATGGGAATCCCAATGGGGAAGTCGATGCTGGTGCTTCTCACCTT
5 CTTGGCCTTCGCCTCGTGCTGCATTGCTGCTTACCGCCCCAGTGAGACCCTGTGCGG
6 CGGGGAGCTGGTGGACACCCTCCAGTTCGTCTGTGGGACCGCGGCTTCTACTTCA
7 GCAGGCCCGCAAGCCGTGTGAGCCGTCGCAGCCGTGGCATCGTTGAGGAGTGCTGT
8 TTCCGCAGCTGTGACCTGGCCCTCCTGGAGACGTA^{CT}GTGCTACCCCCGCAAGTCC
9 GAGAGGGACGTGCGGCCCTCCGGCCGTGCTTCCGGACA^{ACT}TCCCCAGATACCC
0 CGTGGGCGGCGCGCG

1 Amino acid sequence of GILT1-87- Δ GS (SEQ ID NO: 37).
2 MGIPMGKSMVLVLLFLAFASCCIAAYRPSETLCGGELVDTLQFVCGDRGFYFSRPASRV
3 SRRSRGIVEECCFRSCDLALLETYCATPAKSERDVAAPPVLDPNFPYPVGGAP

4 This modified GILT tag was fused to position 70 of GAA yielding pCEP-GILT1-87- Δ GS-
5 GAA70-952. This construct expressed protein with GAA enzymatic activity.

6 [0117] In addition, GILT tag incorporating both R68A and Δ GS modifications was
7 generated (GILT1-87-R68A- Δ GS). The DNA (SEQ ID NO: 38) and amino acid (SEQ ID
8 NO:39) sequences of GILT1-87-R68A- Δ GS are shown below (the mutated sequences are
9 underlined).

20 DNA sequence of GILT1-87-R68A- Δ GS (SEQ ID NO: 38).
21 GGTACCACACCAATGGGAATCCCAATGGGGAAGTCGATGCTGGTGCTTCTCACCTT
22 CTTGGCCTTCGCCTCGTGCTGCATGCTGCTTACCGCCCCAGTGAGACCCTGTGCGG
23 CGGGGAGCTGGTGGACACCCTCCAGTTCGTCTGTGGGACCGCGGCTTCTACTTCA
24 GCAGGCCCGCAAGCCGTGTGAGCCGTCGCAGCCGTGGCATCGTTGAGGAGTGCTGT
25 TTCCGCAGCTGTGACCTGGCCCTCCTGGAGACGTA^{CT}GTGCTACCCCCGCAAGTCC
26 GAGGCGGACGTGCGGCCCTCCGGCCGTGCTTCCGGACA^{ACT}TCCCCAGATACCC
27 CGTGGGCGGCGCGCG

23 Nov 2006

2005212435

1 Amino acid sequence of GILT1-87-R68A-ΔGS (SEQ ID NO: 39).
2 MGIPMGKSMMLVLLTFLAFASCCIAAYRPSETLCGGELVDTLQFVCGDRGFYFSRPASRV
3 SRRSRGIVECCFRSCDLALLETYCATPAKSEADVAAPPVLPDNFPRYPVGGAP

4 The modified GILT1-87-R68A-ΔGS was used to generate constructs pCEP-GILT1-87-
5 R68AΔGS-GAA56-952 and pCEP-GILT1-87-R68AΔGS-GAA70-952. Both constructs
6 expressed protein with GAA enzymatic activity.

7 **[0118]** Western blots were performed on the above proteins with GILT tag fused at GAA
8 position 56. As illustrated in Figure 6, the precursor proteins are full length and contain the
9 IGF-II tag. The ΔGS mutation appears to produce a protein with a slightly faster mobility,
0 consistent with absence of a carbohydrate moiety.

1 Example 6. Additional Constructs with longer and modified GILT tags.

2 **[0119]** In an effort to provide a native folding environment for an IGF-II tag, a precursor
3 form of IGF-II including amino acid 8-156 was used as an internal tag fused at GAA position
4 791. In addition, mutations E67A and D69S were made in IGF-II sequence to introduce a P2/P1
5 proteolysis processing site in order to promote cleavage downstream of position 87 within the
6 IGF-II tag. The resulting construct, pCEP-GAA-791IGF2-P2/P1 yields a protein with GAA
7 enzymatic activity. The DNA and amino acid sequences of pCEP-GAA-791IGF2-P2/P1 are
8 shown below.

19 DNA sequence of pCEP-GAA-791IGF2-P2/P1 (SEQ ID NO: 40)
20 ATGGGAGTGAGGCACCCGCCCTGCTCCCACCGGCTCCTGGCCGTCTGCGCCCTCGTG
21 TCCTTGGAACCGCTGCACTCCTGGGGCACATCCTACTCCATGATTTCCCTGCTGGTT
22 CCCCAGAGCTGAGTGGCTCCTCCCCAGTCCTGGAGGAGACTCACCCAGCTCACCA
23 GCAGGGAGCCAGCAGACCAGGGCCCCGGGATGCCAGGCACACCCCGCCGTCCC
24 AGAGCAGTGCCACACAGTGCGACGTCCCCCAACAGCCGCTTCGATTGCGCCCC
25 TGACAAGGCCATCACCCAGGAACAGTGCGAGGCCCGGGCTGCTGCTACATCCCTG
26 CAAAGCAGGGGCTGCAGGGAGCCAGATGGGGCAGCCCTGGTGTCTTCCCACCC
27 AGCTACCCAGCTACAAGCTGGAGAACCCTGAGCTCCTCTGAAATGGGCTACACGGC
28 CACCCTGACCCGTACCACCCACCTTCTTCCCAAGGACATCCTGACCCTGCGGCT
29 GGACGTGATGATGGAGACTGAGAACCGCCTCCACTTCACGATCAAAGATCCAGCTA
30 ACAGGCGCTACGAGGTGCCCTTGAGACCCCGGTGTCCACAGCCGGGCACCGTCC

CCACTCTACAGCGTGGAGTTCTCIGAGGAGCCCTTCGGGGTGATCGTGCACCGGCAG
CTGGACGGCCCGTGTCTGCTGAACACGACGGTGGCGCCCTGTTCCTTTGCCGACCA
GTTCCCTCAGCTGTCCACCTCGCTGCCCTCGCAGTATATCACAGGCCTCGCCGAGCA
CCTCAGTCCCCTGATGCTCAGCACCAGCTGGACCAGGATCACCTGTGGAACCGGG
5 ACCTTGCGCCACGCCCGGTGCGAACCTCTACGGGTCTCACCTTTCTACCTGGCGC
TGGAGGACGGCGGTGCGCACACGGGGTGTTCCTGCTAAACAGCAATGCCATGGAT
GTGGTCTGCAGCCGAGCCCTGCCCTTAGCTGGAGGTGACAGGTGGGATCCTGGA
TGTCTACATCTTCTGGGCCAGAGCCCAAGAGCGTGGTGCAGCAGTACCTGGACG
TTGTGGGATACCCGTTTATGCGCCATACTGGGGCTGGGCTTCCACCTGTGCCGT
10 GGGGCTACTCTCCACCGCTATCACCCGCCAGTGGTGGAGAACATGACCAGGGCC
CACTTCCCCTGGACGTCCAATGGAACGACCTGGACTACATGGACTCCCGGAGGGA
CTTCACGTTCAACAAGGATGGCTTCCGGGACTTCCCGCCATGGTGCAGGAGCTGC
ACCAGGGCGGCCGGCGCTACATGATGATCGTGGATCCTGCCATCAGCAGCTCGGGC
CCTGCCGGGAGCTACAGGCCCTACGACGAGGGTCTGCGGAGGGGGGTTTTATCAC
15 CAACGAGACCGCCAGCCGCTGATTGGGAAGGTATGCCCGGTCCACTGCCTTCC
CCGACTTACCAACCCACAGCCCTGGCTGGTGGGAGGACATGGTGGCTGAGTTC
CATGACCAGGTGCCCTTCGACGGCATGTGGATTGACATGAACGAGCCTTCAAACCT
CATCAGGGGCTCTGAGGACGGCTGCCCAACAATGAGCTGGAGAACCCACCTACG
TGCTGGGGTGGTGGGGGGACCTCCAGCGGCAACCATCTGTGCCTCCAGCCAC
20 CAGTTTCTCTCCACACACTACAACCTGCACAACCTCTACGGCCTGACCGAAGCCATC
GCCTCCACAGGGCGCTGGTGAAGGCTCGGGGGACACGCCCATTTGTGATCTCCCG
CTCGACCTTTGTGGCCACGGCCGATACGCCGGCACTGGACGGGGGACGTGTGGA
GCTCCTGGGAGCAGCTCGCCTCCTCCGTGCCAGAAATCCTGCAGTTTAACTGCTGG
GGGTGCCTCTGGTCGGGGCCGACGTCTGCGGCTCCTGGGCAACACCTCAGAGGAG
25 CTGTGTGTGCGCTGGACCCAGCTGGGGGCTTCTACCCCTTCATGCGGAACCACAAC
AGCCTGCTCAGTCTGCCCCAGGAGCCGTAACAGCTTACGCGAGCCGGCCAGCAGGC
CATGAGGAAGGCCCTCACCTGCGCTACGCACTCCTCCCCACCTTACACGCTGTT
CCACCAGGCCACGTGCGGGGGAGACCGTGGCCCGGCCCTTCTCTGGAGTTCC
CCAAGGACTCTAGCACCTGGACTGTGGACCACCAGCTCCTGTGGGGGAGGCCCTG
30 CTCATACCCCAGTGTCTCCAGGCCGGAAAGCCGAAGTACTGGCTACTTCCCTT
GGGCACATGGTACGACCTGCAGACGGTGCCAATAGAGGCCCTTGGCAGCCTCCAC
CCCCACCTggcgccgCTGTGCGGCGGGGAGCTGGTGGACACCTCCAGTTCGTCTGT

2005212435 23 Nov 2006

49

1 GGGGACCGCGGCTTCTACTTCAGCAGGCCCGCAAGCCGTGTGAGCCGTCGCAGCCG
2 TGGCATCGTTGAGGAGTGTGTTTCCGCAGCTGTGACCTGGCCCTCCTGGAGACGTA
3 CTGTGCTACCCCGCCAAGTCCGcGAGGtcCGTGTGACCCCTCCGACCGTGCTCCG
4 GACAACTTCCCCAGATACCCCGTGGGCAAGTTCTTCCAATATGACACCTGGAAGCA
5 GTCCACCCAGCGCCTGCGCAGGGGCTGCCTGCCCTCCTGCGTGCCCGCCGGGGTC
6 ACGTGCTCGCCAAGGAGCTCGAGGCGTTCAGGGAGGCCAAACGTACCCGTCCCCTG
7 ATTGCTCTACCCACCCAAGACCCCGCCACGGGGGCGCCCCCAGAGATGGCCAG
8 CAATCGGAAAGggcgccgGCAGCTCCCCGTGAGCCAGCCATCCAACAGCGAGGGGCG
9 TGGGTGACGCTGCCGGCCCCCTGGACACCATCAACGTCCACCTCCGGGCTGGGTA
0 CATCATCCCCCTGCAGGGCCCTGGCCTCACAACCACAGAGTCCCGCCAGCAGCCCA
1 TGGCCCTGGCTGTGGCCCTGACCAAGGGTGGAGAGGCCCGAGGGGAGCTGTTCTGG
2 GACGATGGAGAGAGCCTGGAAGTGTGAGGAGGGGCCCTACACACAGGTCATCT
3 TCCTGGCCAGGAATAACACGATCGTGAATGAGCTGGTACGTGTGACCAGTGAGGGA
4 GCTGGCCTGCAGCTGCAGAAGGTGACTGTCCTGGGCGTGGCCACGGCGCCCCAGCA
5 GGTCTCTCCAACGGTGTCCCTGTCTCCAACCTCACCTACAGCCCCGACACCAAGGT
6 CCTGGACATCTGTGTCTCGCTGTTGATGGGAGAGCAGTTTCTCGTCAGCTGGTGTTA
7 G
3 Amino acid sequence of pCEP-GAA-791IGF2-P2/P1 (SEQ ID NO: 41)
3) MGVRHPPCSHRLLAVALVSLATAALLGHILLHDFLLVPRELSSSPVLEETHPAHQQG
4) ASRPGPRDAQAHPGRPRAVPTQCDVPPNSRFDCAPDKAITQEQQEARGCCYIPAKQGLQ
21 GAQMGQPWCFFPSPYSYKLENLSSSEMGYTATLTRTPTFFPKDILTLRLDVMETEN
22 RLHFTIKDPANRRYEVPLETPRVHSRAPSPLYSVEFSEBPFVIVHRQLDGRVLLNTTVA
23 PLFFADQFLQLSTSLPSQYITGLAEHLSPLMLSTSWTRITLWNRDLAPTPGANLYGSHPF
24 YLALEDGGSAGHVFLNSNAMDVVLQPSPALSWRSTGGILDVYIFLGPEPKSVVQQYL
25 DVVGYPFMPYWGFGFHLCRWGSSTAITRQVVENMTRAHFPLDVQWNDLDYMSR
26 RDFTFNKDGFRDFPAMVQELHQGRRYMMIVDPAISSSGPAGSYRPFYDEGLRRGVFITN
27 ETGQPLIGKVWPGSTAFPFTNPTALAWWEDMVAEFHDQVPFDGMWIDMNEPSNFIR
28 GSEDGCPNNELENPYVPGVVGTLQAATICASSHQFLSTHYNLHNLGLTEAIAISHRA
29 LVKARGTRPFVISRSTFAGHGRYAGHWTGDVWSSWEQLASSVPEILQFNLLGVPLVGA
30 DVCGFLGNTSEELCVRWTQLGAFYPMRNHNSLLSLPQEPYSFSEPAQQAMRKALTLR
31 YALLPHLYTLFHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPVLQAGK
32 AEVTGYFPLGTWYDLQTVPIEALGSLPPPPGAPLCGGELVDTLQFVCGDRGFYFSRPAS

2005212435 23 Nov 2006

1 RVSRRSRGIVEECCFRSCDLALLETYCATPAKSARSVSTPPTVLPDNFPRYPVGGKFFQYD
 2 TWKQSTQRLRRGLPALLRARRGHVLAKELEAFREAKRHRPLIALPTQDPAHGGAPPEM
 3 ASNRKGGAPAAPREPAIHSEGQWVTLPAPLDTINVHLRAGYIIPLQGGPGLTTTESRQQPMA
 4 LAVALTKGGGEARGELFWDDGESLEVLERGAYTQVIFLARNTIVNELVRVTSEGAGLQ
 5 LQKVTVLGVATAPQQVLSNGVPPVSNFTYSPDTKVLDICVSLLMGEQFLVSWC.

6 [0120] To further improve presentation and/or folding of the GILT tag fused at the N-
 7 terminus (e.g., position 70), a spacer with the sequence Gly-Gly-Gly-Gly-Pro (SEQ ID
 8 NO:3) was inserted between an N-terminal GILTA2-7 tag and the GAA fusion point position 70,
 9 yielding pCEP-GILTA2-7-spcr1-GAA70-952. This construct expressed protein with GAA
 10 enzymatic activity.

1 Example 7. GAA construct with AscI restriction site insertion.

2 [0121] Constructs were made to include an insertion of Gly-Ala-Pro sequence (an AscI
 3 restriction site) within GAA region amino acid residues 783-791 according to standard
 4 molecular techniques. As indicated in Table 7, the insertion of AscI restriction site increases
 5 transient GAA enzyme expression levels. This insertion possibly could cause a shift of the
 6 enzyme to a high-affinity form. Normally the precursor GAA matures into the high-affinity
 7 GAA form after cleavage in the 783-791 boundary region (Moreland et al., 2004). It was
 8 reported that, after cleavage, the N-terminal region and the C-terminal region remain associated
 9 (Moreland et al., 2004).

20 Table 7. AscI restriction site insertion increases transient expression.

Enzyme	Transient Expression	Average transient expression from two experiments U/ml
GAA	++	20
GAA-779Asc	++	11
GAA-787Asc	+++	199
GAA-791Asc	+++	243
GAA-796Asc	++ - +++	88
GAA-881Asc	+	5
GAA-920Asc	+	5

[0122] In order to determine if insertion of the three residues promotes cleavage of the precursor GAA, Western blot analysis was performed comparing wild-type GAA and GAA-791Asc proteins using anti-GAA polyclonal antibody. As indicated in Figure 7, GAA-791Asc migrated with a similar mobility to that of wild-type GAA, indicating that the insertion of the
5 three residues does not promote proteolysis.

[0123] An alternate explanation for the increase in enzyme activity is that insertion of the residues within the domain boundary allows a conformational shift to the high-affinity form without cleavage of the two domains. This can be tested using affinity chromatography and comparing the binding affinity of GAA-791Asc and wild-type GAA on a Superdex 200 column
10 as described in Moreland et al., 2004.

[0124] In addition, construct pCEP-GILTA2-7-GAA70-952-791Asc was made to combine the 791AscI site insertion with an N-terminal GILT tag at position 69.

Example 8. Internal GILT tags with engineered proteolysis sites.

[0125] In order to generate an active internal GILT tag, experiments were designed to place the tag within the GAA 779-796 region engineered with a Factor X restriction protease site
15 downstream of the tag. Treatment of the expressed protein with Factor Xa would release the C-terminal portion of GAA and potentially reveal an exposed and active GILT tag.

[0126] Accordingly, GILT tags with a downstream Factor X protease site were placed within GAA at positions 787, 779, and 796. All three resulting proteins had GAA enzymatic
20 activity. In Western analysis, as illustrated in Figure 8, all three proteins contained the GILT tag as probed by an anti-IGF-II antibody. All three protein preparations contained a band with a relative mobility (M_r 120,000-140,000) consistent with the presence of a full-length precursor. All three protein preparations also contained a faster migrating intermediate band (M_r 85,000-100,000) which retained the GILT tag. Upon treatment of the proteins with Factor Xa, most of
25 the full-length band is removed, and the intermediate band is shifted slightly lower. The GILT tag is retained in the Xa-treated intermediate bands.

[0127] It is possible that the presence of the GILTXa tags within the GAA sequence promotes proteolysis at a site downstream of the Factor X site. Position 816 has been reported as a site of GAA processing upon maturation.

2005212435 23 Nov 2006

1 [0128] A possible GAA C-terminal processing model is illustrated in Figure 8.

2 Example 9. Human/Mouse GAA Hybrids.

3 [0129] In order to improve the folding of GILT tag, chimeric proteins composed of N-
4 terminal human GAA and C-terminal mouse GAA were constructed with fusion points at amino
5 acid positions 791, 796, 816, 881, and 920 of human GAA. An Ascl restriction site including
6 sequence Gly-Ala-Pro was introduced at the point of fusion.

7 [0130] Specifically, chimeric human/mouse GAA proteins were made with C-terminal
8 portions of human GAA replaced with corresponding mouse GAA C-terminal sequence. DNA
9 cassettes were constructed by fusing the human and mouse portions at a common linker
0 sequence, ggcgcgccg (SEQ ID NO: 11), that contains a unique Ascl site and encodes the
1 sequence Gly-Ala-Pro (GAP). Mouse portions of the GAA hybrid were generated by PCR with
2 the primers listed below that contain the 5'Ascl site for fusion to the N-terminal human GAA
3 sequence and a 3' NotI site for cloning into the NotI site of the pCEP vector.

Table 8. Human/Mouse GAA hybrids.

Human GAA portions	Linker Sequence	Mouse GAA portions (Human numbering)	Mouse GAA portions (Mouse numbering)	Forward Primer With 5' AseI site	Reverse Primer With 3' NotI site
1-790	GAP	791-952	792-953	gcggcgccgcGCTTCA TCCTTCAGATCTG C (SEQ ID NO:42)	ggcgccgcCTAGGA CCAGCTGATTG AAAC (SEQ ID NO:43)
1-796	GAP	797-952	798-953	gcggcgccgcGCTGTC CAGAGCAAGGGGC (SEQ ID NO:44)	ggcgccgcCTAGGA CCAGCTGATTG AAAC(SEQ ID NO:45)
1-816	GAP	817-952	818-953	gcggcgccgcCACCT GAGGGAGGGGTAC ATC(SEQ ID NO:46)	ggcgccgcCTAGGA CCAGCTGATTG AAAC(SEQ ID NO:47)
1-881	GAP	882-952	883-953	gcggcgccgcAACAA TACCATTGTGAAC AAG(SEQ ID NO:48)	ggcgccgcCTAGGA CCAGCTGATTG AAAC(SEQ ID NO:49)
1-920	GAP	921-952	922-953	gcggcgccgcATCCCT GTCTCCAATTCA CC(SEQ ID NO:50)	ggcgccgcCTAGGA CCAGCTGATTG AAAC(SEQ ID NO:51)

2
3 Mouse GAA nucleotide sequence (SEQ ID NO: 52):
4 ATGAATATACGGAAGCCCCCTGTTTCGAACTCCGTGGTTGGGGCCTGCACCCTTATC
5 TCTCTGACTACAGCGGTATCCTGGGTATCTCATGCTTCGGGAGTTAATGCTGCTT
6 CCCCAGACCTTCATGAGTCTCTCAGGACTGTGGAAGACGTACCGACCTACCA
7 CCAGGAAGGTTACAAGCCAGGCCTCTGCACATCCAGGAGCAGACTGAACAGCCC
8 AAAGAAGCACCCACACAGTGTGATGTGCCCCCAGCAGCCGCTTTGACTGTGCCCC
9 CGACAAAGGCATCTCACAGGAGCAATGCGAGGCCCGCGGCTGCTGCTATGTCCCAG
10 CAGGGCAGGTGCTGAAGGAGCCGAGATAGGGCAGCCCTGGTGTTCCTCCCTCCC
11 AGTACCCAAGCTACCGTCTAGAGAACCTGAGCTCTACAGAGTCGGGTACACAGC

CACCCTGACCCGTACCAGCCCCGACCTTCTTCCCAAAGGATGTGCTGACCTTACAGCT
GGAGGTGCTGATGGAGACAGACAGCCGCTCCACTTCAAGATCAAAGATCCTGCTA
GTAAGCGCTACGAAGTGCCCTGGAGACCCACGTGTGCTGAGCCAGGCACCATCC
CCACTTTACAGCGTGAATTCTCAGAGGAACCCTTTGGAGTGATCGTTCGTAGGAA
5 GCTTGGTGGCCGAGTGTGCTGAACACAACCGTGGCCCCCTGTTCTTCGCTGACCA
GTTCTGCAGCTGTCCACTTCCCTGCCCTCCAGCACATCACAGGCTGGGGGAACA
CCTCAGCCCACTCATGCTCAGCACCGACTGGGCTCGTATCACCTTCGGAACCGGG
ACACACCACCCTCGCAAGGTACCAACCTCTACGGGTCACATCCTTTCTACCTGGCAC
TGGAGGACGGTGGCTTGGCTCACGGTGTCTTCTGCTAAACAGCAATGCCATGGAT
10 GTCATCCTGCAACCCAGCCAGCCCTAACCTGGAGGTCAACGGGCGGGATCCTGGA
TGIGTATGIGTTCCTAGGCCAGAGCCCAAGAGCGTTGTGCAACAATACCTGGATG
TTGTGGGATACCCCTTCATGCCTCCATACTGGGGCTCGGCTTCCACCTCTGCCGCT
GGGGTACTCCTCGACCCGCAATGTCCGCCAGGTAGTGGAGAACATGACCAGGACA
CACTTCCCGCTGGATGTGCAATGGAATGACCTGGACTACATGGACGCCCGAAGAGA
15 CTTACCTTCAACCAGGACAGCTTTGCCGACTTCCCAGACATGGTGGGGAGCTGC
ACCAGGTGGCCGGCGCTACATGATGATCGTGGACCCTGCCATCAGCAGCGCAGGC
CCTGCTGGGAGTTACAGGCCCTACGACGAGGGTCTCGGAGGGGTGTGTTATCAC
CAACGAGACTGGCAGCCGCTGATTGGAAGGTTGGCCCCGAACCACCGCCTTCC
CTGATTTACCAACCCTGAGACCCTTGACTGGTGGCAGGACATGGTGTCTGAGITCC
20 ACGCCCAGGTGCCCTTCGATGGCATGTGGCTCGACATGAACGAACCGTCCAACITC
GTTAGAGGCTCTCAGCAGGGCTGCCCCAACAAATGAACTGGAGAACCCCCCTATGT
GCCCGGGGTGGTTGGCGGGATCTTGCAGGCAGCCACCATCTGTGCCTCCAGCCACC
AATTCTCTCCACACACTACAACCTCCACAACCTGTACGGCCTCACTGAAGCTATCG
CCTCCAGCAGGGCCCTGGTCAAGACTCGGGGAACACGACCCTTGTGATCTCCCGC
25 TCAACCTTCTCGGGCCACGGCCGTACGCTGGTCACTGGACAGGGGATGTGCGGAG
CTTTGGGAGCATCTTGATACTCTGTGCCAGACATCCTGCAGTTCAACCTGCTGGG
CGTCCCCCTGGTGGGGCGGACATCTGCGGCTTCATAGGAGACACGTCAGAAGAGC
TGTGTGTGCGCTGGACCAAGTTGGGGCCTTCTACCCCTTCATGCGGAACCACAATG
ACCTGAATAGCGTGCCCTCAGGAGCCGTACAGTTTCAGCGAGACGGCGCAGCAGGCC
30 ATGAGGAAGGCCTTCGCCTTACGCTATGCCCTTCTGCCCTACCTGTACACTCTCTTC
CACCGCGCCACGTCAGAGGAGACACGGTGGCCCCGGCCCTCTTCTGGAGTTCCC
TGAGGATCCCAGCACCTGGTCTGTGGACCGCCAGCTCTTGTGGGGCCGGCCCTGC

2005212435 23 Nov 2006

1 TCATCACACCTGTGCTTGAGCCTGGGAAAAGTGAAGTGACGGGCTACTTCCCCAAG
2 GGCACGTGGTACAACATGCAGGTGGTGTGTCAGTGGATTCCCTCGGACTCTCCCTTCT
3 CCATCATCGGCTTCATCCTTCAGATCTGCTGTCCAGAGCAAGGGGCAGTGGCTGAC
4 ACTGGAAGCCCCACTGGATACCATCAACGTGCACCTGAGGGAGGGGTACATCATAC
5 CGCTGCAGGGTCCCAGCCTCACAACCACGGAGTCCCGAAAGCAGCCCATGGCTCTG
6 GCTGTGGCATTAAACAGCAAGCGGCGAGGCCGATGGGGAGCTGTTCTGGGACGACG
7 GGGAGAGCCTTGCAGTTCTGGAGCGTGGGGCCTACACACTGGTCACCTTCTCAGCC
8 AAGAACAATACCATTGTGAACAAGTTAGTGCCTGTGACCAAGGAGGGAGCTGAGCT
9 ACAACTGAGGGAGGTGACCGT
0 CTTGGGAGTGGCCACAGCTCCTACCCAGGTCCTTTCCAACGGCATCCCTGTCTCCAA
1 TTTCACCTACAGCCCTGACAACAAGAGCCTGGCCATCCCTGTCTACTGCTGATGGG
2 AGAGCTGTTCAAATCAGCTGGTCTAG

3 Mouse GAA amino acid sequence (SEQ ID NO: 53):
4 MNIRKPLCSNSVVGACTLISLTTAVILGHLMLRELMLLPQDLHESSGLWKTYRPHHQE
5 GYKPGPLHIQEIQTEQPKEAPTQCDVPPSSRFDCAPDKGISQEQCEARGCCYVPAQVVK
6 EPQIQPWCFPPSPSYRLENLSSTESGYTATLRTSPTFFPKDVLTLQLEVLMTDSRL
7 HFKIKDPASKRYEVPLETPRVLSQAPSPLYSEVFSEEPFGVIVRRKLGGRVLLNTTVAPLF
8 FADQFLQLSTSLPSQHITGLGEHLSPLMLSTDWARITLWNRDTPPSQGTNLYGSHPFYLA
9 LEDGGLAHGVFLNNSNAMDVILQSPALTRWRSTGGILDVYVFLGPEPKSVVQQYLDVV
0 GYPFMPYWGFGFLCRWGYSSAIVRQVVENMTRTHFPLDVQWNDLDYMDARRDF
1 TFNQDSFADFPDMVRELHQGRRYMMIVDPAISSAGPAGSYRKYDEGLRRGVFITNETG
2 QPLIGKVVPGTTAFPDFTNPETLDWWQDMVSEFHAQVFPDGMWLDMNEPSNFVRSQ
3 QGCPNNELENPPYVPGVVGILQAATICASSHQFLSTHYNLHNLGLTEAIASSRALVK
4 TRGTRPFVISRSTFSGHGRYAGHWTGDVRSWEHLAYSVPDILQFNLLGVPLVGADICG
5 FIGDTSEELCVRWTQLGAFYPMRNHNDLNSVPQEPYRSETAQQAMRKAFALRYALL
6 PYLYTLFHRAHVRGDTVARPLFLEFPEDPSTWSVDRQLLWGPALLITPVLEPGKTEVTG
7 YFPKGTWYNMQVVSVDLGTLPSPSSASSFRSAVQSKGQWLTLEAPLDTINVHLREGYI
8 IPLQGPSLTTTESRKQPMALAVALTASGEADGELFWDDGESLAVLERGAYTLVTFSAKN
9 NTIVNKLVRVTKEGAELQLREVTVLGVATAPTQVLSNGIPVSNFTYSPDNKSLAIPVSL
0 MGELFQISWS.

2005212435 08 Aug 2006

1 [0131] The chimeric GAA cassettes were transfected into HEK 293 cells as described in
 2 Example 1. GAA expression levels were determined from two stable transfectants. As shown
 3 in Table 8, fusion at position 881 gives the highest enzyme expression levels. Western analysis
 4 of the position 881 fusion hybrid shows that the expressed precursor protein is of similar size to
 5 wild-type GAA.

6 Table 9. Human/Mouse GAA Hybrids Expression.

Fusion position	Stable GAA expression nmol/hr-ml Average of two lines
791	31
796	20
816	11
881	83
920	5

7
 8 [0132] Further experiments were carried out to determine if the presence of the mouse
 9 GAA sequence at the C-terminus of the hybrids was able to accommodate the presence of the
 0 GILT tag. Accordingly, the GILT Δ 1-7 tag was fused to the C-terminus of each of the five full-
 1 length human/mouse hybrids listed above and the expression levels were determined in each
 2 case. Constructs were also made to combine the C-terminal position 881 mouse GAA hybrid
 3 with an N-terminal GILT tag at positions 29, 56, 70, or 81. The expression levels were
 4 determined as described above.

15
 16 [0133] Throughout this specification and the claims which follow, unless the context
 17 requires otherwise, the word "comprise", and variations such as "comprises" and "comprising",
 18 will be understood to imply the inclusion of a stated integer or step or group of integers or steps
 19 but not the exclusion of any other integer or step or group of integers or steps.

20 [0134] The reference to any prior art in this specification is not, and should not be taken
 21 as, an acknowledgement or any form of suggestion that the prior art forms part of the common
 22 general knowledge in Australia.

23 [0135] We claim:

CLAIMS

2005212435 09 Aug 2010

1. A targeted therapeutic fusion protein comprising a therapeutic agent comprising amino acid residues 70-952 of human GAA, residues 1-69 being deleted, and a peptide tag, wherein the peptide tag is directly or indirectly linked to amino acid residue 70 of human GAA and said tag comprises IGF-II residue 1 followed by residues 8-67, residues 2-7 being deleted.
5
2. The targeted therapeutic of claim 1 further comprising a spacer between the therapeutic agent and the peptide tag.
10
3. The targeted therapeutic of claim 2 wherein the spacer is a Gly-Ala-Pro linker.
4. The targeted therapeutic of any one of the preceding claims for use in the treatment of Pompe disease.
15
5. The targeted therapeutic fusion protein according to any one of claims 1-3 and substantially as herein described with reference to the Examples and/or Drawings.

2005212435 09 Aug 2010

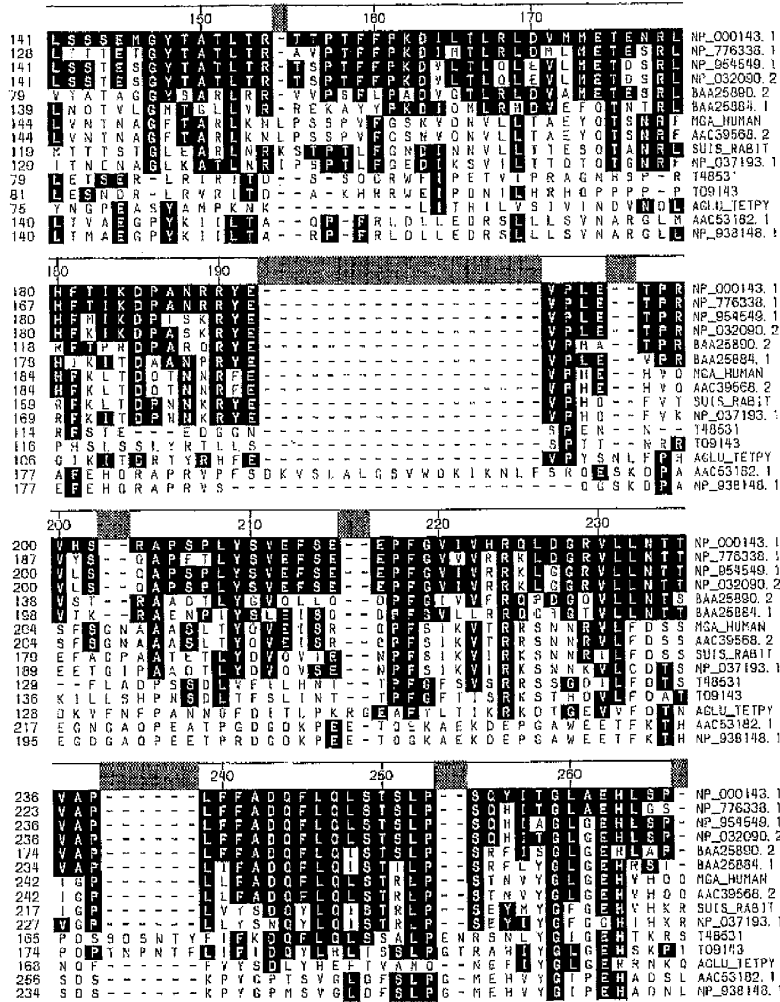


Figure 1-2

2005212435 09 Aug 2010

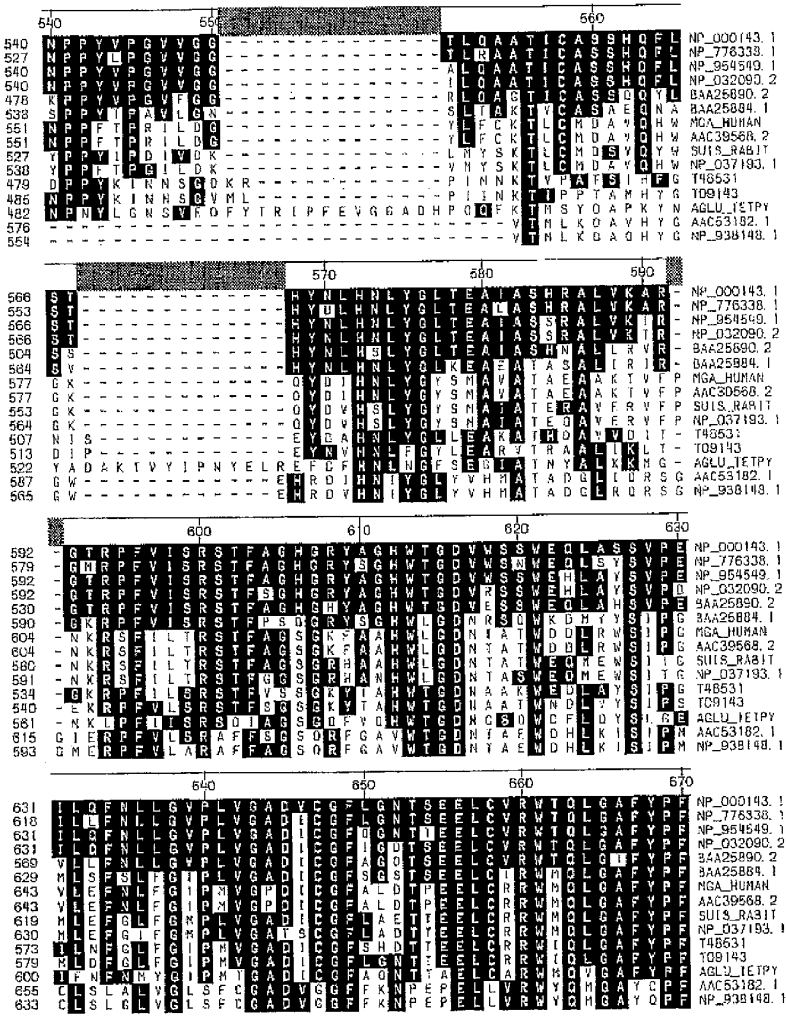


Figure 1-5

2005212435 09 Aug 2010

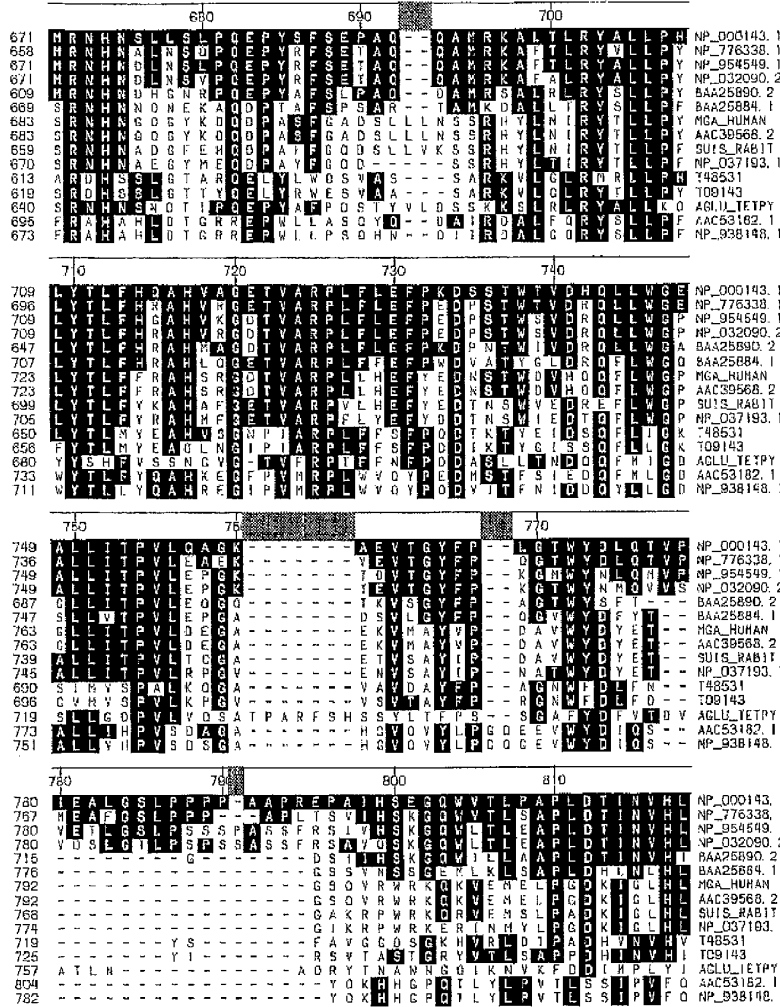


Figure 1-6

2005212435 09 Aug 2010

```

952 NP_000143.1
957 NP_776338.1
953 NP_954549.1
953 NP_032090.2
873 BAA25890.2
932 BAA25884.1
959 RDEEEKIUCYPDENGASAEENCTARDCIWEASNS-SGVVFCY HGA_HUMAN
959 RDEEEKIUCYPDENGASAEENCTARDCIWEASNS-SGVVFCY AAC39568.2
936 LESECKIUCYPDADIAIDAEKCTQRDCIWDITNTVWPRAPFCY SUI5_RAB11
942 SDNREKFTCYPDYGTATEGTCTQRDCIWDQVPSGLSNVPPYY NP_037193.1
902 T48531
903 T09143
923 AGLU_1ETPY
966 AAC53182.1
944 NP_938148.1

```

```

952 NP_000143.1
937 NP_776338.1
953 NP_954549.1
953 NP_032090.2
873 BAA25890.2
932 BAA25884.1
956 FVKK--DLYSYSDVQYNSHGATADISLKSYYANAFPSIPY HGA_HUMAN
956 FVKK--DLYSYSDVQYNSHGATADISLKSYYANAFPSIPY AAC39568.2
976 FPKTDNPPYSVSSIQYSPFGITADLQLNPIRIRITLPSSEPI SUI5_RAB11
982 FPPENNPYTLTISIQPLPIGITAELQLNPPNARIKLPSNPFI NP_037193.1
902 T48531
903 T09143
923 AGLU_1ETPY
966 AAC53182.1
944 NP_938148.1

```

```

952 NP_000143.1
937 NP_776338.1
953 NP_954549.1
953 NP_032090.2
873 BAA25890.2
932 BAA25884.1
1038 NPLRLLDVITYHKNEHLQFKIYDPPNKNRYEVVPLNIPSPNPS HGA_HUMAN
1038 NPLRLLDVITYHKNEHLQFKIYDPPNKNRYEVVPLNIPSPNPS AAC39568.2
1016 INLHVVEVKYHKNDMVOFKIYDPPNKNRYEVVPLNIPSPNPS SUI5_RAB11
1022 SYLRVGVKVKPNQMLQFKIYDAOHKRYEVVPLNIPSPNPS NP_037193.1
902 T48531
903 T09143
923 AGLU_1ETPY
966 AAC53182.1
944 NP_938148.1

```

```

952 NP_000143.1
937 NP_776338.1
953 NP_954549.1
953 NP_032090.2
873 BAA25890.2
932 BAA25884.1
1076 STPEGQLYDYLKKNPFGEIARRKSTGTIIWDSQLLGFIF HGA_HUMAN
1076 STPEGQLYDYLKKNPFGEIARRKSTGTIIWDSQLLGFIF AAC39568.2
1086 STOENRLYDYEIKENPFGEIDJRRRSTGKVIWDSCLYDFAF SUI5_RAB11
1082 SSNE-RLYDYEIKENPFGEIOVRRRHSCKLIWDSRLPBFDF NP_037193.1
902 T48531
903 T09143
923 AGLU_1ETPY
966 AAC53182.1
944 NP_938148.1

```

Figure 1-8

2005212435 09 Aug 2010

952		NP_000143.1
937		NP_776338.1
953		NP_964649.1
953		NP_032090.2
873		BAA25890.2
932		BAA25889.1
1116	SOMFIRISTRRLPSKYL YGFGE TEHRSYRRDLEWHTWGMFS	NGA_HUMAN
1116	SOMFIRISTRRLPSKYL YGFGE TEHRSYRRDLEWHTWGMFS	AAC39668.2
1096	NDOFIOISTRRLPSEYIYGFGEAEHTAFKRDLNWHWTWGMFT	SUIS_RAB1T
1101	NDOFIOISTRRLPSHYLYGFGEVEHTAFKRDLNWHWTWGMFT	NP_037193.1
902		T48531
903		T09143
923		AGLU_IETPY
966		AAC53182.1
944		NP_938148.1

952		NP_000143.1
937		NP_776338.1
953		NP_964649.1
953		NP_032090.2
873		BAA25890.2
932		BAA25889.1
1156	RDOPPGYKKNSTYGVHPYYMGL EEDGSAHG VLL LNSNANDY	NGA_HUMAN
1156	RDOPPGYKKNSTYGVHPYYMGL EEDGSAHG VLL LNSNANDY	AAC39668.2
1136	RDOPPGYKKNSTYGFHPYYMALEDEGN AHG VLL LNSNANDY	SUIS_RAB1T
1141	RDOPPGYKKNSTYGFHPYYHALENEGN AHG VLL LNSNANDY	NP_037193.1
902		T48531
903		T09143
923		AGLU_IETPY
966		AAC53182.1
944		NP_938148.1

952		NP_000143.1
937		NP_776338.1
953		NP_964649.1
953		NP_032090.2
873		BAA25890.2
932		BAA25889.1
1186	TFQPLPALTYRTIGGVLDFYVFLGPTPELVITQQYTELIGR	NGA_HUMAN
1196	TFQPLPALTYRTIGGVLDFYVFLGPTPELVITQQYTELIGR	AAC39668.2
1176	TFMPLPALTYRVIGGILDFYMFIGPTPEVAITQDYHEVIGH	SUIS_RAB1T
1181	TFQPLPALTYRTIGGILDFYHFLGPTPEIATRQDYHEVIGF	NP_037193.1
902		T48531
903		T09143
923		AGLU_IETPY
966		AAC53182.1
944		NP_938148.1

952		NP_000143.1
937		NP_776338.1
953		NP_964649.1
953		NP_032090.2
873		BAA25890.2
932		BAA25889.1
1236	PVHVPPYNSLGFGLCRYGYONDSEIASLYDEMVAADIPYDY	NGA_HUMAN
1236	PVHVPPYNSLGFGLCRYGYONDSEIASLYDEMVAADIPYDY	AAC39668.2
1216	PVXPPYNSLGFGLCRYGYRNTSEIIELYEGMVAADIPYDY	SUIS_RAB1T
1221	PVXPPYWAIGFGLCRYGYRNTSEIEOLYNDHVAANIIPYDY	NP_037193.1
902		T48531
903		T09143
923		AGLU_IETPY
966		AAC53182.1
944		NP_938148.1

Figure 1-9

2005212435 09 Aug 2010

952		NP_000143.1
937		NP_776338.1
953		NP_954549.1
953		NP_032080.2
873		BAA25890.2
932		BAA25884.1
1276	QYSDIDYHERQLDCLFTLSPKFACFPALINRMKADGMRVLLI	NGA_HUMAN
1276	QYSDIDYHERQLDCLFTLSPKFACFPALINRMKADGMRVLLI	AAC39568.2
1256	QYTDIDYHERQLDCLFTLDENFRELPOFVDRIRGEGMRYIII	SUIS_RAB1T
1261	QYTDINMYHERQLDCLFTIGERFKJLPEFVDRIRKDGMRKYIV	NP_037193.1
902		T48531
903		T09143
923		AGLU_IETPY
966		AAC53182.1
944		NP_938148.1
952		NP_000143.1
937		NP_776338.1
953		NP_954549.1
953		NP_032080.2
873		BAA25890.2
932		BAA25884.1
1316	LDPALISGNETOPYPAFTRRGVFDVDFIKYPNDGDIWCKKVV	NGA_HUMAN
1316	LDPALISGNETOPYPAFTRRGVEDDVFIKYPNDGDIWCKKVV	AAC39568.2
1296	LDPALISGNETRPPYPAFDRGEAKDVFVKWPNTSDICWAKKVV	SUIS_RAB1T
1301	LAPALISGNETOPYPAFERGSIQKDVFKWPNTINDICWPKVV	NP_037193.1
902		T48531
903		T09143
923		AGLU_IETPY
966		AAC53182.1
944		NP_938148.1
952		NP_000143.1
937		NP_776338.1
953		NP_954549.1
953		NP_032080.2
873		BAA25890.2
932		BAA25884.1
1356	PDFPDVYVNGSLDWDSSQVELYRAYVAFPDFFRNMSTAKWVK	NGA_HUMAN
1356	PDFPDVYVNGSLDWDSSQVELYRAYVAFPDFFRNMSTAKWVK	AAC39568.2
1336	PDLPNITIDEELTDFAYNASRAHAAFPDFFRNMSTAEWWT	SUIS_RAB1T
1341	PDLPNYITIDEETI)EEDAYNASRAHVAFPDFFRNMSTLEWVA	NP_037193.1
902		T48531
903		T09143
923		AGLU_IETPY
966		AAC53182.1
944		NP_938148.1
952		NP_000143.1
937		NP_776338.1
953		NP_954549.1
953		NP_032080.2
873		BAA25890.2
932		BAA25884.1
1396	REIEEELYNPNPCHPERSLKKFDGMWIDMNEPSSFF--YHGAV	NGA_HUMAN
1396	REIEEELYNPNPCHPERSLKKFDGMWIDMNEPSSFF--YHGAV	AAC39568.2
1376	REILDFYN-----NYMKFDGLWIDMNEPSSFF--YHGTT	SUIS_RAB1T
1381	REIYDFYN-----EKMKFDGLWIDMNEPSSFFGIOMGCKY	NP_037193.1
902		T48531
903		T09143
923		AGLU_IETPY
966		AAC53182.1
944		NP_938148.1

Figure 1-10

2005212435 09 Aug 2010

952		NP_000143.1
937		NP_776338.1
953		NP_954549.1
953		NP_032090.2
873		BAA25890.2
932		BAA25884.1
1433	S P G C R - D A S L N H P P Y M P H L E S R D R G - - - L S S K T L C M E S Q Q	MGA_HUMAN
1433	S P G C R - D A S L N H P P Y M P H L E S R D R G - - - L S S K T L C M E S Q Q	AAC39568.2
1407	T N Y C R - N T E L N Y P P Y F P E L T K R T D G - - - L H F R T M C M E T E H	SUIS_RAB11
1416	L N E C R R M M I L N Y P P Y F S P E L R V K E G E G A S I S E A M C M E T E H	NP_037193.1
902		T48531
903		T09143
923		AGLU_TETPY
966		AAC53182.1
944		NP_938148.1
952		NP_000143.1
937		NP_776338.1
953		NP_954549.1
953		NP_032090.2
873		BAA25890.2
932		BAA25884.1
1469	I L P D G S L V O H Y N V H N L Y G W S D I R P T Y E A V D E V T G O R G V V I	MGA_HUMAN
1469	I L P D G S L V O H Y N V H N L Y G W S D I R P T Y E A V D E V T G O R G V V I	AAC39568.2
1443	I L S D G S S V L H Y D V H N L Y G W S Q A K P T Y D A L Q K T T G K R G I V I	SUIS_RAB11
1465	I L I D G S S V L Q Y D V H N L Y G W S Q V K P T L D A L C N T T G L R G I V I	NP_037193.1
902		T48531
903		T09143
923		AGLU_TETPY
966		AAC53182.1
944		NP_938148.1
952		NP_000143.1
937		NP_776338.1
953		NP_954549.1
953		NP_032090.2
873		BAA25890.2
932		BAA25884.1
1509	T R S T F P S S G R W A C H W L G D N T A A W D Q L K K S I I G H M E F S L F G	MGA_HUMAN
1509	T R S T F P S S G R W A C H W L G D N T A A W D Q L K K S I I G H M E F S L F G	AAC39568.2
1483	S R S T Y P T A G R W A C H W L G D N Y A R W D N M D K S I I G H M E F S L F G	SUIS_RAB11
1495	S R S T Y P T T G R W C G H W L G D N Y T T W D N L E K S L I G H L E L N L F G	NP_037193.1
902		T48531
903		T09143
923		AGLU_TETPY
966		AAC53182.1
944		NP_938148.1
952		NP_000143.1
937		NP_776338.1
953		NP_954549.1
953		NP_032090.2
873		BAA25890.2
932		BAA25884.1
1549	I S Y T G A D I C G F F Q D A E Y - E M C V R W M O L G A F Y P F S R N - H N T	MGA_HUMAN
1549	I S Y T G A D I C G F F Q D A E Y - E M C V R W M O L G A F Y P F S R N - H N T	AAC39568.2
1523	I S Y T G A D I C G F F N D S E Y - H L C T R W T Q L Q A F Y P F A R N - H N I	SUIS_RAB11
1535	I P Y I G A D I C G V F H D S G Y P S L Y F Y G I O Y G A F Y P Y P R E S P T I	NP_037193.1
902		T48531
903		T09143
923		AGLU_TETPY
966		AAC53182.1
944		NP_938148.1

Figure 1-11

2005212435 09 Aug 2010

952 NP_000143.1
 937 NP_776338.1
 953 NP_954549.1
 873 NP_032090.2
 932 BAA25890.2
 1587 ICTRRRQDPVSWOVAFVNIISRTVLDQTRYTLIPYLYTLMHKA MGA_HUMAN
 1587 IGTTRRDQDPVSWOVAFVNIISRTVLDQTRYTLIPYLYTLMHKA AAC39568.2
 1561 QFTTRRDQDPVSWOVAFVNIISRTVLDQTRYTLIPYLYTLMHKA SUIS_RAB1T
 1576 NFFTRSDQDPVSWMKLLLDMSKKVLEIRYTLIPYLYTLMHKA NP_037193.1
 902 T48531
 903 TQ9143
 923 AGLU_TETPY
 966 AAC63182.1
 944 NP_938148.1

952 NP_000143.1
 937 NP_776338.1
 953 NP_954549.1
 873 NP_032090.2
 932 BAA25890.2
 1627 HTECVTVVVRPLLHEFFVSDQVITWDIDISDFLLGPAFLVSPYL MGA_HUMAN
 1627 HTECVTVVVRPLLHEFFVSDQVITWDIDISDFLLGPAFLVSPYL AAC39568.2
 1611 HAHCQGTVIRPLMHEFFVDQRTITWDIFLDQFLWGP AFMYIPVL SUIS_RAB1T
 1615 HAHCQGTVIRPLMHEFFVDQRTITWDIFLDQFLWGP AFMYIPVL NP_037193.1
 902 T48531
 903 TQ9143
 923 AGLU_TETPY
 966 AAC63182.1
 944 NP_938148.1

952 NP_000143.1
 937 NP_776338.1
 953 NP_954549.1
 873 NP_032090.2
 932 BAA25890.2
 1667 ERNARNVYTAFFPRARWYDYIIGVDINARGEWKITLPAPLDH MGA_HUMAN
 1667 ERNARNVYTAFFPRARWYDYIIGVDINARGEWKITLPAPLDH AAC39568.2
 1641 EPYTITVVRGCVYVPKARWFDYHTGSDIGIRSDYQDLTLLMNA SUIS_RAB1T
 1656 EPTFRITSVITGVYVPKARWFDYHTGSDIGIRSDYQDLTLLMNA NP_037193.1
 902 T48531
 903 TQ9143
 923 AGLU_TETPY
 966 AAC63182.1
 944 NP_938148.1

952 NP_000143.1
 937 NP_776338.1
 953 NP_954549.1
 873 NP_032090.2
 932 BAA25890.2
 1707 INLHVRCGGYIILPWOEPALNTHLSRQKFMGFKJALDDEGTA MGA_HUMAN
 1707 INLHVRCGGYIILPWOEPALNTHLSRQKFMGFKJALDDEGTA AAC39568.2
 1681 INLHVRCGGYIILPWOEPARTTIFLSRQKYMKLLIYAADDNQHHA SUIS_RAB1T
 1695 INLHVRCGGYIILPWOEPARNTHLSRQNYMKLLIYAADDNQHHA NP_037193.1
 902 T48531
 903 TQ9143
 923 AGLU_TETPY
 966 AAC63182.1
 944 NP_938148.1

Figure 1-12

2005212435 09 Aug 2010

952		NP_000143.1
937		NP_776338.1
953		NP_954548.1
873		NP_032080.2
932		BAA25890.2
1747	GGWLFWDGQSSJDTYGGKGLYYLASFSASONTMQSHIIFNN	BAA25894.1
1747	GGWLFWDGQSSJDTYGGKGLYYLASFSASONTMQSHIIFNN	NGA_HUMAN
1721	GGSLFWDDGDTJIDYERDLYLSVQFNLNKTTLLTSTLLKTG	AAC39568.2
1735	GGYLFWDGDESIIDTYERDLYLSVQFNLNKTTLLTSTYLANG	SUIS_RAB1T
902		NP_037193.1
903		T48531
923		T09143
966		AGLU_1ETPY
944		AAC53182.1
		NP_938148.1

952		NP_000143.1
937		NP_776338.1
953		NP_954548.1
873		NP_032080.2
932		BAA25890.2
1767	YITCTHPLKLGYTEIHWGYGSPVVTSSVSYSGMYITPSFN	BAA25894.1
1767	YITCTHPLKLGYTEIHWGYGSPVVTSSVSYSGMYITPSFN	NGA_HUMAN
1761	YJN-KTEIRLSYVHWGIGNTLINENLNYNEINYPLIFN	AAC39568.2
1775	YKN-KDEMRLLSSIHVWCKGTLRISNANLVYBGRKHQPPFI	SUIS_RAB1T
902		NP_037193.1
903		T48531
923		T09143
966		AGLU_1ETPY
944		AAC53182.1
		NP_938148.1

952		NP_000143.1
937		NP_776338.1
953		NP_954548.1
873		NP_032080.2
932		BAA25890.2
1827	NOPTTIOVLSIDVYDRNHSLSHNFSTSLTWLSTL	BAA25894.1
1827	NOPTTIOVLSIDVYDRNHSLSHNFSTSLTWLSTL	NGA_HUMAN
1800	QTQAQELNIDLTAEVTLDDPIEISWS	AAC39568.2
1814	QEEAKETLIFDLKNNMNVILDEPIDITWS	SUIS_RAB1T
902		NP_037193.1
903		T48531
923		T09143
966		AGLU_1ETPY
944		AAC53182.1
		NP_938148.1

Decoration 'Decoration #1': Shado (with solid black) residues that match NP_000143.1 exactly.

Figure 1-13

2005212435 09 Aug 2010

Key for alignments:

NP_000143.1 :acid alpha-glucosidase preproprotein; lysosomal alpha-glucosidase; acid maltase [Homo sapiens]. (SEQ ID NO: 54)
NP_776338.1 glucosidase, alpha; acid [Bos Taurus]. (SEQ ID NO: 55)
NP_954549.1 glucosidase, alpha; acid (Pompe disease, glycogen storage disease type II) [Rattus norvegicus]. (SEQ ID NO: 56)
NP_032090.2 glucosidase, alpha, acid [Mus musculus]. (SEQ ID NO: 57)
BAA25890.2 acid alpha glucosidase [Coturnix japonica]. (SEQ ID NO: 58)
BAA25884.1 another acid alpha glucosidase [Coturnix japonica]. (SEQ ID NO: 59)
MGA_HUMAN Maltase-glucoamylase, intestinal [Includes: Maltase (Alpha-glucosidase); Glucoamylase (Glucan 1,4-alpha-glucosidase)]. (SEQ ID NO: 60)
AAC39568.2 maltase-glucoamylase [Homo sapiens]. (SEQ ID NO: 61)
SUITS_RABIT Sucrase-isomaltase, intestinal Oryctolagus cuniculus (rabbit). (SEQ ID NO: 62)
NP_037193.1 sucrase-isomaltase [Rattus norvegicus]. (SEQ ID NO: 63)
T48531 alpha-glucosidase 1 - Arabidopsis thaliana. (SEQ ID NO: 64)
T09143 alpha-glucosidase (EC 3.2.1.20) - spinach. (SEQ ID NO: 65)
AGLU_TETPY LYSOSOMAL ACID ALPHA-GLUCOSIDASE PRECURSOR (ACID MALTASE). (SEQ ID NO: 66)
AAC53182.1 alpha glucosidase II, alpha subunit [Mus musculus]. (SEQ ID NO: 67)
ACCESSION AAC53182
NP_938148.1 alpha glucosidase II alpha subunit [Homo sapiens]. (SEQ ID NO: 68)

Figure 1 - 14

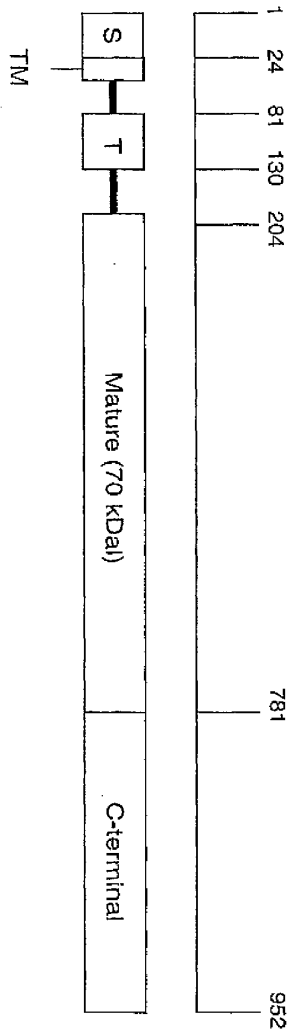


Figure 2

2005212435 09 Aug 2010

2005212435 09 Aug 2010

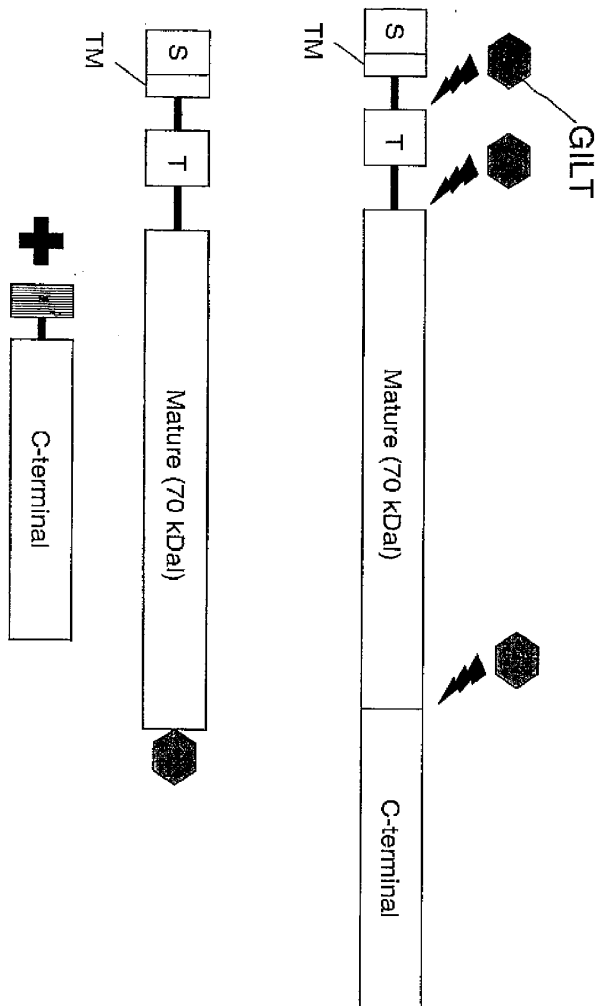


Figure 3

2005212435 09 Aug 2010

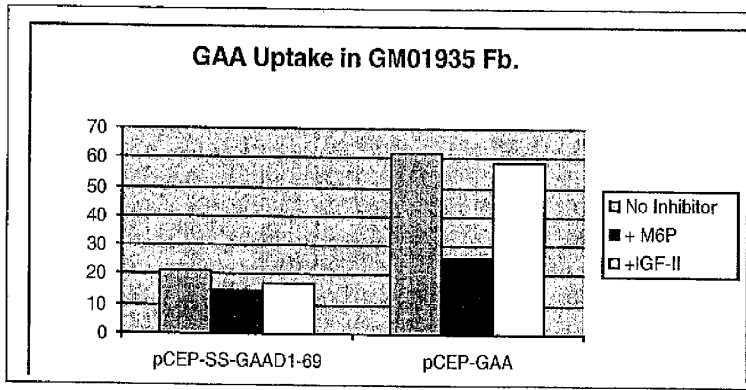


FIGURE 4

2005212435 09 Aug 2010

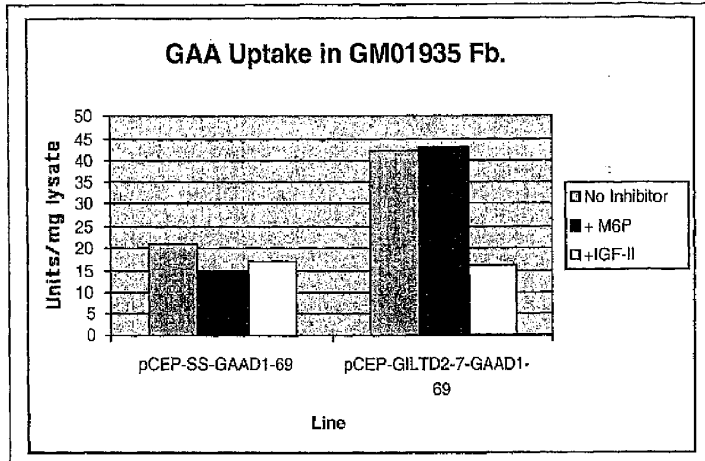


FIGURE 5

2005212435 09 Aug 2010

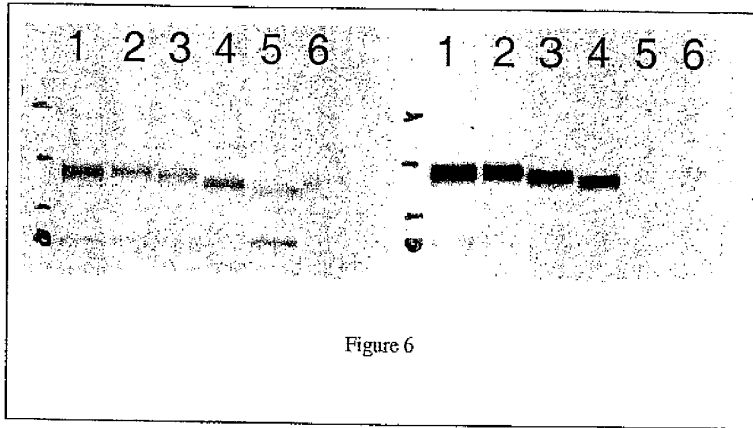


Figure 6

2005212435 09 Aug 2010

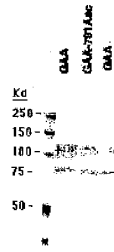


Figure 7

2005212435 09 Aug 2010

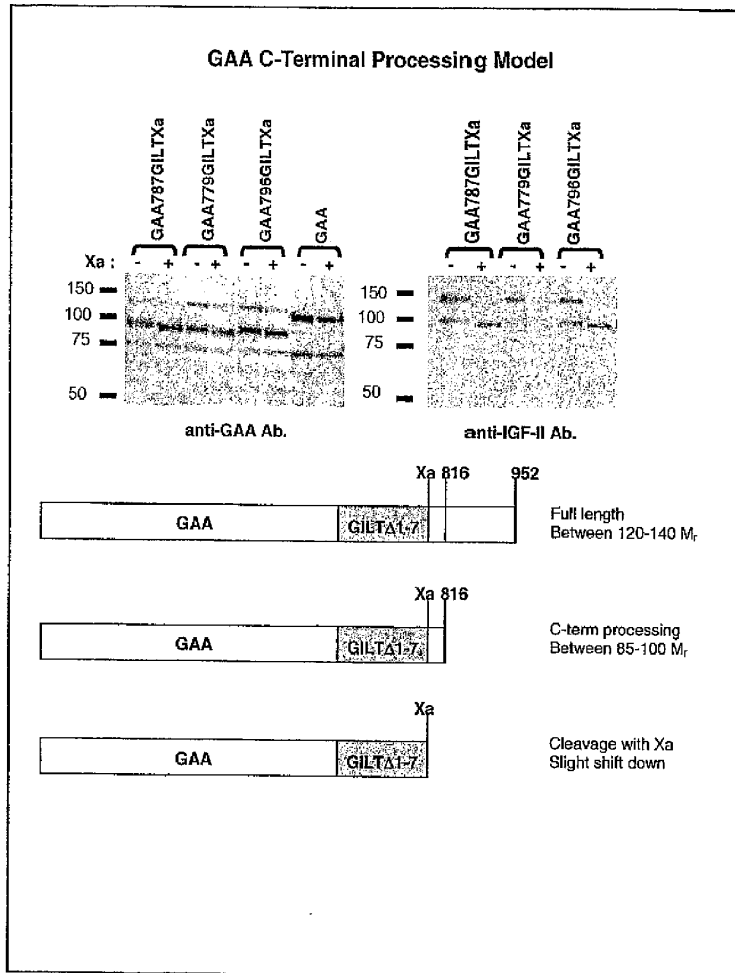


Figure 8