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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 January 2014 (03.01.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/005019 A2

(51) International Patent Classification:
C12N 9/16 (2006.01)

(21) International Application Number:
PCT/US2013/048571

(22) International Filing Date:
28 June 2013 (28.06.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/666,719 29 June 2012 (29.06.2012) US

(71) Applicant: **SHIRE HUMAN GENETIC THERAPIES, INC.** [US/US]; 300 Shire Way, Lexington, Massachusetts 02421 (US).

(72) Inventors: **BOLGOG, Ferenc**; 300 Shire Way, Lexington, Massachusetts 02421 (US). **HEARTLEIN, Mike**; 167 Reed Farm Road, Boxborough, Massachusetts 01719 (US).

(74) Agent: **HUDDLESON, Justin P.**; Choate, hall & Stewart LLP, Two International Place, Boston, Massachusetts 02110 (US).

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

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

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2014/005019 A2

(54) Title: CELLS FOR PRODUCING RECOMBINANT IDURONATE-2-SULFATASE

(57) Abstract: The present invention provides, among other things, methods and compositions for production of recombinant I2S protein with improved potency and activity using cells co- express I2S and FGE protein. In some embodiments, cells according to the present invention are engineered to simultaneously over-express recombinant I2S and FGE proteins. Cells according to the invention are adaptable to various cell culture conditions. In some embodiments, cells of the present invention adaptable to a large-scale suspension serum- free culture.

CELLS FOR PRODUCING RECOMBINANT IDURONATE-2-SULFATASE

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application Serial No. 61/666,719, filed June 29, 2012; the entirety of which is hereby incorporated by reference.

SEQUENCE LISTING

[0002] The present specification makes reference to a Sequence Listing submitted in electronic form as an ASCII .txt file named "2006685-00340_SEQ_LIST" on June 27, 2013. The .txt file was generated on June 25, 2013 and is 25 KB in size. The entire contents of the Sequence Listing are herein incorporated by reference.

BACKGROUND

[0003] Mucopolysaccharidosis type II (MPS II, Hunter syndrome) is an X-chromosome-linked recessive lysosomal storage disorder that results from a deficiency in the enzyme iduronate-2-sulfatase (I2S). I2S cleaves the terminal 2-O-sulfate moieties from the glycosaminoglycans (GAG) dermatan sulfate and heparan sulfate. Due to the missing or defective I2S enzyme in patients with Hunter syndrome, GAG progressively accumulate in the lysosomes of a variety of cell types, leading to cellular engorgement, organomegaly, tissue destruction, and organ system dysfunction.

[0004] Generally, physical manifestations for people with Hunter syndrome include both somatic and neuronal symptoms. For example, in some cases of Hunter syndrome, central nervous system involvement leads to developmental delays and nervous system problems. While the non-neuronal symptoms of Hunter Syndrome are generally absent at birth, over time the progressive accumulation of GAG in the cells of the body can have a dramatic impact on the peripheral tissues of the body. GAG accumulation in the peripheral tissue leads to a distinctive coarseness in the facial features of a patient and is responsible for the prominent forehead, flattened bridge and enlarged tongue, the defining hallmarks of a Hunter patient. Similarly, the accumulation of GAG can adversely affect the organ systems of the body. Manifesting initially as a thickening of the wall of the heart, lungs and airways,

and abnormal enlargement of the liver, spleen and kidneys, these profound changes can ultimately lead to widespread catastrophic organ failure. As a result, Hunter syndrome is always severe, progressive, and life-limiting.

[0005] Enzyme replacement therapy (ERT) is an approved therapy for treating Hunter syndrome (MPS II), which involves administering exogenous replacement I2S enzyme to patients with Hunter syndrome.

SUMMARY OF THE INVENTION

[0006] The present invention provides, among other things, improved methods and compositions for production of recombinant I2S protein that allows more effective enzyme replacement therapy for Hunter syndrome. The present invention encompasses the discovery that more potent recombinant I2S protein can be produced by mammalian cells engineered to co-express a recombinant I2S protein and a formylglycine generating enzyme (FGE). Unexpectedly, recombinant I2S protein produced by such engineered cells has an unusually high level of C_α-formylglycine (FGly) conversion percentage (e.g., greater than 70% and up to 100%), resulting in significantly improved enzymatic activity of recombinant I2S protein. In addition, mammalian cells co-expressing I2S and FGE proteins according to the present invention have been successfully adapted to grow in suspension culture at a large scale. Therefore, the present invention allows more efficient large scale production of highly potent recombinant I2S protein.

[0007] Thus, in one aspect, the present invention provides a cell containing a first nucleic acid encoding an iduronate-2-sulfatase (I2S) protein having an amino acid sequence at least about 50% (e.g., at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%) identical to SEQ ID NO:1; and a second nucleic acid encoding a formylglycine generating enzyme (FGE) protein comprising an amino acid sequence at least about 50% (e.g., at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%) identical to SEQ ID NO:5, wherein the first and/or the second nucleic acid are exogenous and wherein the cell, once cultivated under a cell culture condition (e.g., suspension or adherent culture), produces the I2S protein comprising at least about 70% (e.g., at least about 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) conversion of the cysteine residue corresponding to Cys59 of SEQ ID NO:1 to Cα-formylglycine (FGly).

[0008] In another aspect, the present invention provides a cell containing a first nucleic acid encoding an iduronate-2-sulfatase (I2S) protein having an amino acid sequence at least about 50% (e.g., at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%) identical to SEQ ID NO:1; and a second nucleic acid encoding a formylglycine generating enzyme (FGE) protein comprising an amino acid sequence at least about 50% (e.g., at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%) identical to SEQ ID NO:5, wherein the first and/or the second nucleic acid are exogenous and wherein the cell, once cultivated under a cell culture condition, produces I2S protein comprising at least about 50% (e.g., at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) conversion of the cysteine residue corresponding to Cys59 of SEQ ID NO:1 to Cα-formylglycine (FGly) and at a specific productivity rate of great than about 10 picogram/cell/day (e.g., greater than about 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 picogram/cell/day).

[0009] In some embodiments, the first nucleic acid encodes an I2S protein having an amino acid sequence identical to SEQ ID NO:1. In some embodiments, the second nucleic acid encodes an FGE protein having an amino acid sequence identical to SEQ ID NO:5.

[0010] In some embodiments, the first and/or the second nucleic acid is operably linked to a hCMV promoter.

[0011] In some embodiments, the first and/or second nucleic acid are codon optimized. In some embodiments, the first nucleic acid has a sequence at least about 50% (e.g., at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%) identical to SEQ ID NO:7. In particular embodiments, the first nucleic acid has a sequence of SEQ ID NO:7.

[0012] In some embodiments, the second nucleic acid comprises a sequence at least about 50% (e.g., at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%) identical to SEQ ID NO:8. In some embodiments, the second nucleic acid has a sequence identical to SEQ ID NO:8.

[0013] In some embodiments, both of the first and second nucleic acids are exogenous (also referred to as recombinant). In some embodiments, the first and/or second nucleic acids are integrated (e.g., stably) in the genome of the cell. In some embodiments,

the first and/or second nucleic acids are present in one or more extra-chromosomal constructs.

[0014] In some embodiments, a cell of the present invention is a mammalian cell. In certain embodiments, a suitable mammalian cell is a human cell. In certain embodiments, a suitable mammalian cell is a CHO cell.

[0015] In some embodiments, a cell according to the invention is adaptable to suspension culture. In other embodiments, a cell according to the invention is adherent.

[0016] In a further aspect, the present invention provides a method of producing recombinant iduronate-2-sulfatase (I2S) protein by cultivating a cell described in various embodiments herein under conditions such that the recombinant I2S and FGE proteins are co-expressed in the cell. In some embodiments, the cell is cultivated at a large scale. In some embodiments, a large scale suitable for the present invention is a bioreactor process. In some embodiments, a bioreactor suitable for the invention is at a scale selected from 10L, 200L, 500L, 1000L, 1500L, 2000L. In some embodiments, a large scale (e.g., bioreactor) process suitable for the present invention involves a perfusion process. In some embodiments, a large scale (e.g., bioreactor) process suitable for the present invention involves a batch culture. In some embodiments, a large scale process suitable for the present invention is a roller bottle process. In some embodiments, a cell according to the present invention is cultivated in suspension. In other embodiments, a cell according to the present invention is cultivated adherent.

[0017] In some embodiments, a cell according to the present invention is cultivated in a serum-free medium (e.g., animal-free, chemically-defined, or protein-free medium). In other embodiments, a cell according to the present invention is cultivated in a serum-containing medium.

[0018] In various embodiments, a method according to the invention further includes a step of purifying the recombinant I2S protein.

[0019] In still another aspect, the present invention provides a recombinant iduronate-2-sulfatase (I2S) protein produced by a cell or method described in various embodiments herein.

[0020] In some embodiments, the present invention provides a preparation of recombinant iduronate-2-sulfatase (I2S) protein, in which said recombinant I2S protein has an amino acid sequence at least about 50% (e.g., at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%) identical to SEQ ID NO:1; and containing at least about 70% (e.g., at least about 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 100%) conversion of the cysteine residue corresponding to Cys59 of SEQ ID NO:1 to C_α-formylglycine (FGly). In some embodiments, the recombinant I2S protein has an amino acid sequence identical to SEQ ID NO:1. In some embodiments, the recombinant I2S protein has specific activity of at least about 20 U/mg, 30 U/mg, 40 U/mg, 50 U/mg, 60 U/mg, 70 U/mg, 80 U/mg, 90 U/mg, or 100 U/mg as determined by an *in vitro* sulfate release activity assay using heparin disaccharide as substrate.

[0021] Among other things, the present invention also provides a pharmaceutical composition containing a recombinant I2S protein described in various embodiments herein and a pharmaceutically acceptable carrier and a method of treating Hunter syndrome by administering into a subject in need of treatment recombinant I2S protein described herein or a pharmaceutical composition containing the same.

[0022] As used herein, the terms "I2S protein," "I2S," "I2S enzyme," or grammatical equivalents, refer to a preparation of recombinant I2S protein molecules unless otherwise specifically indicated.

[0023] As used in this application, the terms "about" and "approximately" are used as equivalents. Any numerals used in this application with or without about/approximately are meant to cover any normal fluctuations appreciated by one of ordinary skill in the relevant art.

[0024] Other features, objects, and advantages of the present invention are apparent in the detailed description that follows. It should be understood, however, that the detailed description, while indicating embodiments of the present invention, is given by way of illustration only, not limitation. Various changes and modifications within the scope of the invention will become apparent to those skilled in the art from the detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] The Figures described below, that together make up the Drawings, are for illustration purposes only, not for limitation.

[0026] *Figure 1* depicts the amino acid sequence (SEQ ID NO:1) encoding the mature form of human iduronate-2-sulfatase (I2S) protein and indicates potential sites within the protein sequence for N-linked glycosylation and cysteine conversion.

[0027] *Figure 2* depicts exemplary construct designs for co-expression of I2S and FGE (i.e., SUMF1). (A) Expression units on separate vectors (for co-transfection or subsequent transfections); (B) Expression units on the same vector (one transfection): (1) Separate cistrons and (2) Transcriptionally linked cistrons.

[0028] *Figure 3* depicts exemplary levels of I2S specific activity observed as correlated to percent formylglycine conversion.

[0029] *Figure 4* depicts an exemplary glycan profile generated for recombinant I2S enzyme produced using the I2S-AF 2D and 4D cell lines grown under serum-free cell culture conditions as compared to a reference recombinant I2S enzyme.

DEFINITIONS

[0030] In order for the present invention to be more readily understood, certain terms are first defined. Additional definitions for the following terms and other terms are set forth throughout the specification.

[0031] *Amino acid*: As used herein, term "amino acid," in its broadest sense, refers to any compound and/or substance that can be incorporated into a polypeptide chain. In some embodiments, an amino acid has the general structure H_2N–C(H)(R)–COOH. In some embodiments, an amino acid is a naturally occurring amino acid. In some embodiments, an amino acid is a synthetic amino acid; in some embodiments, an amino acid is a D-amino acid; in some embodiments, an amino acid is an L-amino acid. "Standard amino acid" refers to any of the twenty standard L-amino acids commonly found in naturally occurring peptides. "Nonstandard amino acid" refers to any amino acid, other than the standard amino acids,

regardless of whether it is prepared synthetically or obtained from a natural source. As used herein, "synthetic amino acid" encompasses chemically modified amino acids, including but not limited to salts, amino acid derivatives (such as amides), and/or substitutions. Amino acids, including carboxy- and/or amino-terminal amino acids in peptides, can be modified by methylation, amidation, acetylation, protecting groups, and/or substitution with other chemical groups that can change the peptide's circulating half-life without adversely affecting their activity. Amino acids may participate in a disulfide bond. Amino acids may comprise one or posttranslational modifications, such as association with one or more chemical entities (*e.g.*, methyl groups, acetate groups, acetyl groups, phosphate groups, formyl moieties, isoprenoid groups, sulfate groups, polyethylene glycol moieties, lipid moieties, carbohydrate moieties, biotin moieties, *etc.* In some embodiments, amino acids of the present invention may be provided in or used to supplement medium for cell cultures. In some embodiments, amino acids provided in or used to supplement cell culture medium may be provided as salts or in hydrate form.

[0032] *Approximately:* As used herein, the term "approximately" or "about," as applied to one or more values of interest, refers to a value that is similar to a stated reference value. In certain embodiments, the term "approximately" or "about" refers to a range of values that fall within 25%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, or less in either direction (greater than or less than) of the stated reference value unless otherwise stated or otherwise evident from the context (except where such number would exceed 100% of a possible value).

[0033] *Batch culture*: The term "batch culture" as used herein refers to a method of culturing cells in which all the components that will ultimately be used in culturing the cells, including the medium (see definition of "medium" below) as well as the cells themselves, are provided at the beginning of the culturing process. A batch culture is typically stopped at some point and the cells and/or components in the medium are harvested and optionally purified.

[0034] *Bioavailability:* As used herein, the term "bioavailability" generally refers to the percentage of the administered dose that reaches the blood stream of a subject.

[0035] *Biologically active*: As used herein, the phrase "biologically active" refers to a characteristic of any substance that has activity in a biological system (*e.g.*, cell culture,

organism, *etc.*). For instance, a substance that, when administered to an organism, has a biological effect on that organism, is considered to be biologically active. Biological activity can also be determined by *in vitro* assays (for example, in vitro enzymatic assays such as sulfate release assays). In particular embodiments, where a protein or polypeptide is biologically active, a portion of that protein or polypeptide that shares at least one biological activity of the protein or polypeptide is typically referred to as a "biologically active" portion. In some embodiments, a protein is produced and/or purified from a cell culture system, which displays biologically activity when administered to a subject. In some embodiments, a protein requires further processing in order to become biologically active. In some embodiments, a protein requires posttranslational modification such as, but is not limited to, glycosylation (e.g., sialyation), farnysylation, cleavage, folding, formylglycine conversion and combinations thereof, in order to become biologically active. In some embodiments, a protein produced as a proform (i.e. immature form), may require additional modification to become biologically active.

[0036] *Bioreactor*: The term "bioreactor" as used herein refers to a vessel used for the growth of a host cell culture. A bioreactor can be of any size so long as it is useful for the culturing of mammalian cells. Typically, a bioreactor will be at least 1 liter and may be 10, 100, 250, 500, 1000, 2500, 5000, 8000, 10,000, 12,0000 liters or more, or any volume in between. Internal conditions of a bioreactor, including, but not limited to pH, osmolarity, CO_2 saturation, O_2 saturation, temperature and combinations thereof, are typically controlled during the culturing period. A bioreactor can be composed of any material that suitable for holding cells in media under the culture conditions of the present invention, including glass, plastic or metal. In some embodiments, a bioreactor may be used for performing animal cell culture. In some embodiments, a bioreactor may be used for performing mammalian cell culture. In some embodiments, a bioreactor may used with cells and/or cell lines derived from such organisms as, but not limited to, mammalian cell, insect cells, bacterial cells, yeast cells and human cells. In some embodiments, a bioreactor is used for large-scale cell culture production and is typically at least 100 liters and may be 200, 500, 1000, 2500, 5000, 8000, 10,000, 12,0000 liters or more, or any volume in between. One of ordinary skill in the art will be aware of and will be able to choose suitable bioreactors for use in practicing the present invention.

[0037] *Cell culture*: These terms as used herein refer to a cell population that is gown in a medium under conditions suitable to survival and/or growth of the cell population. As will be clear to those of ordinary skill in the art, these terms as used herein may refer to the combination comprising the cell population and the medium in which the population is grown.

[0038] *Cultivation*: As used herein, the term "cultivation" or grammatical equvilents refers to a process of maintaining cells under conditions favoring growth or survival. The terms "cultivation" and "cell culture" or any synonyms are used inter-changeably in this application.

[0039] *Culture vessel*: As used herein, the term "culture vessel" refers to any container that can provide an aseptic environment for culturing cells. Exemplary culture vessels include, but are not limited to, glass, plastic, or metal containers.

[0040] *Enzyme replacement therapy (ERT)*: As used herein, the term "enzyme replacement therapy (ERT)" refers to any therapeutic strategy that corrects an enzyme deficiency by providing the missing enzyme. In some embodiments, the missing enzyme is provided by intrathecal administration. In some embodiments, the missing enzyme is provided by infusing into bloodstream. Once administered, enzyme is taken up by cells and transported to the lysosome, where the enzyme acts to eliminate material that has accumulated in the lysosomes due to the enzyme deficiency. Typically, for lysosomal enzyme replacement therapy to be effective, the therapeutic enzyme is delivered to lysosomes in the appropriate cells in target tissues where the storage defect is manifest.

[0041] *Expression*: As used herein, "expression" of a nucleic acid sequence refers to one or more of the following events: (1) production of an RNA template from a DNA sequence (*e.g.*, by transcription); (2) processing of an RNA transcript (*e.g.*, by splicing, editing, 5' cap formation, and/or 3' end formation); (3) translation of an RNA into a polypeptide or protein; and/or (4) post-translational modification of a polypeptide or protein.

[0042] *Fed-batch culture*: The term "fed-batch culture" as used herein refers to a method of culturing cells in which additional components are provided to the culture at some time subsequent to the beginning of the culture process. The provided components typically comprise nutritional supplements for the cells which have been depleted during the culturing

process. A fed-batch culture is typically stopped at some point and the cells and/or components in the medium are harvested and optionally purified.

[0043] *Fragment*: The term "fragment" as used herein refers to polypeptides and is defined as any discrete portion of a given polypeptide that is unique to or characteristic of that polypeptide. The term as used herein also refers to any discrete portion of a given polypeptide that retains at least a fraction of the activity of the full-length polypeptide. Preferably the fraction of activity retained is at least 10% of the activity of the full-length polypeptide. More preferably the fraction of activity retained is at least 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90% of the activity of the full-length polypeptide. More preferably still the fraction of activity retained is at least 95%, 96%, 97%, 98% or 99% of the activity of the full-length polypeptide. Most preferably, the fraction of activity retained is 100% of the activity of the full-length polypeptide. The term as used herein also refers to any portion of a given polypeptide that includes at least an established sequence element found in the full-length polypeptide. Preferably, the sequence element spans at least 4-5, more preferably at least about 10, 15, 20, 25, 30, 35, 40, 45, 50 or more amino acids of the full-length polypeptide.

[0044] *Gene*: The term "gene" as used herein refers to any nucleotide sequence, DNA or RNA, at least some portion of which encodes a discrete final product, typically, but not limited to, a polypeptide, which functions in some aspect of a cellular process. The term is not meant to refer only to the coding sequence that encodes the polypeptide or other discrete final product, but may also encompass regions preceding and following the coding sequence that modulate the basal level of expression, as well as intervening sequences ("introns") between individual coding segments ("exons"). In some embodiments, a gene may include regulatory sequences (*e.g.*, promoters, enhancers, poly adenylation sequences, termination sequences, kozac sequences, tata box, *etc.*) and/or modification sequences. In some embodiments, a gene may include references to nucleic acids that do not encode proteins but rather encode functional RNA molecules such as tRNAs, RNAi-inducing agents, *etc.*

[0045] *Gene product* or *expression product*: As used herein, the term "gene product" or "expression product" generally refers to an RNA transcribed from the gene (pre-and/or post-processing) or a polypeptide (pre- and/or post-modification) encoded by an RNA transcribed from the gene.

[0046] *Genetic control element*: The term "genetic control element" as used herein refers to any sequence element that modulates the expression of a gene to which it is operably linked. Genetic control elements may function by either increasing or decreasing the expression levels and may be located before, within or after the coding sequence. Genetic control elements may act at any stage of gene expression by regulating, for example, initiation, elongation or termination of transcription, mRNA splicing, mRNA editing, mRNA stability, mRNA localization within the cell, initiation, elongation or termination of translation, or any other stage of gene expression. Genetic control elements may function individually or in combination with one another.

[0047] *Homology*: As used herein, the term "homology" refers to the overall relatedness between polymeric molecules, *e.g.*, between nucleic acid molecules (*e.g.*, DNA molecules and/or RNA molecules) and/or between polypeptide molecules. In some embodiments, polymeric molecules are considered to be "homologous" to one another if their sequences are at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% identical. In some embodiments, polymeric molecules are considered to be "homologous" to one another if their sequences are at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% similar.

[0048] *Identity*: As used herein, the term "identity" refers to the overall relatedness between polymeric molecules, *e.g.*, between nucleic acid molecules (*e.g.*, DNA molecules and/or RNA molecules) and/or between polypeptide molecules. Calculation of the percent identity of two nucleic acid sequences, for example, can be performed by aligning the two sequences for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second nucleic acid sequences for optimal alignment and non-identical sequences can be disregarded for comparison purposes). In certain embodiments, the length of a sequence aligned for comparison purposes is at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or substantially 100% of the length of the reference sequence. The nucleotides at corresponding nucleotide positions are then compared. When a position in the first sequence is occupied by the same nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which needs to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. For example, the percent identity between two nucleotide sequences can be determined using the algorithm of Meyers and Miller (CABIOS, 1989, 4: 11-17), which has been incorporated into the ALIGN program (version 2.0) using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. The percent identity between two nucleotide sequences can, alternatively, be determined using the GAP program in the GCG software package using an NWSgapdna.CMP matrix. Various other sequence alignment programs are available and can be used to determine sequence identity such as, for example, Clustal.

[0049] *Improve, increase, or reduce*: As used herein, the terms "improve," "increase" or "reduce," or grammatical equivalents, indicate values that are relative to a baseline measurement, such as a measurement in the same individual prior to initiation of the treatment described herein, or a measurement in a control individual (or multiple control individuals) in the absence of the treatment described herein. A "control individual" is an individual afflicted with the same form of lysosomal storage disease as the individual being treated, who is about the same age as the individual being treated (to ensure that the stages of the disease in the treated individual and the control individual(s) are comparable).

[0050] *Intrathecal administration*: As used herein, the term "intrathecal administration" or "intrathecal injection" refers to an injection into the spinal canal (intrathecal space surrounding the spinal cord). Various techniques may be used including, without limitation, lateral cerebroventricular injection through a burrhole or cisternal or lumbar puncture or the like. In some embodiments, "intrathecal administration" or "intrathecal delivery" according to the present invention refers to IT administration or delivery via the lumbar area or region, i.e., lumbar IT administration or delivery. As used herein, the term "lumbar region" or "lumbar area" refers to the area between the third and fourth lumbar (lower back) vertebrae and, more inclusively, the L2-S1 region of the spine.

[0051] *Isolated*: As used herein, the term "isolated" refers to a substance and/or entity that has been (1) separated from at least some of the components with which it was associated when initially produced (whether in nature and/or in an experimental setting), and/or (2) produced, prepared, and/or manufactured by the hand of man. Isolated substances and/or entities may be separated from about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 91%, about 92%, about 93%,

about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more than about 99% of the other components with which they were initially associated. In some embodiments, isolated agents are about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more than about 99% pure. As used herein, a substance is "pure" if it is substantially free of other components. As used herein, calculation of percent purity of isolated substances and/or entities should not include excipients (*e.g.*, buffer, solvent, water, *etc.*)

[0052] *Medium*: The terms as used herein refer to a solution containing nutrients which nourish growing cells. Typically, these solutions provide essential and non-essential amino acids, vitamins, energy sources, lipids, and trace elements required by the cell for minimal growth and/or survival. The solution may also contain components that enhance growth and/or survival above the minimal rate, including hormones and growth factors. In some embodiments, medium is formulated to a pH and salt concentration optimal for cell survival and proliferation. In some embodiments, medium may be a "chemically defined medium" – a serum-free media that contains no proteins, hydrolysates or components of unknown composition. In some embodiment, chemically defined medium is free of animal-derived components and all components within the medium have a known chemical structure. In some embodiments, medium may be a "serum based medium" – a medium that has been supplemented animal derived components such as, but not limited to, fetal calf serum, horse serum, goat serum, donkey serum and/or combinations thereof.

[0053] *Nucleic acid*: As used herein, the term "nucleic acid," in its broadest sense, refers to a compound and/or substance that is or can be incorporated into an oligonucleotide chain. In some embodiments, a nucleic acid is a compound and/or substance that is or can be incorporated into an oligonucleotide chain via a phosphodiester linkage. In some embodiments, "nucleic acid" refers to individual nucleic acid residues (*e.g.*, nucleotides and/or nucleosides). In some embodiments, "nucleic acid" refers to an oligonucleotide chain comprising individual nucleic acid residues. As used herein, the terms "oligonucleotide" and "polynucleotide" can be used interchangeably. In some embodiments, "nucleic acid" encompasses RNA as well as single and/or double-stranded DNA and/or cDNA. Furthermore, the terms "nucleic acid," "DNA," "RNA," and/or similar terms include nucleic acid analogs, *i.e.*, analogs having other than a phosphodiester backbone. For example, the so-called "peptide nucleic acids," which are known in the art and have peptide bonds instead of

phosphodiester bonds in the backbone, are considered within the scope of the present invention. The term "nucleotide sequence encoding an amino acid sequence" includes all nucleotide sequences that are degenerate versions of each other and/or encode the same amino acid sequence. Nucleotide sequences that encode proteins and/or RNA may include introns. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, *etc.* Where appropriate, *e.g.*, in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, backbone modifications, *etc.* A nucleic acid sequence is presented in the 5' to 3' direction unless otherwise indicated. The term "nucleic acid segment" is used herein to refer to a nucleic acid sequence that is a portion of a longer nucleic acid sequence. In many embodiments, a nucleic acid segment comprises at least 3, 4, 5, 6, 7, 8, 9, 10, or more residues. In some embodiments, a nucleic acid is or comprises natural nucleosides (*e.g.*, adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (*e.g.*, 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, C-5 propynyl-cytidine, C-5 propynyl-uridine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (*e.g.*, methylated bases); intercalated bases; modified sugars (*e.g.*, 2'-fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (*e.g.*, phosphorothioates and 5'-*N*-phosphoramidite linkages). In some embodiments, the present invention is specifically directed to "unmodified nucleic acids," meaning nucleic acids (*e.g.*, polynucleotides and residues, including nucleotides and/or nucleosides) that have not been chemically modified in order to facilitate or achieve delivery.

[0054] *Perfusion process*: The term "perfusion process" as used herein refers to a method of culturing cells in which additional components are provided continuously or semi-continuously to the culture subsequent to the beginning of the culture process. The provided components typically comprise nutritional supplements for the cells which have been depleted during the culturing process. A portion of the cells and/or components in the medium are typically harvested on a continuous or semi-continuous basis and are optionally purified. Typically, a cell culture process involving a perfusion process is referred to as

"perfusion culture." Typically, nutritional supplements are provided in a fresh medium during a perfusion process. In some embodiments, a fresh medium may be identical or similar to the base medium used in the cell culture process. In some embodiments, a fresh medium may be different than the base medium but containing desired nutritional supplements. In some embodiments, a fresh medium is a chemically-defined medium.

[0055] *Protein*: As used herein, the term "protein" refers to a polypeptide (*i.e.*, a string of at least two amino acids linked to one another by peptide bonds). Proteins may include moieties other than amino acids (*e.g.*, may be glycoproteins, proteoglycans, *etc.*) and/or may be otherwise processed or modified. Those of ordinary skill in the art will appreciate that a "protein" can be a complete polypeptide chain as produced by a cell (with or without a signal sequence), or can be a characteristic portion thereof. In some embodiments, a protein can sometimes include more than one polypeptide chain, for example linked by one or more disulfide bonds or associated by other means. In some embodiments, polypeptides may contain L-amino acids, D-amino acids, or both and may contain any of a variety of amino acid modifications or analogs known in the art. Useful modifications include, *e.g.*, terminal acetylation, amidation, methylation, *etc.* In some embodiments, proteins may comprise natural amino acids, non-natural amino acids, synthetic amino acids, and combinations thereof. The term "peptide" is generally used to refer to a polypeptide having a length of less than about 100 amino acids, less than about 50 amino acids, less than 20 amino acids, or less than 10 amino acids. In some embodiments, proteins are antibodies, antibody fragments, biologically active portions thereof, and/or characteristic portions thereof.

[0056] *Recombinant protein* and *Recombinant polypeptide*: These terms as used herein refer to a polypeptide expressed from a host cell, that has been genetically engineered to express that polypeptide. In some embodiments, a recombinant protein may be expressed in a host cell derived from an animal. In some embodiments, a recombinant protein may be expressed in a host cell derived from an insect. In some embodiments, a recombinant protein may be expressed in a host cell derived from a yeast. In some embodiments, a recombinant protein may be expressed in a host cell derived from a prokaryote. In some embodiments, a recombinant protein may be expressed in a host cell derived from an mammal. In some embodiments, a recombinant protein may be expressed in a host cell derived from a human. In some embodiments, the recombinantly expressed polypeptide may be identical or similar to a polypeptide that is normally expressed in the host cell. In some embodiments, the

recombinantly expressed polypeptide may be foreign to the host cell, i.e. heterologous to peptides normally expressed in the host cell. Alternatively, in some embodiments the recombinantly expressed polypeptide can be a chimeric, in that portions of the polypeptide contain amino acid sequences that are identical or similar to polypeptides normally expressed in the host cell, while other portions are foreign to the host cell.

[0057] *Replacement enzyme*: As used herein, the term "replacement enzyme" refers to any enzyme that can act to replace at least in part the deficient or missing enzyme in a disease to be treated. In some embodiments, the term "replacement enzyme" refers to any enzyme that can act to replace at least in part the deficient or missing lysosomal enzyme in a lysosomal storage disease to be treated. In some embodiments, a replacement enzyme is capable of reducing accumulated materials in mammalian lysosomes or that can rescue or ameliorate one or more lysosomal storage disease symptoms. Replacement enzymes suitable for the invention include both wild-type or modified lysosomal enzymes and can be produced using recombinant and synthetic methods or purified from nature sources. A replacement enzyme can be a recombinant, synthetic, gene-activated or natural enzyme.

[0058] *Vector*: As used herein, "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it is associated. In some embodiment, vectors are capable of extra-chromosomal replication and/or expression of nucleic acids to which they are linked in a host cell such as a eukaryotic and/or prokaryotic cell. Vectors capable of directing the expression of operatively linked genes are referred to herein as "expression vectors."

DETAILED DESCRIPTION OF THE INVENTION

[0059] The present invention provides, among other things, methods and compositions for production of recombinant I2S protein with improved potency and activity using cells co-expressing I2S and FGE protein. In some embodiments, cells according to the present invention are engineered to simultaneously over-express recombinant I2S and FGE proteins. Cells according to the invention are adaptable to various cell culture conditions. In some embodiments, cells of the present invention are adaptable to a large-scale suspension serum-free culture.

[0060] Various aspects of the invention are described in further detail in the following subsections. The use of subsections is not meant to limit the invention. Each subsection may apply to any aspect of the invention. In this application, the use of "or" means "and/or" unless stated otherwise.

Iduronate-2-sulfatase (I2S)

[0061] As used herein, an I2S protein is any protein or a portion of a protein that can substitute for at least partial activity of naturally-occurring Iduronate-2-sulfatase (I2S) protein or rescue one or more phenotypes or symptoms associated with I2S-deficiency. As used herein, the terms "an I2S enzyme" and "an I2S protein", and grammatical equivalents, are used inter-changeably.

[0062] Typically, the human I2S protein is produced as a precursor form. The precursor form of human I2S contains a signal peptide (amino acid residues 1-25 of the full length precursor), a pro-peptide (amino acid residues 26-33 of the full length precursor), and a chain (residues 34-550 of the full length precursor) that may be further processed into the 42 kDa chain (residues 34-455 of the full length precursor) and the 14 kDa chain (residues 446-550 of the full length precursor). Typically, the precursor form is also referred to as full-length precursor or full-length I2S protein, which contains 550 amino acids. The amino acid sequences of the mature form (SEQ ID NO:1) having the signal peptide removed and full-length precursor (SEQ ID NO:2) of a typical wild-type or naturally-occurring human I2S protein are shown in Table 1. The signal peptide is underlined. In addition, the amino acid sequences of human I2S protein isoform a and b precursor are also provided in Table 1, SEQ ID NO:3 and 4, respectively.

Table 1. Human Iduronate-2-sulfatase

Mature Form	SETQANSTTDALNVLLIIVDDLRPSLGCYGDKLVRSPNIDQLASHSLLFQNAFA QQAVCAPSRVSFLTGRRPDTTRLYDFNSYWRVHAGNFSTIPQYFKENGYVTMSV GKVFHPGISSNHTDDSPYSWSFPPYHPSSEKYENTKTCRGPDGELHANLLCPVD VLDVPEGTLPDKQSTEQAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKL YPLENITLAPDPEVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVDFQRK IRQSYFASVSYLDTQVGRLLSALDDLQLANSTIIAFTSDHGWALGEHGEWAKYS NFDVATHVPLIFYVPGRTASLPEAGEKLFPYLDPFDSASQLMEPGRQSMDLVEL VSLFPTLAGLAGLQVPPRCPVPSFHVELCREGKNLLKHFRFRDLEEDPYLPGNP RELIAYSQYPRPSDIPQWNSDKPSLKDIKIMGYSIRTIDYRYTVWVGFNPDEFL ANFSDIHAGELYFVDSDPLQDHNMYNDSQGGDLFQLLMP(SEQ ID NO:1)
Full-Length	MPPPRTGRGLLWLGLVLSSVCVALGSETQANSTTDALNVLLIIVDDLRPSLGCY GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYDFNSY

Precursor (Isoform a)	WRVHAGNFSTIPQYFKENGYVTMSVGKVFHPGISSNHTDDSPYSWSFPPYHPSS EKYENTKTCRGPDGELHANLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSA SPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDPEVPDGLPPVAYNPWMDI RQREDVQALNISVPYGPIPVDFQRKIRQSYFASVSYLDTQVGRLLSALDDLQLA NSTIIAFTSDHGWALGEHGEWAKYSNFDVATHVPLIFYVPGRTASLPEAGEKLF PYLDPFDSASQLMEPGRQSMDLVELVSLFPTLAGLAGLQVPPRCPVPSFHVELC REGKNLLKHFRFRDLEEDPYLPGNPRELIAYSQYPRPSDIPQWNSDKPSLKDIK IMGYSIRTIDYRYTVWVGFNPDEFLANFSDIHAGELYFVDSDPLQDHNMYNDSQ GGDLFQLLMP(SEQ ID NO:2)
Isoform b Precursor	MPPPRTGRGLLWLGLVLSSVCVALGSETQANSTTDALNVLLIIVDDLRPSLGCY GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYDFNSY WRVHAGNFSTIPQYFKENGYVTMSVGKVFHPGISSNHTDDSPYSWSFPPYHPSS EKYENTKTCRGPDGELHANLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSA SPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDPEVPDGLPPVAYNPWMDI RQREDVQALNISVPYGPIPVDFQEDQSSTGFRLKTSSTRKYK (SEQ ID NO:3)
Isoform c Precursor	MPPPRTGRGLLWLGLVLSSVCVALGSETQANSTTDALNVLLIIVDDLRPSLGCY GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYDFNSY WRVHAGNFSTIPQYFKENGYVTMSVGKVFHPGISSNHTDDSPYSWSFPPYHPSS EKYENTKTCRGPDGELHANLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSA SPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDPEVPDGLPPVAYNPWMDI RQREDVQALNISVPYGPIPVDFQRKIRQSYFASVSYLDTQVGRLLSALDDLQLA NSTIIAFTSDHGFLMRTNT(SEQ ID No:4)

[0063] Thus, in some embodiments, an I2S enzyme is mature human I2S protein (SEQ ID NO:1). As disclosed herein, SEQ ID NO:1 represents the canonical amino acid sequence for the human I2S protein. In some embodiments, the I2S protein may be a splice isoform and/or variant of SEQ ID NO:1, resulting from transcription at an alternative start site within the 5' UTR of the I2S gene. In some embodiments, a suitable replacement enzyme may be a homologue or an analogue of mature human I2S protein. For example, a homologue or an analogue of mature human I2S protein may be a modified mature human I2S protein containing one or more amino acid substitutions, deletions, and/or insertions as compared to a wild-type or naturally-occurring I2S protein (e.g., SEQ ID NO:1), while retaining substantial I2S protein activity. Thus, in some embodiments, a replacement enzyme suitable for the present invention is substantially homologous to mature human I2S protein (SEQ ID NO:1). In some embodiments, a replacement enzyme suitable for the present invention has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homologous to SEQ ID NO:1. In some embodiments, a replacement enzyme suitable for the present invention is substantially identical to mature human I2S protein (SEQ ID NO:1). In some embodiments, a replacement enzyme suitable for the present invention has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%,

98%, 99% or more identical to SEQ ID NO:1. In some embodiments, a replacement enzyme suitable for the present invention contains a fragment or a portion of mature human I2S protein.

[0064] Alternatively, an I2S enzyme is full-length I2S protein. In some embodiments, an I2S enzyme may be a homologue or an analogue of full-length human I2S protein. For example, a homologue or an analogue of full-length human I2S protein may be a modified full-length human I2S protein containing one or more amino acid substitutions, deletions, and/or insertions as compared to a wild-type or naturally-occurring full-length I2S protein (e.g., SEQ ID NO:2), while retaining substantial I2S protein activity. Thus, In some embodiments, an I2S enzyme is substantially homologous to full-length human I2S protein (SEQ ID NO:2). In some embodiments, an I2S enzyme suitable for the present invention has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homologous to SEQ ID NO:2. In some embodiments, an I2S enzyme suitable for the present invention is substantially identical to SEQ ID NO:2. In some embodiments, an I2S enzyme suitable for the present invention has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:2. In some embodiments, an I2S enzyme suitable for the present invention contains a fragment or a portion of full-length human I2S protein. As used herein, a full-length I2S protein typically contains signal peptide sequence.

[0065] In some embodiments, an I2S enzyme suitable for the present invention is human I2S isoform a protein. In some embodiments, a suitable I2S enzyme may be a homologue or an analogue of human I2S isoform a protein. For example, a homologue or an analogue of human I2S isoform a protein may be a modified human I2S isoform a protein containing one or more amino acid substitutions, deletions, and/or insertions as compared to a wild-type or naturally-occurring human I2S isoform a protein (e.g., SEQ ID NO:3), while retaining substantial I2S protein activity. Thus, in some embodiments, an I2S enzyme is substantially homologous to human I2S isoform a protein (SEQ ID NO:3). In some embodiments, an I2S enzyme has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homologous to SEQ ID NO:3. In some embodiments, an I2S enzyme is substantially identical to SEQ ID NO:3. In some embodiments, an I2S enzyme suitable for the present

invention has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:3. In some embodiments, an I2S enzyme suitable for the present invention contains a fragment or a portion of human I2S isoform a protein. As used herein, a human I2S isoform a protein typically contains a signal peptide sequence.

[0066] In some embodiments, an I2S enzyme is human I2S isoform b protein. In some embodiments, an I2S enzyme may be a homologue or an analogue of human I2S isoform b protein. For example, a homologue or an analogue of human I2S isoform b protein may be a modified human I2S isoform b protein containing one or more amino acid substitutions, deletions, and/or insertions as compared to a wild-type or naturally-occurring human I2S isoform b protein (e.g., SEQ ID NO:4), while retaining substantial I2S protein activity. Thus, In some embodiments, an I2S enzyme is substantially homologous to human I2S isoform b protein (SEQ ID NO:4). In some embodiments, an I2S enzyme has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homologous to SEQ ID NO:4. In some embodiments, an I2S enzyme is substantially identical to SEQ ID NO:4. In some embodiments, an I2S enzyme has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:4. In some embodiments, an I2S enzyme suitable for the present invention contains a fragment or a portion of human I2S isoform b protein. As used herein, a human I2S isoform b protein typically contains a signal peptide sequence.

[0067] Homologues or analogues of human I2S proteins can be prepared according to methods for altering polypeptide sequence known to one of ordinary skill in the art such as are found in references that compile such methods. In some embodiments, conservative substitutions of amino acids include substitutions made among amino acids within the following groups: (a) M, I, L, V; (b) F, Y, W; (c) K, R, H; (d) A, G; (e) S, T; (f) Q, N; and (g) E, D. In some embodiments, a "conservative amino acid substitution" refers to an amino acid substitution that does not alter the relative charge or size characteristics of the protein in which the amino acid substitution is made.

[0068] In some embodiments, I2S enzymes contain a moiety that binds to a receptor on the surface of cells to facilitate cellular uptake and/or lysosomal targeting. For example, such a receptor may be the cation-independent mannose-6-phosphate receptor (CI-MPR)

which binds the mannose-6-phosphate (M6P) residues. In addition, the CI-MPR also binds other proteins including IGF-II. A suitable lysosomal targeting moiety can be IGF-I, IGF-II, RAP, p97, and variants, homologues or fragments thereof (e.g., including those peptide having a sequence at least 70%, 75%, 80%, 85%, 90%, or 95% identical to a wild-type mature human IGF-I, IGF-II, RAP, p97 peptide sequence). In some embodiments, a suitable receptor that the M6P residues bind may be cation-dependent.

Formylglycine Generating Enzyme (FGE)

[0069] Typically, the enzyme activity of I2S is influenced by a post-translational modification of a conserved cysteine (e.g., corresponding to amino acid 59 of the mature human I2S (SEQ ID NO:1)) to formylglycine, which is also referred to as 2-amino-3-oxopropionic acid, or oxo-alanine. This post-translational modification generally occurs in the endoplasmic reticulum during protein synthesis and is catalyzed by Formylglycine Generating Enzyme (FGE). The specific enzyme activity of I2S is typically positively correlated with the extent to which the I2S has the formylglycine modification. For example, an I2S protein preparation that has a relatively high amount of formylglycine modification typically has a relatively high specific enzyme activity; whereas an I2S protein preparation that has a relatively low amount of formylglycine modification typically has a relatively low specific enzyme activity.

[0070] Thus, cells suitable for producing recombinant I2S protein according to the present invention typically also express FGE protein. In some embodiments, suitable cells express an endogenous FGE protein. In some embodiments, suitable cells are engineered to express an exogenous (also referred to as recombinant) Formylglycine Generating Enzyme (FGE) in combination with recombinant I2S. In some embodiments, suitable cells are engineered to activate an endogenous FGE gene such that the expression level or activity of the FGE protein is increased.

[0071] Typically, the human FGE protein is produced as a precursor form. The precursor form of human FGE contains a signal peptide (amino acid residues 1-33 of the full length precursor) and a chain (residues 34-374 of the full length precursor). Typically, the precursor form is also referred to as full-length precursor or full-length FGE protein, which contains 374 amino acids. The amino acid sequences of the mature form (SEQ ID NO:5)

having the signal peptide removed and full-length precursor (SEQ ID NO:6) of a typical wild-type or naturally-occurring human FGE protein are shown in Table 2.

Table 2. Human Formylglycine Generating Enzyme (FGE)

Mature Form	SQEAGTGAGAGSLAGSCGCGTPQRPGAHGSSAAAHRYSREANAPGPVPGERQLA HSKMVPIPAGVFTMGTDDPQIKQDGEAPARRVTIDAFYMDAYEVSNTEFEKFVN STGYLTEAEKFGDSFVFEGMLSEQVKTNIQQAVAAAPWWLPVKGANWRHPEGPD STILHRPDHPVLHVSWNDAVAYCTWAGKRLPTEAEWEYSCRGGLHNRLFPWGNK LQPKGQHYANIWQGEFPVTNTGEDGFQGTAPVDAFPPNGYGLYNIVGNAWEWTS DWWTVHHSVEETLNPKGPPSGKDRVKKGGSYMCHRSYCYRYRCAARSQNTPDSS ASNLGFRCAADRLPTMD (SEQ ID NO:5)
Full-Length Precursor	MAAPALGLVCGRCPELGLVLLLLLLSLLCGAAGSQEAGTGAGAGSLAGSCGCGT PQRPGAHGSSAAAHRYSREANAPGPVPGERQLAHSKMVPIPAGVFTMGTDDPQI KQDGEAPARRVTIDAFYMDAYEVSNTEFEKFVNSTGYLTEAEKFGDSFVFEGML SEQVKTNIQQAVAAAPWWLPVKGANWRHPEGPDSTILHRPDHPVLHVSWNDAVA YCTWAGKRLPTEAEWEYSCRGGLHNRLFPWGNKLQPKGQHYANIWQGEFPVTNT GEDGFQGTAPVDAFPPNGYGLYNIVGNAWEWTSDWWTVHHSVEETLNPKGPPSG KDRVKKGGSYMCHRSYCYRYRCAARSQNTPDSSASNLGFRCAADRLPTMD (SEQ ID NO:6)

[0072] Thus, in some embodiments, an FGE enzyme suitable for the present invention is mature human FGE protein (SEQ ID NO:5). In some embodiments, a suitable FGE enzyme may be a homologue or an analogue of mature human FGE protein. For example, a homologue or an analogue of mature human FGE protein may be a modified mature human FGE protein containing one or more amino acid substitutions, deletions, and/or insertions as compared to a wild-type or naturally-occurring FGE protein (e.g., SEQ ID NO:5), while retaining substantial FGE protein activity. Thus, in some embodiments, an FGE enzyme suitable for the present invention is substantially homologous to mature human FGE protein (SEQ ID NO:5). In some embodiments, an FGE enzyme suitable for the present invention has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homologous to SEQ ID NO:5. In some embodiments, an FGE enzyme suitable for the present invention is substantially identical to mature human FGE protein (SEQ ID NO:5). In some embodiments, an FGE enzyme suitable for the present invention has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:5. In some embodiments, an FGE enzyme suitable for the present invention contains a fragment or a portion of mature human FGE protein.

[0073] Alternatively, an FGE enzyme suitable for the present invention is full-length FGE protein. In some embodiments, an FGE enzyme may be a homologue or an analogue of

full-length human FGE protein. For example, a homologue or an analogue of full-length human FGE protein may be a modified full-length human FGE protein containing one or more amino acid substitutions, deletions, and/or insertions as compared to a wild-type or naturally-occurring full-length FGE protein (e.g., SEQ ID NO:6), while retaining substantial FGE protein activity. Thus, in some embodiments, an FGE enzyme suitable for the present invention is substantially homologous to full-length human FGE protein (SEQ ID NO:6). In some embodiments, an FGE enzyme suitable for the present invention has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homologous to SEQ ID NO:4. In some embodiments, an FGE enzyme suitable for the present invention is substantially identical to SEQ ID NO:6. In some embodiments, an FGE enzyme suitable for the present invention has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:6. In some embodiments, an FGE enzyme suitable for the present invention contains a fragment or a portion of full-length human FGE protein. As used herein, a full-length FGE protein typically contains signal peptide sequence.

[0074] Exemplary nucleic acid sequences and amino acid sequences encoding exemplary FGE proteins are disclosed US Publication No. 20040229250, the entire contents of which is incorporated herein by reference.

Cells Co-Expressing I2S and FGE

[0075] The present invention recognizes the need for the high-level, commercial production of biologically active I2S using a cell culture system. Because a large number of production factors can influence the selection of a specific host cell, nucleic acid molecules disclosed in the present specification are directed toward a wide range of prokaryotic and eukaryotic cells and/or cell lines including, without limitation, cell lines derived from bacteria strains, yeast strains, insect cells, animal cells, mammalian cells and human cells. Aspects of the present invention also provide for expression constructs and the generation of recombinant stable cell lines useful for expressing naturally occurring, as well as, modified I2S and/or FGE proteins which are disclosed in the present specification. In addition, aspects of the present invention also provide methods for producing cell lines that express I2S and FGE using the disclosed nucleic acid sequences of the present specification.

Nucleic Acids Encoding I2S and/or FGE Proteins

[0076] In some embodiments, nucleic acid molecules are provided comprising nucleic acid sequences encoding for a recombinant gene of interest (herein referred to as a transgene) such as an I2S and/or FGE protein described in various embodiments herein. In some embodiments, the nucleic acid encoding a transgene may be modified to provide increased expression of the encoded I2S and/or FGE protein, which is also referred to as codon optimization. For example, the nucleic acid encoding a transgene can be modified by altering the open reading frame for the coding sequence. As used herein, the term "open reading frame" is synonymous with "ORF" and means any nucleotide sequence that is potentially able to encode a protein, or a portion of a protein. An open reading frame usually begins with a start codon (represented as, e.g. AUG for an RNA molecule and ATG in a DNA molecule in the standard code) and is read in codon-triplets until the frame ends with a STOP codon (represented as, e.g. UAA, UGA or UAG for an RNA molecule and TAA, TGA or TAG in a DNA molecule in the standard code). As used herein, the term "codon" means a sequence of three nucleotides in a nucleic acid molecule that specifies a particular amino acid during protein synthesis; also called a triplet or codon-triplet. For example, of the 64 possible codons in the standard genetic code, two codons, GAA and GAG encode the amino acid Glutamine whereas the codons AAA and AAG specify the amino acid Lysine. In the standard genetic code three codons are stop codons, which do not specify an amino acid. As used herein, the term "synonymous codon" means any and all of the codons that code for a single amino acid. Except for Methionine and Tryptophan, amino acids are coded by two to six synonymous codons. For example, in the standard genetic code the four synonymous codons that code for the amino acid Alanine are GCA, GCC, GCG and GCU, the two synonymous codons that specify Glutamine are GAA and GAG and the two synonymous codons that encode Lysine are AAA and AAG.

[0077] In some embodiments, a nucleic acid encoding the open reading frame of an I2S and/or FGE protein may be modified using standard codon optimization methods. Various commercial algorithms for codon optimization are available and can be used to practice the present invention. Typically, codon optimization does not alter the encoded amino acid sequences. In some embodiments, codon optimization may lead to amino acids

alteration such as substitution, deletion or insertion. Typically, such amino acid alteration does not substantially alter the protein activity.

[0078] Exemplary nucleic acid sequences encoding an I2S and FGE proteins, respectively are shown in SEQ ID NO:7 and 8 below.

SEQ ID NO:7 Exemplary nucleic acid sequence encoding iduronate 2-sulfatase (I2S)

ATGCCCCCGCCCCGCACCGGCCGCGGCCTGCTGTGGCTGGGCCTGGTGCTGAGCAGCGTGTGCGTG
GCCCTGGGCAGCGAGACCCAGGCCAACAGCACCACCGACGCCCTGAACGTGCTGCTGATCATCGT
GGACGACCTGCGCCCCAGCCTGGGCTGCTACGGCGACAAGCTGGTGCGCAGCCCCAACATCGACC
AGCTGGCCAGCCACAGCCTGCTGTTCCAGAACGCCTTCGCCCAGCAGGCCGTGTGCGCCCCCAGCC
GCGTGAGCTTCCTGACCGGCCGCCGCCCCGACACCACCCGCCTGTACGACTTCAACAGCTACTGGC
GCGTGCACGCCGGCAACTTCAGCACCATCCCCCAGTACTTCAAGGAGAACGGCTACGTGACCATG
AGCGTGGGCAAGGTGTTCCACCCCGGCATCAGCAGCAACCACACCGACGACAGCCCCTACAGCTG
GAGCTTCCCCCCCTACCACCCCAGCAGCGAGAAGTACGAGAACACCAAGACCTGCCGCGGCCCCG
ACGGCGAGCTGCACGCCAACCTGCTGTGCCCCGTGGACGTGCTGGACGTGCCCGAGGGCACCCTG
CCCGACAAGCAGAGCACCGAGCAGGCCATCCAGCTGCTGGAGAAGATGAAGACCAGCGCCAGCC
CCTTCTTCCTGGCCGTGGGCTACCACAAGCCCCACATCCCCTTCCGCTACCCCAAGGAGTTCCAGA
AGCTGTACCCCCTGGAGAACATCACCCTGGCCCCCGACCCCGAGGTGCCCGACGGCCTGCCCCCCG
TGGCCTACAACCCCTGGATGGACATCCGCCAGCGCGAGGACGTGCAGGCCCTGAACATCAGCGTG
CCCTACGGCCCCATCCCCGTGGACTTCCAGCGCAAGATCCGCCAGAGCTACTTCGCCAGCGTGAGC
TACCTGGACACCCAGGTGGGCCGCCTGCTGAGCGCCCTGGACGACCTGCAGCTGGCCAACAGCAC
CATCATCGCCTTCACCAGCGACCACGGCTGGGCCCTGGGCGAGCACGGCGAGTGGGCCAAGTACA
GCAACTTCGACGTGGCCACCCACGTGCCCCTGATCTTCTACGTGCCCGGCCGCACCGCCAGCCTGC
CCGAGGCCGGCGAGAAGCTGTTCCCCTACCTGGACCCCTTCGACAGCGCCAGCCAGCTGATGGAG
CCCGGCCGCCAGAGCATGGACCTGGTGGAGCTGGTGAGCCTGTTCCCCACCCTGGCCGGCCTGGCC
GGCCTGCAGGTGCCCCCCCGCTGCCCCGTGCCCAGCTTCCACGTGGAGCTGTGCCGCGAGGGCAA
GAACCTGCTGAAGCACTTCCGCTTCCGCGACCTGGAGGAGGACCCCTACCTGCCCGGCAACCCCCG
CGAGCTGATCGCCTACAGCCAGTACCCCCGCCCCAGCGACATCCCCCAGTGGAACAGCGACAAGC
CCAGCCTGAAGGACATCAAGATCATGGGCTACAGCATCCGCACCATCGACTACCGCTACACCGTG
TGGGTGGGCTTCAACCCCGACGAGTTCCTGGCCAACTTCAGCGACATCCACGCCGGCGAGCTGTAC
TTCGTGGACAGCGACCCCCTGCAGGACCACAACATGTACAACGACAGCCAGGGCGGCGACCTGTT
CCAGCTGCTGATGCCCTAG

SEQ ID NO:8 Exemplary nucleic acid sequence encoding full-length precursor formylglycine generating enzyme (FGE)

ATGGCTGCGCCCGCACTAGGGCTGGTGTGTGGACGTTGCCCTGAGCTGGGTCTCGTCCTCTTGCTG
CTGCTGCTCTCGCTGCTGTGTGGAGCGGCAGGGAGCCAGGAGGCCGGGACCGGTGCGGGCGCGGG
GTCCCTTGCGGGTTCTTGCGGCTGCGGCACGCCCCAGCGGCCTGGCGCCCATGGCAGTTCGGCAGC
CGCTCACCGATACTCGCGGGAGGCTAACGCTCCGGGCCCCGTACCCGGAGAGCGGCAACTCGCGC

ACTCAAAGATGGTCCCCATCCCTGCTGGAGTATTTACAATGGGCACAGATGATCCTCAGATAAAGC
AGGATGGGGAAGCACCTGCGAGGAGAGTTACTATTGATGCCTTTTACATGGATGCCTATGAAGTC
AGTAATACTGAATTTGAGAAGTTTGTGAACTCAACTGGCTATTTGACAGAGGCTGAGAAGTTTGGC
GACTCCTTTGTCTTTGAAGGCATGTTGAGTGAGCAAGTGAAGACCAATATTCAACAGGCAGTTGCA
GCTGCTCCCTGGTGGTTACCTGTGAAAGGCGCTAACTGGAGACACCCAGAAGGGCCTGACTCTACT
ATTCTGCACAGGCCGGATCATCCAGTTCTCCATGTGTCCTGGAATGATGCGGTTGCCTACTGCACTT
GGGCAGGGAAGCGGCTGCCCACGGAAGCTGAGTGGGAATACAGCTGTCGAGGAGGCCTGCATAA
TAGACTTTTCCCCTGGGGCAACAAACTGCAGCCCAAAGGCCAGCATTATGCCAACATTTGGCAGG
GCGAGTTTCCGGTGACCAACACTGGTGAGGATGGCTTCCAAGGAACTGCGCCTGTTGATGCCTTCC
CTCCCAATGGTTATGGCTTATACAACATAGTGGGGAACGCATGGGAATGGACTTCAGACTGGTGG
ACTGTTCATCATTCTGTTGAAGAAACGCTTAACCCAAAAGGTCCCCCTTCTGGGAAAGACCGAGTG
AAGAAAGGTGGATCCTACATGTGCCATAGGTCTTATTGTTACAGGTATCGCTGTGCTGCTCGGAGC
CAGAACACACCTGATAGCTCTGCTTCGAATCTGGGATTCCGCTGTGCAGCCGACCGCCTGCCCACC
ATGGACTGA

[0079] In some embodiments, a nucleotide change may alter a synonymous codon within the open reading frame in order to agree with the endogenous codon usage found in a particular heterologous cell selected to express I2S and/or FGE. Alternatively or additionally, a nucleotide change may alter the G+C content within the open reading frame to better match the average G+C content of open reading frames found in endogenous nucleic acid sequence present in the heterologous host cell. A nucleotide change may also alter a polymononucleotide region or an internal regulatory or structural site found within an I2S or FGE sequence. Thus, a variety of modified or optimized nucleotide sequences are envisioned including, without limitation, nucleic acid sequences providing increased expression of I2S and/or FGE proteins in a prokaryotic cell; yeast cell; insect cell; and in a mammalian cell.

[0080] Thus, in some embodiments, a nucleic acid encoding an I2S protein suitable for the present invention has a nucleotide sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:7. In some embodiments, a nucleic acid encoding an FGE protein suitable for the present invention has a nucleotide sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:8. Typically, a modified nucleic acid encodes an I2S and/or FGE protein with or without amino acid sequence alteration. In the event there is amino acid alteration, such alteration typically does not substantially alter the I2S or FGE protein activity.

Expression Vectors

[0081] A nucleic acid sequence encoding an I2S and/or FGE protein as described in the present application, can be molecularly cloned (inserted) into a suitable vector for

propagation or expression in a host cell. A wide variety of expression vectors can be used to practice the present invention, including, without limitation, a prokaryotic expression vector; a yeast expression vector; an insect expression vector and a mammalian expression vector. Exemplary vectors suitable for the present invention include, but are not limited to, viral based vectors (e.g., AAV based vectors, retrovirus based vectors, plasmid based vectors). In some embodiments, nucleic acid sequences encoding an I2S and FGE proteins, respectively can be inserted in separate vectors. In some embodiments, nucleic acid sequences encoding an I2S and FGE proteins, respectively can be inserted in a same vector. Typically, a nucleic acid encoding an I2S or FGE protein is operably linked to various regulatory sequences or elements.

Regulatory Sequences or Elements

[0082] Various regulatory sequences or elements may be incorporated in an expression vector suitable for the present invention. Exemplary regulatory sequences or elements include, but are not limited to, promoters, enhancers, repressors or suppressors, 5' untranslated (or non-coding) sequences, introns, 3' untranslated (or non-coding) sequences.

[0083] As used herein, a "Promoter" or "Promoter sequence" is a DNA regulatory region capable of binding an RNA polymerase in a cell (e.g., directly or through other promoter bound proteins or substances) and initiating transcription of a coding sequence. A promoter sequence is, in general, bound at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at any level. The promoter may be operably associated with or operably linked to the expression control sequences, including enhancer and repressor sequences or with a nucleic acid to be expressed. In some embodiments, the promoter may be inducible. In some embodiments, the inducible promoter may be unidirectional or bio-directional. In some embodiments, the promoter may be a constitutive promoter. In some embodiments, the promoter can be a hybrid promoter, in which the sequence containing the transcriptional regulatory region is obtained from one source and the sequence containing the transcription initiation region is obtained from a second source. Systems for linking control elements to coding sequence within a transgene are well known in the art (general molecular biological and recombinant DNA techniques are described in Sambrook, Fritsch, and Maniatis, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor

Laboratory Press, Cold Spring Harbor, NY, 1989, which is incorporated herein by reference). Commercial vectors suitable for inserting a transgene for expression in various host cells under a variety of growth and induction conditions are also well known in the art.

[0084] In some embodiments, a specific promoter may be used to control expression of the transgene in a mammalian host cell such as, but are not limited to, SRα-promoter (Takebe et al., Molec. and Cell. Bio. 8:466-472 (1988)), the human CMV immediate early promoter (Boshart et al., Cell 41:521-530 (1985); Foecking et al., Gene 45:101-105 (1986)), human CMV promoter, the human CMV5 promoter, the murine CMV immediate early promoter, the EF1-α-promoter, a hybrid CMV promoter for liver specific expression (e.g., made by conjugating CMV immediate early promoter with the transcriptional promoter elements of either human α-1-antitrypsin (HAT) or albumin (HAL) promoter), or promoters for hepatoma specific expression (e.g., wherein the transcriptional promoter elements of either human albumin (HAL; about 1000 bp) or human α-1-antitrypsin (HAT, about 2000 bp) are combined with a 145 long enhancer element of human α-1-microglobulin and bikunin precursor gene (AMBP); HAL-AMBP and HAT-AMBP); the SV40 early promoter region (Benoist at al., Nature 290:304-310 (1981)), the Orgyia pseudotsugata immediate early promoter, the herpes thymidine kinase promoter (Wagner at al., Proc. Natl. Acad. Sci. USA 78:1441-1445 (1981)); or the regulatory sequences of the metallothionein gene (Brinster et al., Nature 296:39-42 (1982)). In some embodiments, the mammalian promoter is a is a constitutive promoter such as, but not limited to, the hypoxanthine phosphoribosyl transferase (HPTR) promoter, the adenosine deaminase promoter, the pyruvate kinase promoter, the beta-actin promoter as well as other constitutive promoters known to those of ordinary skill in the art.

[0085] In some embodiments, a specific promoter may be used to control expression of a transgene in a prokaryotic host cell such as, but are not limited to, the β-lactamase promoter (Villa-Komaroff et al., Proc. Natl. Acad. Sci. USA 75:3727-3731 (1978)); the tac promoter (DeBoer et al., Proc. Natl. Acad. Sci. USA 80:21-25 (1983)); the T7 promoter, the T3 promoter, the M13 promoter or the M16 promoter; in a yeast host cell such as, but are not limited to, the GAL1, GAL4 or GAL10 promoter, the ADH (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, alkaline phosphatase promoter, glyceraldehyde-3-phosphate dehydrogenase III (TDH3) promoter, glyceraldehyde-3-phosphate dehydrogenase II (TDH2) promoter, glyceraldehyde-3-phosphate dehydrogenase I

(TDH1) promoter, pyruvate kinase (PYK), enolase (ENO), or triose phosphate isomerase (TPI).

[0086] In some embodiments, the promoter may be a viral promoter, many of which are able to regulate expression of a transgene in several host cell types, including mammalian cells. Viral promoters that have been shown to drive constitutive expression of coding sequences in eukaryotic cells include, for example, simian virus promoters, herpes simplex virus promoters, papilloma virus promoters, adenovirus promoters, human immunodeficiency virus (HIV) promoters, Rous sarcoma virus promoters, cytomegalovirus (CMV) promoters, the long terminal repeats (LTRs) of Moloney murine leukemia virus and other retroviruses, the thymidine kinase promoter of herpes simplex virus as well as other viral promoters known to those of ordinary skill in the art.

[0087] In some embodiments, the gene control elements of an expression vector may also include 5' non-transcribing and 5' non-translating sequences involved with the initiation of transcription and translation, respectively, such as a TATA box, capping sequence, CAAT sequence, Kozak sequence and the like. Enhancer elements can optionally be used to increase expression levels of a polypeptide or protein to be expressed. Examples of enhancer elements that have been shown to function in mammalian cells include the SV40 early gene enhancer, as described in Dijkema et al., EMBO J. (1985) 4: 761 and the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus (RSV), as described in Gorman et al., Proc. Natl. Acad. Sci. USA (1982b) 79:6777 and human cytomegalovirus, as described in Boshart et al., Cell (1985) 41:521. Genetic control elements of an expression vector will also include 3' non-transcribing and 3'non-translating sequences involved with the termination of transcription and translation. Respectively, such as a poly polyadenylation (polyA) signal for stabilization and processing of the 3' end of an mRNA transcribed from the promoter. Poly A signals included, for example, the rabbit beta globin polyA signal, bovine growth hormone polyA signal, chicken beta globin terminator/polyA signal, or SV40 late polyA region.

Selectable Markers

[0088] Expression vectors will preferably but optionally include at least one selectable marker. In some embodiments, the selectable maker is a nucleic acid sequence

encoding a resistance gene operably linked to one or more genetic regulatory elements, to bestow upon the host cell the ability to maintain viability when grown in the presence of a cyctotoxic chemical and/or drug. In some embodiments, a selectable agent may be used to maintain retention of the expression vector within the host cell. In some embodiments, the selectable agent is may be used to prevent modification (i.e. methylation) and/or silencing of the transgene sequence within the expression vector. In some embodiments, a selectable agent is used to maintain episomal expression of the vector within the host cell. In some embodiments, the selectable agent is used to promote stable integration of the transgene sequence into the host cell genome. In some embodiments, an agent and/or resistance gene may include, but is not limited to, methotrexate (MTX), dihydrofolate reductase (DHFR, U.S. Pat. Nos. 4,399,216; 4,634,665; 4,656,134; 4,956,288; 5,149,636; 5,179,017, ampicillin, neomycin (G418), zeomycin, mycophenolic acid, or glutamine synthetase (GS, U.S. Pat. Nos. 5,122,464; 5,770,359; 5,827,739) for eukaryotic host cell; tetracycline, ampicillin, kanamycin or chlorampenichol for a prokaryotic host cell; and URA3, LEU2, HIS3, LYS2, HIS4, ADE8, CUP1 or TRP1 for a yeast host cell.

[0089] Expression vectors may be transfected, transformed or transduced into a host cell. As used herein, the terms "transfection," "transformation" and "transduction" all refer to the introduction of an exogenous nucleic acid sequence into a host cell. In some embodiments, expression vectors containing nucleic acid sequences encoding for I2S and/or FGE are transfected, transformed or transduced into a host cell at the same time. In some embodiments, expression vectors containing nucleic acid sequences encoding for I2S and/or FGE are transfected, transformed or transduced into a host cell sequentially. For example, a vector encoding an I2S protein may be transfected, transformed or transduced into a host cell first, followed by the transfection, transformation or transduction of a vector encoding an FGE protein, and vice versa. Examples of transformation, transfection and transduction methods, which are well known in the art, include liposome delivery, i.e., lipofectamine™ (Gibco BRL) Method of Hawley-Nelson, Focus 15:73 (1193), electroporation, $CaPO_4$ delivery method of Graham and van der Erb, *Virology*, 52:456-457 (1978), DEAE-Dextran medicated delivery, microinjection, biolistic particle delivery, polybrene mediated delivery, cationic mediated lipid delivery, transduction, and viral infection, such as, e.g., retrovirus, lentivirus, adenovirus adeno-associated virus and Baculovirus (Insect cells). General aspects of cell host transformations have been described in the art, such as by Axel in U.S. Pat. No. 4,399,216; Sambrook, supra, Chapters 1-4 and 16-18; Ausubel, supra, chapters 1, 9, 13, 15,

and 16. For various techniques for transforming mammalian cells, see Keown et al., *Methods in Enzymology* (1989), Keown et al., *Methods in Enzymology,* 185:527-537 (1990), and Mansour et al., *Nature*, 336:348-352 (1988).

[0090] Once introduced inside cells, expression vectors may be integrated stably in the genome or exist as extra-chromosomal constructs. Vectors may also be amplified and multiple copies may exist or be integrated in the genome. In some embodiments, cells of the invention may contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or more copies of nucleic acids encoding an I2S protein. In some embodiments, cells of the invention may contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or more copies of nucleic acids encoding an FGE protein. In some embodiments, cells of the invention may contain multiple copies (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or more) of nucleic acids encoding both I2S and FGE proteins.

Host Cells

[0091] As used herein, the term "host cells" refers to cells that can be used to produce recombinant I2S enzyme. In particular, host cells are suitable for producing recombinant I2S enzyme at a large scale. Suitable host cells can be derived from a variety of organisms, including, but not limited to, mammals, plants, birds (e.g., avian systems), insects, yeast, and bacteria. In some embodiments, host cells are mammalian cells. In some embodiments, a suitable host cell is not a endosomal acidification-deficient cell.

Mammalian Cell Lines

[0092] Any mammalian cell or cell type susceptible to cell culture, and to expression of polypeptides, may be utilized in accordance with the present invention as a host cell. Non-limiting examples of mammalian cells that may be used in accordance with the present invention include human embryonic kidney 293 cells (HEK293), HeLa cells; BALB/c mouse myeloma line (NSO/l, ECACC No: 85110503); human retinoblasts (PER.C6 (CruCell, Leiden, The Netherlands)); monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., J. Gen Virol., 36:59 (1977)); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells +/-DHFR (CHO, Urlaub and Chasin, Proc. Natl. Acad. Sci. USA, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol.

Reprod., 23:243-251 (1980)); monkey kidney cells (CV1 ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1 587); human cervical carcinoma cells (HeLa, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor (MMT 060562, ATCC CCL51); TRI cells (Mather et al., Annals N.Y. Acad. Sci., 383:44-68 (1982)); MRC 5 cells; FS4 cells; and a human hepatoma line (Hep G2). In some embodiments, a suitable mammalian cell is not a endosomal acidification-deficient cell.

[0093] Additionally, any number of commercially and non-commercially available hybridoma cell lines that express polypeptides or proteins may be utilized in accordance with the present invention. One skilled in the art will appreciate that hybridoma cell lines might have different nutrition requirements and/or might require different culture conditions for optimal growth and polypeptide or protein expression, and will be able to modify conditions as needed.

Non-Mammalian Cell Lines

[0094] Any non-mammalian derived cell or cell type susceptible to cell culture, and to expression of polypeptides, may be utilized in accordance with the present invention as a host cell. Non-limiting examples of non-mammalian host cells and cell lines that may be used in accordance with the present invention include cells and cell lines derived from *Pichia pastoris*, *Pichia methanolica*, *Pichia angusta, Schizosacccharomyces pombe*, *Saccharomyces cerevisiae,* and *Yarrowia lipolytica* for yeast; *Sodoptera frugiperda*, *Trichoplusis ni*, *Drosophila melangoster* and *Manduca sexta* for insects; and *Escherichia coli*, *Salmonella typhimurium*, *Bacillus subtilis*, *Bacillus lichenifonnis*, *Bacteroides fragilis*, *Clostridia perfringens*, *Clostridia difficile* for bacteria; and *Xenopus Laevis* from amphibian.

Adaptable to Adherent vs Suspension Growth

[0095] In certain embodiments, a host cell is selected for generating a cell line based on certain preferable attributes or growth under particular conditions chosen for culturing cells. It will be appreciated by one skilled in the art, such attributes may be ascertained based on known characteristic and/or traits of an established line (i.e. a characterized commercially available cell line) or though empirical evaluation. In some embodiments, a cell line may be

selected for its ability to grow on a feeder layer of cells. In some embodiments, a cell line may be selected for its ability to grow in suspension. In some embodiments, a cell line may be selected for its ability to grow as an adherent monolayer of cells. In some embodiments, such cells can be used with any tissue culture vessel or any vessel treated with a suitable adhesion substrate. In some embodiments, a suitable adhesion substrate is selected from the group consisting of collagen (e.g. collagen I, II, II, or IV), gelatin, fibronectin, laminin, vitronectin, fibrinogen, BD Matrigel™, basement membrane matrix, dermatan sulfate proteoglycan, Poly-D-Lysine and/or combinations thereof. In some embodiments, an adherent host cell may be selected and modified under specific growth conditions to grow in suspension. Such methods of modifying an adherent cell to grown in suspension are known in the art. For example, a cell may be conditioned to grow in suspension culture, by gradually removing animal serum from the growth media over time.

Cell Line Selection and Evaluation

[0096] According to the present invention, cells engineered to express recombinant I2S protein are selected for its ability to produce the recombinant I2S protein at commercially viable scale. In particular, engineered cells according to the present invention are able to produce recombinant I2S at a high level and/or with high enzymatic activity. In some embodiments, desirable cells, once cultivated under a cell culture condition (e.g., a standard large scale suspension or adherent culture condition), can produce I2S enzyme in an amount of or greater than about 5 picogram/cell/day (e.g., greater than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 picogram/cell/day). In some embodiments, desired cells, once cultivated under a cell culture condition (e.g., a standard large scale suspension or adherent culture condition), are able to produce I2S enzyme in an amount ranging from about 5-100 picogram/cell/day (e.g., about 5-90 picogram/cell/day, about 5-80 picogram/cell/day, about 5-70 picogram/cell/day, about 5-60 picogram/cell/day, about 5-50 picogram/cell/day, about 5-40 picogram/cell/day, about 5-30 picogram/cell/day, about 10-90 picogram/cell/day, about 10-80 picogram/cell/day, about 10-70 picogram/cell/day, about 10-60 picogram/cell/day, about 10-50 picogram/cell/day, about 10-40 picogram/cell/day, about 10-30 picogram/cell/day, about 20-90 picogram/cell/day, about 20-80 picogram/cell/day, about 20-70 picogram/cell/day, about 20-60 picogram/cell/day, about 20-50 picogram/cell/day, about 20-40 picogram/cell/day, about 20-30 picogram/cell/day).

[0097] As discussed above, typically, the enzyme activity of I2S is influenced by a post-translational modification of a conserved cysteine (e.g., at amino acid 59) to formylglycine. This post-translational modification generally occurs in the endoplasmic reticulum during protein synthesis and is catalyzed by FGE. The enzyme activity of I2S is typically positively correlated with the extent to which the I2S has the formylglycine modification. For example, an I2S preparation that has a relatively high amount of formylglycine modification typically has a relatively high specific enzyme activity; whereas an I2S preparation that has a relatively low amount of formylglycine modification typically has a relatively low specific enzyme activity.

[0098] It is further contemplated that the ratio between the I2S and FGE protein or mRNA may also affect formylglycine modification on the produced recombinant I2S protein. In some embodiments, the I2S and FGE expressed in a desired cell have different protein and/or mRNA expression levels. In some embodiments, the I2S protein or mRNA expression level is at least 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8, 9, or 10-fold higher than the protein or mRNA level of FGE. In some embodiments the recombinant FGE protein or mRNA expression level is at least 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8, 9, or 10-fold higher than the protein or mRNA level of I2S.

[0099] In some embodiments, desirable cells, once cultivated under a cell culture condition (e.g., a standard large scale suspension or adherent culture condition), can produce I2S protein comprising at least about 50% (e.g., at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) conversion of the cysteine residue corresponding to Cys59 of SEQ ID NO:1 to Cα-formylglycine (FGly). In some embodiments, desirable cells, once cultivated under a cell culture condition (e.g., a standard large scale suspension or adherent culture condition), can produce I2S enzyme comprising at least about 50% (e.g., at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) conversion of the cysteine residue corresponding to Cys59 of SEQ ID NO:1 to Cα-formylglycine (FGly) and in an amount of or greater than about 5 picogram/cell/day (e.g., greater than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 picogram/cell/day).

FGly Conversion Percentage

[0100] Various methods are known and can be used to determine the FGly conversion percentage. Generally, the percentage of formylglycine conversion (%FG) can be calculated using the following formula:

$$\%\mathbf{FG}\ (\text{of DS}) = \frac{\text{Number of active I2S molecules}}{\text{Number of total (active+inactive) I2S molecules}} \times 100$$

For example 50% FG means half of the purified recombinant I2S is enzymatically inactive without any therapeutic effect. Various methods may be used to calculate %FG. For example, peptide mapping may be used. Briefly, an I2S protein may be digested into short peptides using a protease (e.g., trypsin or chymotrypsin). Short peptides may be separated and characterized using chromatography (e.g., HPLC) such that the nature and quantity of each peptide (in particular the peptide containing the position corresponding to position 59 of the mature human I2S) may be determined, as compared to a control (e.g., an I2S protein without FGly conversion or an I2S protein with 100% FGly conversion). The amount of peptides containing FGly (corresponding to number of active I2S molecules) and the total amount of peptides with both FGly and Cys (corresponding to number of total I2S molecules) may be determined and the ratio reflecting %FG calculated.

Specific Activity

[0101] As discussed above, typically, the enzyme activity of I2S is influenced by a post-translational modification of a conserved cysteine (e.g., at amino acid 59) to formylglycine. Thus, the enzyme activity of I2S is typically positively correlated with the extent to which the I2S has the formylglycine modification. For example, an I2S preparation that has a relatively high amount of formylglycine modification typically has a relatively high specific enzyme activity; whereas an I2S preparation that has a relatively low amount of formylglycine modification typically has a relatively low specific enzyme activity.

[0102] As can be appreciated by one skilled in the art, the enzymatic activity of recombinant I2S protein produced by cells of the present invention may be measured by various *in vitro* and *in vivo* assays. In some embodiments, a desired enzymatic activity, as

measured by *in vitro* sulfate release activity assay using heparin disaccharide as substrate, of the produced recombinant I2S protein is at least about 20 U/mg, 30 U/mg, 40 U/mg, 50 U/mg, 60 U/mg, 70 U/mg, 80 U/mg, 90 U/mg, or 100 U/mg. In some embodiments, a desired enzymatic activity, as measured by *in vitro* sulfate release activity assay using heparin disaccharide as substrate, of the produced recombinant I2S protein ranges from about 20-100 U/mg (e.g., about 20-90 U/mg, about 20-80 U/mg, about 20-70 U/mg, about 20-60 U/mg, about 20-50 U/mg, about 20-40 U/mg, about 20-30 U/mg, about 30-100 U/mg, about 30-90 U/mg, about 30-80 U/mg, about 30-70 U/mg, about 30-60 U/mg, about 30-50 U/mg, about 30-40 U/mg, about 40-100 U/mg, about 40-90 U/mg, about 40-80 U/mg, about 40-70 U/mg, about 40-60 U/mg, about 40-50 U/mg). Exemplary conditions for performing *in vitro* sulfate release activity assay using heparin disaccharide as substrate are provided below. Typically, this assay measures the ability of I2S to release sulfate ions from a naturally derived substrate, heparin diasaccharide. The released sulfate may be quantified by ion chromatography. In some cases, ion chromatography is equipped with a conductivity detector. As a non-limiting example, samples are first buffer exchanged to 10 mM Na acetate, pH 6 to remove inhibition by phosphate ions in the formulation buffer. Samples are then diluted to 0.075 mg/ml with reaction buffer (10 mM Na acetate, pH 4.4) and incubated for 2 hrs at 37°C with heparin disaccharide at an enzyme to substrate ratio of 0.3 μg I2S/100 μg substrate in a 30 μL reaction volume. The reaction is then stopped by heating the samples at 100°C for 3 min. The analysis is carried out using a Dionex IonPac AS18 analytical column with an IonPac AG18 guard column. An isocratic method is used with 30 mM potassium hydroxide at 1.0 mL/min for 15 minutes. The amount of sulfate released by the I2S sample is calculated from the linear regression analysis of sulfate standards in the range of 1.7 to 16.0 nmoles. The reportable value is expressed as Units per mg protein, where 1 unit is defined as 1 μmoles of sulfate released per hour and the protein concentration is determined by A280 measurements.

[0103] In some embodiments, the enzymatic activity of recombinant I2S protein produced by cells of the present invention may also be determined using various other methods known in the art such as, for example, 4-MUF assay which measures hydrolysis of 4-methylumbelliferyl-sulfate to sulfate and naturally fluorescent 4-methylumbelliferone (4-MUF). In some embodiments, a desired enzymatic activity, as measured by *in vitro* 4-MUF assay, of the produced recombinant I2S protein is at least about 2 U/mg, 4 U/mg, 6 U/mg, 8 U/mg, 10 U/mg, 12 U/mg, 14 U/mg, 16 U/mg, 18 U/mg, or 20 U/mg. In some embodiments,

a desired enzymatic activity, as measured by *in vitro* 4-MUF assay, of the produced recombinant I2S protein ranges from about 0-50 U/mg (e.g., about 0-40 U/mg, about 0-30 U/mg, about 0-20 U/mg, about 0-10 U/mg, about 2-50 U/mg, about 2-40 U/mg, about 2-30 U/mg, about 2-20 U/mg, about 2-10 U/mg, about 4-50 U/mg, about 4-40 U/mg, about 4-30 U/mg, about 4-20 U/mg, about 4-10 U/mg, about 6-50 U/mg, about 6-40 U/mg, about 6-30 U/mg, about 6-20 U/mg, about 6-10 U/mg). Exemplary conditions for performing *in vitro* 4-MUF assay are provided below. Typically, a 4-MUF assay measures the ability of an I2S protein to hydrolyze 4-methylumbelliferyl-sulfate (4-MUF-SO_4) to sulfate and naturally fluorescent 4-methylumbelliferone (4-MUF). One milliunit of activity is defined as the quantity of enzyme required to convert one nanomole of 4-MUF-SO_4 to 4-MUF in one minute at 37°C. Typically, the mean fluorescence units (MFU) generated by I2S test samples with known activity can be used to generate a standard curve, which can be used to calculate the enzymatic activity of a sample of interest.

Cell Culture Medium and Condition

[0104] Various cell culture medium and conditions may be used to produce a recombinant I2S protein using engineered cells according to the present invention. For example, a recombinant I2S protein may be produced in serum-containing or serum-free medium. In some embodiments, a recombinant I2S protein is produced in serum-free medium. In some embodiments, a recombinant I2S protein is produced in an animal free medium, i.e., a medium that lacks animal-derived components. In some embodiments, a recombinant I2S protein is produced in a chemically defined medium. As used herein, the term "chemically-defined nutrient medium" refers to a medium of which substantially all of the chemical components are known. In some embodiments, a chemically defined nutrient medium is free of animal-derived components such as serum, serum derived proteins (e.g., albumin or fetuin), and other components. In some cases, a chemically-defined medium comprises one or more proteins (*e.g.*, protein growth factors or cytokines.) In some cases, a chemically-defined nutrient medium comprises one or more protein hydrolysates. In other cases, a chemically-defined nutrient medium is a protein-free media, i.e., a serum-free media that contains no proteins, hydrolysates or components of unknown composition.

[0105] In some embodiments, a chemically defined medium may be supplemented by one or more animal derived components. Such animal derived components include, but are not limited to, fetal calf serum, horse serum, goat serum, donkey serum, human serum, and serum derived proteins such as albumins (e.g., bovine serum albumin or human serum albumin).

[0106] Various cell culture conditions may be used to produce recombinant I2S proteins at large scale including, but not limited to, roller bottle cultures, bioreactor batch cultures and bioreactor fed-batch cultures. In some embodiments, recombinant I2S protein is produced by cells cultured in suspension. In some embodiments, recombinant I2S protein is produced by adherent cells.

[0107] Exemplary cell media and culture conditions are described in the Examples sections. Additional exemplary methods and compositions for producing recombinant I2S protein are described in the provisional application entitled "Methods and Compositions for

Producing Recombinant Iduronate-2-Sulfatase" filed herewith on even date, the entire disclosure of which is hereby incorporated by reference.

Purification of Expressed I2S Protein

[0108] Various methods may be used to purify or isolate I2S protein produced according to various methods described herein. In some embodiments, the expressed I2S protein is secreted into the medium and thus cells and other solids may be removed, as by centrifugation or filtering for example, as a first step in the purification process. Alternatively or additionally, the expressed I2S protein is bound to the surface of the host cell. In this embodiment, the host cells expressing the polypeptide or protein are lysed for purification. Lysis of mammalian host cells can be achieved by any number of means well known to those of ordinary skill in the art, including physical disruption by glass beads and exposure to high pH conditions.

[0109] The I2S protein may be isolated and purified by standard methods including, but not limited to, chromatography (e.g., ion exchange, affinity, size exclusion, and hydroxyapatite chromatography), gel filtration, centrifugation, or differential solubility, ethanol precipitation or by any other available technique for the purification of proteins (See, e.g., Scopes, Protein Purification Principles and Practice 2nd Edition, Springer-Verlag, New York, 1987; Higgins, S. J. and Hames, B. D. (eds.), Protein Expression: A Practical Approach, Oxford Univ Press, 1999; and Deutscher, M. P., Simon, M. I., Abelson, J. N. (eds.), Guide to Protein Purification: Methods in Enzymology (Methods in Enzymology Series, Vol 182), Academic Press, 1997, all incorporated herein by reference). For immunoaffinity chromatography in particular, the protein may be isolated by binding it to an affinity column comprising antibodies that were raised against that protein and were affixed to a stationary support. Alternatively, affinity tags such as an influenza coat sequence, poly-histidine, or glutathione-S-transferase can be attached to the protein by standard recombinant techniques to allow for easy purification by passage over the appropriate affinity column. Protease inhibitors such as phenyl methyl sulfonyl fluoride (PMSF), leupeptin, pepstatin or aprotinin may be added at any or all stages in order to reduce or eliminate degradation of the polypeptide or protein during the purification process. Protease inhibitors are particularly desired when cells must be lysed in order to isolate and purify the expressed polypeptide or protein.

[0110] Exemplary purification methods are described in the Examples sections below. Additional purification methods are described in the provisional application entitled "Purification of Recombinant I2S Protein" filed on herewith on even date, the entire disclosure of which is hereby incorporated by reference.

Pharmaceutical Composition and Administration

[0111] Purified recombinant I2S protein may be administered to a Hunter Syndrome patient in accordance with known methods. For example, purified recombinant I2S protein may be delivered intravenously, subcutaneously, intramuscularly, parenterally, transdermally, or transmucosally (*e.g.*, orally or nasally)).

[0112] In some embodiments, a recombinant I2S or a pharmaceutical composition containing the same is administered to a subject by intravenous administration.

[0113] In some embodiments, a recombinant I2S or a pharmaceutical composition containing the same is administered to a subject by intrathecal administration. As used herein, the term "intrathecal administration" or "intrathecal injection" refers to an injection into the spinal canal (intrathecal space surrounding the spinal cord). Various techniques may be used including, without limitation, lateral cerebroventricular injection through a burrhole or cisternal or lumbar puncture or the like. In some embodiments, "intrathecal administration" or "intrathecal delivery" according to the present invention refers to IT administration or delivery via the lumbar area or region, i.e., lumbar IT administration or delivery. As used herein, the term "lumbar region" or "lumbar area" refers to the area between the third and fourth lumbar (lower back) vertebrae and, more inclusively, the L2-S1 region of the spine.

[0114] In some embodiments, a recombinant I2S or a pharmaceutical composition containing the same is administered to the subject by subcutaneous (i.e., beneath the skin) administration. For such purposes, the formulation may be injected using a syringe. However, other devices for administration of the formulation are available such as injection devices (e.g., the Inject-ease™ and Genject™ devices); injector pens (such as the GenPen™); needleless devices (e.g., MediJector™ and BioJector™); and subcutaneous patch delivery systems.

[0115] In some embodiments, intrathecal administration may be used in conjunction with other routes of administration (e.g., intravenous, subcutaneously, intramuscularly, parenterally, transdermally, or transmucosally (*e.g.*, orally or nasally)).

[0116] The present invention contemplates single as well as multiple administrations of a therapeutically effective amount of a recombinant I2S or a pharmaceutical composition containing the same described herein. A recombinant I2S or a pharmaceutical composition containing the same can be administered at regular intervals, depending on the nature, severity and extent of the subject's condition (e.g., a lysosomal storage disease). In some embodiments, a therapeutically effective amount of a recombinant I2S or a pharmaceutical composition containing the same may be administered periodically at regular intervals (e.g., once every year, once every six months, once every five months, once every three months, bimonthly (once every two months), monthly (once every month), biweekly (once every two weeks), weekly, daily or continuously).

[0117] A recombinant I2S or a pharmaceutical composition containing the same can be formulated with a physiologically acceptable carrier or excipient to prepare a pharmaceutical composition. The carrier and therapeutic agent can be sterile. The formulation should suit the mode of administration.

[0118] Suitable pharmaceutically acceptable carriers include but are not limited to water, salt solutions (*e.g.*, NaCl), saline, buffered saline, alcohols, glycerol, ethanol, gum arabic, vegetable oils, benzyl alcohols, polyethylene glycols, gelatin, carbohydrates such as lactose, amylose or starch, sugars such as mannitol, sucrose, or others, dextrose, magnesium stearate, talc, silicic acid, viscous paraffin, perfume oil, fatty acid esters, hydroxymethylcellulose, polyvinyl pyrolidone, *etc.*, as well as combinations thereof. The pharmaceutical preparations can, if desired, be mixed with auxiliary agents (*e.g.*, lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic pressure, buffers, coloring, flavoring and/or aromatic substances and the like) which do not deleteriously react with the active compounds or interference with their activity. In some embodiments, a water-soluble carrier suitable for intravenous administration is used.

[0119] The composition or medicament, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents. The composition can be a liquid solution, suspension, emulsion, tablet, pill, capsule, sustained release formulation, or powder.

The composition can also be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulation can include standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, polyvinyl pyrollidone, sodium saccharine, cellulose, magnesium carbonate, *etc.*

[0120] The composition or medicament can be formulated in accordance with the routine procedures as a pharmaceutical composition adapted for administration to human beings. For example, in some embodiments, a composition for intravenous administration typically is a solution in sterile isotonic aqueous buffer. Where necessary, the composition may also include a solubilizing agent and a local anesthetic to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampule or sachette indicating the quantity of active agent. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water, saline or dextrose/water. Where the composition is administered by injection, an ampule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

[0121] As used herein, the term "therapeutically effective amount" is largely determined base on the total amount of the therapeutic agent contained in the pharmaceutical compositions of the present invention. Generally, a therapeutically effective amount is sufficient to achieve a meaningful benefit to the subject (e.g., treating, modulating, curing, preventing and/or ameliorating the underlying disease or condition). For example, a therapeutically effective amount may be an amount sufficient to achieve a desired therapeutic and/or prophylactic effect, such as an amount sufficient to modulate lysosomal enzyme receptors or their activity to thereby treat such lysosomal storage disease or the symptoms thereof (e.g., a reduction in or elimination of the presence or incidence of "zebra bodies" or cellular vacuolization following the administration of the compositions of the present invention to a subject). Generally, the amount of a therapeutic agent (e.g., a recombinant lysosomal enzyme) administered to a subject in need thereof will depend upon the characteristics of the subject. Such characteristics include the condition, disease severity, general health, age, sex and body weight of the subject. One of ordinary skill in the art will be readily able to determine appropriate dosages depending on these and other related factors.

In addition, both objective and subjective assays may optionally be employed to identify optimal dosage ranges.

[0122] A therapeutically effective amount is commonly administered in a dosing regimen that may comprise multiple unit doses. For any particular therapeutic protein, a therapeutically effective amount (and/or an appropriate unit dose within an effective dosing regimen) may vary, for example, depending on route of administration, on combination with other pharmaceutical agents. Also, the specific therapeutically effective amount (and/or unit dose) for any particular patient may depend upon a variety of factors including the disorder being treated and the severity of the disorder; the activity of the specific pharmaceutical agent employed; the specific composition employed; the age, body weight, general health, sex and diet of the patient; the time of administration, route of administration, and/or rate of excretion or metabolism of the specific fusion protein employed; the duration of the treatment; and like factors as is well known in the medical arts.

[0123] Additional exemplary pharmaceutical compositions and administration methods are described in PCT Publication WO2011/163649 entitled "Methods and Compositions for CNS Delivery of Iduronate-2-Sulfatase;" and provisional application serial no. 61/618,638 entitled "Subcutaneous administration of iduronate 2 sulfatase" filed on March 30, 2012, the entire disclosures of both of which are hereby incorporated by reference.

[0124] It is to be further understood that for any particular subject, specific dosage regimens should be adjusted over time according to the individual need and the professional judgment of the person administering or supervising the administration of the enzyme replacement therapy and that dosage ranges set forth herein are exemplary only and are not intended to limit the scope or practice of the claimed invention.

EXAMPLES

Example 1. Generation of Optimized Cell Line Co-expressing recombinant I2S and FGE

[0125] This example illustrates an exemplary optimized cell line co-expressing recombinant I2S and FGE that can be used to produce recombinant I2S protein. It will be clear to one skilled in the art, that a number of alternative approaches, expression vectors and cloning techniques are available.

[0126] A typical mature form of human iduronate-2-sulfatase enzyme (I2S) is a 525-amino acid glycoprotein that undergoes extensive processing and post translational modification for enzyme activation, such as glycosylation and cysteine conversion to formylglycine (Figure 1). In mammalian cells, conserved cysteine residues within the I2S (i.e., at amino acid 59) enzyme are converted to formylglycine by the formylglycine generating enzyme (FGE). The conversion of cysteine to formylglycine within the active site of the I2S enzyme is an important step in generating the active form of the human sulfatase enzyme. The purpose of this experiment was to engineer an optimized human cell line co-expressing I2S and FGE for generating active recombinant I2S.

[0127] Figure 2 illustrates a number of exemplary construct designs for co-expression of I2S and FGE. For example, expression units of I2S and FGE can be located on separate vectors and the separate vectors can be co-transfected or transfected separately (Figure 2A). Alternatively, expression units of I2S and FGE can be located on the same vector (Figure 2B). In one configuration, I2S and FGE can be on the same vector but under the control of separate promoters, also referred to as separate cistrons (Figure 2B(1)). Alternatively, I2S and FGE can be designed as transcriptionally linked cistrons, that is, I2S and FGE are designed as one open reading frame under the control of a same promoter (Figure 2B(2)). Typically, an internal ribosome entry site (IRES) is designed to allow cap independent translation initiation of the messenger RNA (Figure 2B(2)).

[0128] A human cell line was engineered to co-express human I2S protein with the amino acid sequence shown in SEQ ID NO:2 and human formylglycine generating enzyme (FGE) with the amino acid sequence shown in SEQ ID NO:6.

SEQ ID NO: 2
> Full-lenth Precursor iduronate 2-sulfatase
MPPPRTGRGLLWLGLVLSSVCVALGSETQANSTTDALNVLLIIVDDLRPSLGCYGDK
LVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYDFNSYWRVHAG
NFSTIPQYFKENGYVTMSVGKVFHPGISSNHTDDSPYSWSFPPYHPSSEKYENTKTCR
GPDGELHANLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSASPFFLAVGYHKPH
IPFRYPKEFQKLYPLENITLAPDPEVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPI
PVDFQRKIRQSYFASVSYLDTQVGRLLSALDDLQLANSTIIAFTSDHGWALGEHGEW
AKYSNFDVATHVPLIFYVPGRTASLPEAGEKLFPYLDPFDSASQLMEPGRQSMDLVE
LVSLFPTLAGLAGLQVPPRCPVPSFHVELCREGKNLLKHFRFRDLEEDPYLPGNPREL
IAYSQYPRPSDIPQWNSDKPSLKDIKIMGYSIRTIDYRYTVWVGFNPDEFLANFSDIHA
GELYFVDSDPLQDHNMYNDSQGGDLFQLLMP

SEQ ID NO:6
Full-length human FGE precursor:
MAAPALGLVCGRCPELGLVLLLLLLSLLCGAAGSQEAGTGAGAGSLAGSCGCGTPQ
RPGAHGSSAAAHRYSREANAPGPVPGERQLAHSKMVPIPAGVFTMGTDDPQIKQDG
EAPARRVTIDAFYMDAYEVSNTEFEKFVNSTGYLTEAEKFGDSFVFEGMLSEQVKTN
IQQAVAAAPWWLPVKGANWRHPEGPDSTILHRPDHPVLHVSWNDAVAYCTWAGK
RLPTEAEWEYSCRGGLHNRLFPWGNKLQPKGQHYANIWQGEFPVTNTGEDGFQGT
APVDAFPPNGYGLYNIVGNAWEWTSDWWTVHHSVEETLNPKGPPSGKDRVKKGGS
YMCHRSYCYRYRCAARSQNTPDSSASNLGFRCAADRLPTMD

[0129] To generate an I2S expressing cell line, cells were stably transfected with a codon optimized nucleic acid sequence (SEQ ID NO. 7) encoding an I2S protein with the amino acid sequence shown in SEQ ID NO:2 and a nucleic acid sequence (SEQ ID NO. 8) encoding the human FGE enzyme as set forth in SEQ ID NO. 6.

SEQ ID NO: 7
> Homo sapiens codon optimized iduronate 2-sulfatase (IDS), transcript variant 1, mRNA

ATGCCCCCGCCCCGCACCGGCCGCGGCCTGCTGTGGCTGGGCCTGGTGCTGAGCAGCGTGTGCGTG
GCCCTGGGCAGCGAGACCCAGGCCAACAGCACCACCGACGCCCTGAACGTGCTGCTGATCATCGT
GGACGACCTGCGCCCCAGCCTGGGCTGCTACGGCGACAAGCTGGTGCGCAGCCCCAACATCGACC
AGCTGGCCAGCCACAGCCTGCTGTTCCAGAACGCCTTCGCCCAGCAGGCCGTGTGCGCCCCCAGCC
GCGTGAGCTTCCTGACCGGCCGCCGCCCCGACACCACCCGCCTGTACGACTTCAACAGCTACTGGC
GCGTGCACGCCGGCAACTTCAGCACCATCCCCCAGTACTTCAAGGAGAACGGCTACGTGACCATG
AGCGTGGGCAAGGTGTTCCACCCCGGCATCAGCAGCAACCACACCGACGACAGCCCCTACAGCTG
GAGCTTCCCCCCCTACCACCCCAGCAGCGAGAAGTACGAGAACACCAAGACCTGCCGCGGCCCCG
ACGGCGAGCTGCACGCCAACCTGCTGTGCCCCGTGGACGTGCTGGACGTGCCCGAGGGCACCCTG
CCCGACAAGCAGAGCACCGAGCAGGCCATCCAGCTGCTGGAGAAGATGAAGACCAGCGCCAGCC
CCTTCTTCCTGGCCGTGGGCTACCACAAGCCCCACATCCCCTTCCGCTACCCCAAGGAGTTCCAGA
AGCTGTACCCCCTGGAGAACATCACCCTGGCCCCCGACCCCGAGGTGCCCGACGGCCTGCCCCCCG
TGGCCTACAACCCCTGGATGGACATCCGCCAGCGCGAGGACGTGCAGGCCCTGAACATCAGCGTG
CCCTACGGCCCCATCCCCGTGGACTTCCAGCGCAAGATCCGCCAGAGCTACTTCGCCAGCGTGAGC
TACCTGGACACCCAGGTGGGCCGCCTGCTGAGCGCCCTGGACGACCTGCAGCTGGCCAACAGCAC
CATCATCGCCTTCACCAGCGACCACGGCTGGGCCCTGGGCGAGCACGGCGAGTGGGCCAAGTACA

GCAACTTCGACGTGGCCACCCACGTGCCCCTGATCTTCTACGTGCCCGGCCGCACCGCCAGCCTGC
CCGAGGCCGGCGAGAAGCTGTTCCCCTACCTGGACCCCTTCGACAGCGCCAGCCAGCTGATGGAG
CCCGGCCGCCAGAGCATGGACCTGGTGGAGCTGGTGAGCCTGTTCCCCACCCTGGCCGGCCTGGCC
GGCCTGCAGGTGCCCCCCCGCTGCCCCGTGCCCAGCTTCCACGTGGAGCTGTGCCGCGAGGGCAA
GAACCTGCTGAAGCACTTCCGCTTCCGCGACCTGGAGGAGGACCCCTACCTGCCCGGCAACCCCCG
CGAGCTGATCGCCTACAGCCAGTACCCCCGCCCCAGCGACATCCCCCAGTGGAACAGCGACAAGC
CCAGCCTGAAGGACATCAAGATCATGGGCTACAGCATCCGCACCATCGACTACCGCTACACCGTG
TGGGTGGGCTTCAACCCCGACGAGTTCCTGGCCAACTTCAGCGACATCCACGCCGGCGAGCTGTAC
TTCGTGGACAGCGACCCCCTGCAGGACCACAACATGTACAACGACAGCCAGGGCGGCGACCTGTT
CCAGCTGCTGATGCCCTAG

SEQ ID NO: 8
> Homo sapiens Full-length Precursor formylglycine generating enzyme (FGE), mRNA

ATGGCTGCGCCCGCACTAGGGCTGGTGTGTGGACGTTGCCCTGAGCTGGGTCTCGTCCTCTTGCTG
CTGCTGCTCTCGCTGCTGTGTGGAGCGGCAGGGAGCCAGGAGGCCGGGACCGGTGCGGGCGCGGG
GTCCCTTGCGGGTTCTTGCGGCTGCGGCACGCCCCAGCGGCCTGGCGCCCATGGCAGTTCGGCAGC
CGCTCACCGATACTCGCGGGAGGCTAACGCTCCGGGCCCCGTACCCGGAGAGCGGCAACTCGCGC
ACTCAAAGATGGTCCCCATCCCTGCTGGAGTATTTACAATGGGCACAGATGATCCTCAGATAAAGC
AGGATGGGGAAGCACCTGCGAGGAGAGTTACTATTGATGCCTTTTACATGGATGCCTATGAAGTC
AGTAATACTGAATTTGAGAAGTTTGTGAACTCAACTGGCTATTTGACAGAGGCTGAGAAGTTTGGC
GACTCCTTTGTCTTTGAAGGCATGTTGAGTGAGCAAGTGAAGACCAATATTCAACAGGCAGTTGCA
GCTGCTCCCTGGTGGTTACCTGTGAAAGGCGCTAACTGGAGACACCCAGAAGGGCCTGACTCTACT
ATTCTGCACAGGCCGGATCATCCAGTTCTCCATGTGTCCTGGAATGATGCGGTTGCCTACTGCACTT
GGGCAGGGAAGCGGCTGCCCACGGAAGCTGAGTGGGAATACAGCTGTCGAGGAGGCCTGCATAA
TAGACTTTTCCCCTGGGGCAACAAACTGCAGCCCAAAGGCCAGCATTATGCCAACATTTGGCAGG
GCGAGTTTCCGGTGACCAACACTGGTGAGGATGGCTTCCAAGGAACTGCGCCTGTTGATGCCTTCC
CTCCCAATGGTTATGGCTTATACAACATAGTGGGGAACGCATGGGAATGGACTTCAGACTGGTGG
ACTGTTCATCATTCTGTTGAAGAAACGCTTAACCCAAAAGGTCCCCCTTCTGGGAAAGACCGAGTG
AAGAAAGGTGGATCCTACATGTGCCATAGGTCTTATTGTTACAGGTATCGCTGTGCTGCTCGGAGC
CAGAACACACCTGATAGCTCTGCTTCGAATCTGGGATTCCGCTGTGCAGCCGACCGCCTGCCCACC
ATGGACTGA

[0130] Both I2S- and FGE- encoding nucleic acid sequences are controlled by a human CMV promoter. Translation of I2S mRNA results in synthesis of a 550 amino acid full length I2S protein (SEQ ID NO:2), which includes a 25 amino acid signal peptide. The signal peptide is removed and a soluble enzyme is secreted from the cell.

[0131] The bacterial neomycin phosphotransferase (neo) coding sequence and/or Blasticidin S Deaminase (BSD) gene were used to allow for selection of transfected cells using the neomycin analog G418 and/or blasticidin, respectively. In addition, the mouse dihydrofolate reductase (DHFR) gene was used on the I2S- and/or FGE-encoding vector(s) to allow for isolation of cell lines containing increased copies of the I2S- and/or FGE-encoding sequences by methotrexate (MTX) selection.

[0132] Cells producing I2S were isolated and subjected to appropriate drug selection to isolate cells with an increased number of copies of the transfected I2S and/or FGE genes. Quantification of I2S was performed by ELISA.

[0133] The cell population was also subjected to step-wise selection in methotrexate (MTX) to isolate cells with increased I2S productivity. I2S productivity was monitored during MTX selection by ELISA.

[0134] After several rounds of propagation, several I2S producing clones were then subjected to suspension adaptation in serum-free media through a stepwise reduction from DMEM containing 10% calf serum to serum free chemically defined media. Several individual clonal populations were established through limited dilution cloning. Colonies were screened by I2S enzyme activity assay and ELISA. Two stable cell lines 2D and 4D showed high percent viability and robust expression of I2S and were selected for further development.

Example 2. Evaluation of Stable Cell Lines Co-expressing I2S and FGE

[0135] Additional experiments were carried out to characterize two cell lines 2D and 4D co-expressing I2S and FGE.

Specific Activity

[0136] First specific activity of the I2S enzyme was evaluated. I2S enzyme produced from the 2D and 4D cell lines were analyzed for specific activity using a fluorescence based 4-MUF assay. Briefly, the assay measures the hydrolysis of I2S substrate 4-methylumbelliferyl-sulfate (4-MUF-SO_4). Upon cleavage of the 4-MUF-SO_4 substrate by I2S, the molecule is converted to sulfate and naturally fluorescent 4-methylumbelliferone (4-MUF). As a result, I2S enzyme activity can be determined by evaluating the overall change in fluorescent signal over time. For this experiment, purified I2S enzyme produced from the I2S-AF 2D and 4D human cell lines were incubated with a solution of 4-methylumbelliferyl-sulfate (4-MUF-SO_4), Potassium Salt, Sigma Cat. # M-7133). Calibration of the assay was performed using a series of control reference samples, using commercially available I2S enzyme diluted at 1:100, 1:200 and 1:20,000 of the stock solution. The enzymatic assay was run at 37°C and assayed using a calibrated fluorometer. Using the fluorescence values obtained for each reference standard, the percent coefficient of variation was determined using the following equation:

$$\%CV = \frac{S\tan dard\ Deviation of\ Raw Fluorescenc Values(N=3)}{Averagel\ Fluorescence Value} X\,100\%$$

[0137] The percent CV values were then used to calculate the Corrected Average Fluorescence for each sample, in order to determine the reportable enzyme activity, expressed in mU/mL using the following formula:

$$mU/mL = (CFU)\left(\frac{1nmole/L}{10FU}\right)\left(\frac{1L}{10^3 mL}\right)\left(\frac{2.11mL}{0.01mL}\right)\left(\frac{1hour}{60\min}\right)\left(\frac{1mU}{nmole}\right)(DF)$$

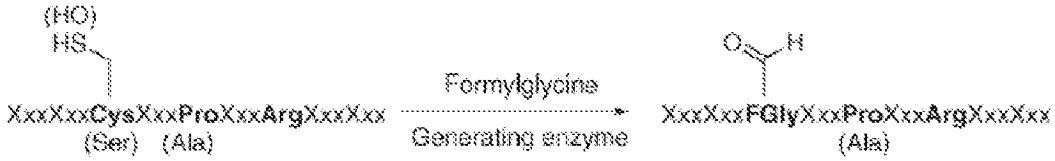
CFU = Negative corrected average fluorescence

DF - Dilution Factor

[0138] One milliunit of activity is the quantity of enzyme required to convert 1 nanomole of 4-methylumbelliferyl-sulfate to 4-methylumbelliferone in 1 minute at 37°C.

Percent Formylglycine Conversion

[0139] Peptide mapping can be used to determine Percent FGly conversion. I2S activation requires Cysteine (corresponding to position 59 of mature human I2S) to formylglycine conversion by formylglycine generating enzyme (FGE) as shown below:



Therefore, the percentage of formylglycine conversion (%FG) can be calculated using the following formula:

$$\%FG\ (\text{of DS}) = \frac{\text{Number of active I2S molecules}}{\text{Number of total (active+inactive) I2S molecules}} \times 100$$

[0140] For example 50% FG means half of the purified recombinant I2S is enzymatically inactive without any therapeutic effect.

[0141] Peptide mapping was used to calculate %FG. Briefly, a recombinant I2S protein was digested into short peptides using a protease (e.g., trypsin or chymotrypsin). Short peptides were separated and characterized using HPLC. The peptide containing the position corresponding to position 59 of the mature human I2S was characterized to determine if the Cys at position 59 was converted to a FGly as compared to a control (e.g., an I2S protein without FGly conversion or an I2S protein with 100% FGly conversion). The amount of peptides containing FGly (corresponding to number of active I2S molecules) and the total amount of peptides with both FGly and Cys (corresponding to number of total I2S molecules) may be determined based on the corresponding peak areas and the ratio reflecting %FG was calculated.

Correlation between Percentage FGly Conversion and Specific Activity

[0142] Exemplary correlation between percentage FGly conversion and specific activity is shown in Figure 3. As can be seen, the data suggest that a higher percentage of formylglycine conversion results in higher I2S enzyme activity.

Glycan Map

[0143] The glycan composition of recombinant I2S protein produced by cell line 2D and 4D was determined. Quantification of the glycan composition was performed, using anion exchange chromatography. As described below, the glycan map of recombinant I2S generated under these conditions consists of seven peak groups, eluting according to an increasing amount of negative charges, at least partly derived from sialic acid and mannose-6-phosphate glycoforms resulting from enzymatic digest. Briefly, purified recombinant I2S obtained using the serum-free cell culture method (I2S-AF 2D Serum-free and I2S-AF 4D Serum-free) and reference recombinant I2S produced, were treated with either (1) purified neuraminidase enzyme (isolated from Arthrobacter Ureafaciens (10 mU/μL), Roche Biochemical (Indianapolis, IN), Cat. # 269 611 (1U/100 μL)) for the removal of sialic acid residues, (2) alkaline phosphatase for 2 hours at 37±1°C for complete release of mannose-6-phosphate residues, (3) alkaline phosphatase + neuraminidase, or (4) no treatment. Each enzymatic digest was analyzed by High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAE-PAD) using a CarboPac PA1 Analytical Column

equipped with a Dionex CarboPac PA1 Guard Column. A series of sialic acid and mannose-6-phosphate standards in the range of 0.4 to 2.0 nmoles were run for each assay. An isocratic method using 48 mM sodium acetate in 100 mM sodium hydroxide was run for a minimum of 15 minutes at a flow rate of 1.0 mL/min at ambient column temperature to elute each peak. The data generated from each individual run, for both the I2S-AF and reference I2S samples, were each combined into a single chromatograph to represent the glycan map for each respective recombinant protein. As indicated in Figure 4, an exemplary glycan map for I2S produced by cell line 2D and 4D displayed representative elution peaks (in the order of elution) constituting neutrals, mono-, disialyated, monophosphorylated, trisialyated and hybrid (monosialyated and capped mannose-6-phosphate), tetrasialylated and hybrid (disilaylated and capped mannose-6-phosphate) and diphosphorylated glycans.

Example 3. Serum-free Suspension Cell Culture

[0144] This example demonstrates that a large scale serum free suspension culture may be developed to cultivate an optimized cell line to produce recombinant I2S.

Serum-free suspension cell culture system

[0145] Briefly, a seed culture was established using the 2D or 4D cell line of Example 1. Cells were transferred to a 250 mL tissue culture shake flask containing serum-free chemically defined expansion medium supplemented with MTX for selection, adjusted with sodium bicarbonate to a pH of 7.3 and grown at 37°C at 5% CO_2 for several days. Once the culture reached a sufficient cell density and viability, the initial seed culture was used to inoculate the first of a series of step-wise cell culture expansions in 500 mL tissue culture shake flasks followed by 1 L tissue culture shake flasks.

[0146] A batch culture expansion was performed by transferring each of the 1L cultures into a 10L Cellbag bioreactor® (Wave Europe), and adding expansion medium. After reaching a sufficient cell density, new expansion medium was added and the cells grown to a sufficient density. The 10L Cellbag was transferred to a Wave bioreactor® system (Wave Europe) and culture conditions were modified to allow for growth under continuous medium perfusion. Expansion growth medium was delivered and samples were collected for off-line metabolite analysis of pH, glutamine, glutamate, glucose, ammonium, lactate, pCO_2 and osmolarity.

[0147] Upon reaching a sufficient cell density, the entire 10L cell culture was transferred to a 50L Wave Cellbag bioreactor®, containing fresh expansion medium, and grown to a sufficient cell density using a Wave bioreactor® system.

[0148] Cell expansion was next performed using a 200L disposable bioreactor and centrifuge perfusion device (Centritech® CELL II unit, Pneumatic Scale Corporation), which was designed to concentrate cells and clarify media for recycling during perfusion mediated cell culture. Expansion medium (adjusted to pH 7.10) was inoculated with a portion of the 50L culture and grown to a sufficient cell density.

[0149] Next a portion of the 200L culture was used to seed a 2000L disposable bioreactor and centrifuge perfusion device (Centritech® CELL II unit, Pneumatic Scale Corporation) in production medium (adjusted to pH 7.20). Cells were grown under batch growth conditions. Following the two day growth, conditions were adjusted for continuous perfusion, until a transition phase was reached. Cells were grown under perfusion growth conditions for the 24 hour transition phase.

[0150] For the production phase, two Centritech CELL II units were used. Production phase was started approximately 24 hours after the start of the transition phase and maintained for a desired period, by regulating the bleed rate.

[0151] While certain compounds, compositions and methods described herein have been described with specificity in accordance with certain embodiments, the following examples serve only to illustrate the compounds of the invention and are not intended to limit the same.

[0152] The articles "a" and "an" as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to include the plural referents. Claims or descriptions that include "or" between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in

which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention also includes embodiments in which more than one, or the entire group members are present in, employed in, or otherwise relevant to a given product or process. Furthermore, it is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, descriptive terms, etc., from one or more of the listed claims is introduced into another claim dependent on the same base claim (or, as relevant, any other claim) unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise. Where elements are presented as lists, (e.g., in Markush group or similar format) it is to be understood that each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It should be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements, features, etc., certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements, features, etc. For purposes of simplicity those embodiments have not in every case been specifically set forth in so many words herein. It should also be understood that any embodiment or aspect of the invention can be explicitly excluded from the claims, regardless of whether the specific exclusion is recited in the specification. The publications, websites and other reference materials referenced herein to describe the background of the invention and to provide additional detail regarding its practice are hereby incorporated by reference.

We claim:

1. A cell comprising

a first nucleic acid encoding an iduronate-2-sulfatase (I2S) protein comprising an amino acid sequence at least 70% identical to SEQ ID NO:1; and

a second nucleic acid encoding a formylglycine generating enzyme (FGE) protein comprising an amino acid sequence at least 70% identical to SEQ ID NO:5,

wherein the first and/or the second nucleic acid are exogenous and wherein the cell, once cultivated under a cell culture condition, produces I2S protein comprising at least about 70% conversion of the cysteine residue corresponding to Cys59 of SEQ ID NO:1 to Cα-formylglycine (FGly).

2. A cell comprising

a first nucleic acid encoding an iduronate-2-sulfatase (I2S) protein comprising an amino acid sequence at least 70% identical to SEQ ID NO:1; and

a second nucleic acid encoding a formylglycine generating enzyme (FGE) protein comprising an amino acid sequence at least 70% identical to SEQ ID NO:5,

wherein the first and/or the second nucleic acid are exogenous and wherein the cell, once cultivated under a cell culture condition, produces I2S protein comprising at least about 60% conversion of the cysteine residue corresponding to Cys59 of SEQ ID NO:1 to Cα-formylglycine (FGly) and at a specific productivity rate of great than about 30 picogram/cell/day.

3. The cell of claim 1 or 2, wherein the cell, once cultivated under a cell culture condition, produces the I2S protein comprising at least about 80% conversion of the cysteine residue corresponding to Cys59 of SEQ ID NO:1 to Cα-formylglycine (FGly).

4. The cell of any one of the preceding claims, wherein the cell, once cultivated under a cell culture condition, produces the I2S protein comprising at least about 90% conversion of the cysteine residue corresponding to Cys59 of SEQ ID NO:1 to Cα-formylglycine (FGly).

5. The cell of any one of the preceding claims, wherein the cell, once cultivated under a cell culture condition, produces the I2S protein comprising at least about 95% conversion of the cysteine residue corresponding to Cys59 of SEQ ID NO:1 to Cα-formylglycine (FGly).

6. The cell of any one of the preceding claims, wherein the cell, once cultivated under a cell culture condition, produces the I2S protein comprising at least about 97% conversion of the cysteine residue corresponding to Cys59 of SEQ ID NO:1 to Cα-formylglycine (FGly).

7. The cell of any one of the preceding claims, wherein the first and/or the second nucleic acid is operably linked to a hCMV promoter.

8. The cell of any one of the preceding claims, wherein the first nucleic acid encodes an I2S protein having an amino acid sequence identical to SEQ ID NO:1.

9. The cell of any one of the preceding claims, wherein the second nucleic acid encodes an FGE protein having an amino acid sequence identical to SEQ ID NO:5.

10. The cell of any one of the preceding claims, wherein the first nucleic acid comprises a sequence at least 70% identical to SEQ ID NO:7

11. The cell of any one of the preceding claims, wherein the first nucleic acid comprises a sequence of SEQ ID NO:7.

12. The cell of any one of the preceding claims, wherein the second nucleic acid comprises a sequence at least 70% identical to SEQ ID NO:8

13. The cell of any one of the preceding claims, wherein the second nucleic acid comprises a sequence identical to SEQ ID NO:8.

14. The cell of any one of the preceding claims, wherein both of the first and second nucleic acids are exogenous.

15. The cell of any one of the preceding claims, wherein the first and/or second nucleic acids are integrated in the genome of the cell.

16. The cell of any one of the preceding claims, wherein the first and/or second nucleic acids are present in one or more extra-chromosomal constructs.

17. The cell of any one of the preceding claims, wherein the cell is a mammalian cell.

18. The cell of claim 17, wherein the mammalian cell is a human cell.

19. The cell of claim 17, wherein the mammalian cell is a CHO cell.

20. The cell of any one of the preceding claims, wherein the cell is adaptable to suspension culture.

21. A method of producing recombinant iduronate-2-sulfatase (I2S) protein comprising cultivating a cell of any one of the preceding claims under conditions such that the recombinant I2S and FGE proteins are co-expressed in the cell.

22. The method of claim 21, wherein the cell is cultivated at a large scale.

23. The method of claim 22, wherein the large scale is a bioreactor process.

24. The method of claim 23, wherein the bioreactor process is a perfusion process.

25. The method of claim 23 or 24, wherein the bioreactor is at a scale selected from 10L, 200L, 500L, 1000L, 1500L, or 2000L.

26. The method of claim 22, wherein the large scale is a roller bottle process.

27. The method of any one of claims 21-26, wherein the cell is cultivated in a serum-free medium.

28. The method of any one of claims 21-26, wherein the cell is cultivated in a serum-containing medium.

29. The method of any one of claims 21-26, wherein the cell is cultivated in suspension.

30. The method of any one of claims 21-26, wherein the cell is cultivated adherent.

31. The method of any one of claims 21-30, wherein the method further comprises a step of purifying the recombinant I2S protein.

32. A recombinant iduronate-2-sulfatase (I2S) protein produced by a method of any one of claims 21-31.

33. A recombinant iduronate-2-sulfatase (I2S) protein produced by a cell of any one of claims 1-20.

SEQ ID NO: 1

1	Ser	Glu	Thr	Gln	Ala	Asn	Ser	Thr	Thr	Asp	Ala	Leu	Asn	Val	Leu	Leu	Ile	Ile	Val	Asp
21	Asp	Leu	Arg	Pro	Ser	Leu	Gly	Cys	Tyr	Gly	Asp	Lys	Leu	Val	Arg	Ser	Pro	Asn	Ile	Asp
41	Gln	Leu	Ala	Ser	His	Ser	Leu	Leu	Phe	Gln	Asn	Ala	Phe	Ala	Gln	Gln	Ala	Val	Cys	Ala
61	Pro	Ser	Arg	Bal	Ser	Phe	Leu	Thr	Gly	Arg	Arg	Pro	Asp	Thr	Thr	Arg	Leu	Tyr	Asp	Phe
81	Asn	Ser	Tyr	Trp	Arg	Val	His	Ala	Gly	Asn	Phe	Ser	Thr	Ile	Pro	Gln	Tyr	Phe	Lys	Glu
101	Asn	Gly	Tyr	Val	Thr	Met	Ser	Val	Gly	Lys	Val	Phe	His	Pro	Gly	Ile	Ser	Ser	Asn	His
121	Thr	Asp	Asp	Ser	Pro	Tyr	Ser	Trp	Ser	Phe	Pro	Pro	Tyr	His	Pro	Ser	Ser	Glu	Lys	Tyr
141	Glu	Asn	Thr	Lys	Thr	Cys	Arg	Gly	Pro	Asp	Gly	Glu	Leu	His	Ala	Asn	Leu	Leu	Cys	Pro
161	Val	Asp	Val	Leu	Asp	Val	Pro	Glu	Gly	Thr	Leu	Pro	Asp	Lys	Gln	Ser	Thr	Glu	Gln	Ala
181	Ile	Gln	Leu	Leu	Glu	Lys	Met	Lys	Thr	Ser	Ala	Ser	Pro	Phe	Phe	Leu	Ala	Val	Gly	Tyr
201	His	Lys	Pro	His	Ile	Pro	Phe	Arg	Tyr	Pro	Lys	Glu	Phe	Gln	Lys	Leu	Tyr	Pro	Leu	Glu
221	Asn	Ile	Thr	Leu	Ala	Pro	Asp	Pro	Glu	Val	Pro	Asp	Gly	Leu	Pro	Pro	Val	Ala	Tyr	Asn
241	Pro	Trp	Met	Asp	Ile	Arg	Gln	Arg	Glu	Asp	Val	Gln	Ala	Leu	Asn	Ile	Ser	Val	Pro	Tyr
261	Gly	Pro	Ile	Pro	Val	Asp	Phe	Gln	Arg	Lys	Ile	Arg	Gln	Ser	Tyr	Phe	Ala	Ser	Val	Ser

FIG. 1A

281 Tyr Leu Asp Thr Gln Val Gly Arg Leu Leu Ser Ala Leu Asp Asp Leu Gln Leu Ala Asn
301 Ser Thr Ile Ile Ala Phe Thr Ser Asp His Gly Trp Ala Leu Gly Glu His Gly Glu Trp
321 Ala Lys Tyr Ser Asn Phe Asp Val Ala Thr His Val Pro Leu Ile Phe Tyr Val Pro Gly
341 Arg Thr Ala Ser Leu Pro Glu Ala Gly Glu Lys Leu Phe Pro Tyr Leu Asp Pro Phe Asp
361 Ser Ala Ser Gln Leu Met Glu Pro Gly Arg Gln Ser Met Asp Leu Val Glu Leu Val Ser
381 Leu Phe Pro Thr Leu Ala Gly Leu Ala Gly Leu Gln Val Pro Pro Arg Cys Pro Val Pro
401 Ser Phe His Val Glu Leu Cys Arg Clu Gly Lys Asn Leu Leu Lys His Phe Arg Phe Arg
421 Asp Leu Glu Glu Asp Pro Tyr Leu Pro Gly Asn Pro Arg Glu Leu Ile Ala Tyr Ser Gln
441 Tyr Pro Arg Pro Ser Asp Ile Pro Gln Trp Asn Ser Asp Lys Pro Ser Leu Lys Asp Ile
461 Lys Ile Met Gly Tyr Ser Ile Arg Thr Ile Asp Tyr Arg Tyr Thr Val Trp Val Gly Phe
481 Asn Pro Asp Glu Phe Leu Ala Asn Phe Ser Asp Ile His Ala Gly Glu Leu Tyr Phe Val
501 Asp Ser Asp Pro Leu Gln Asp His Asn Met Tyr Asn Asp Ser Gln Gly Gly Asp Leu Phe
521 Gln Leu Leu Met Pro

Asn - marks sites of N-linked glycosylation

Cys - example site of cysteine conversion

FIG. 1B

FIG. 2A

I2S and SUMF1 co-expression options

Expression units on separate vectors (co-transfection or subsequent transfections)

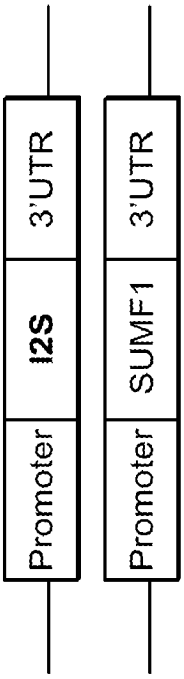
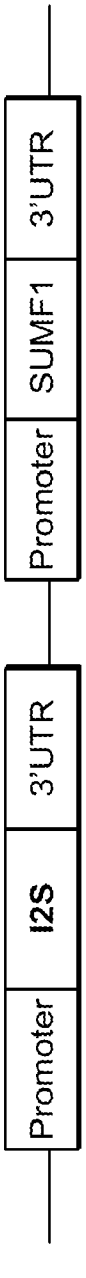


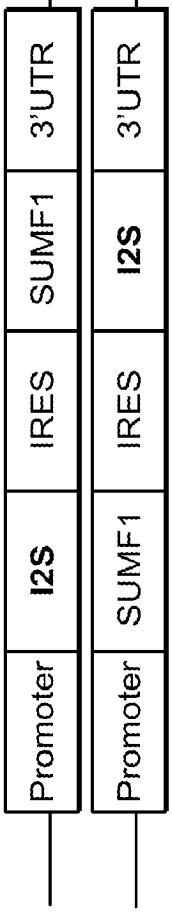
FIG. 2B

Expression units on the same vector (one transfection)

1) Separate cistrons



2) Transcriptionally linked cistrons



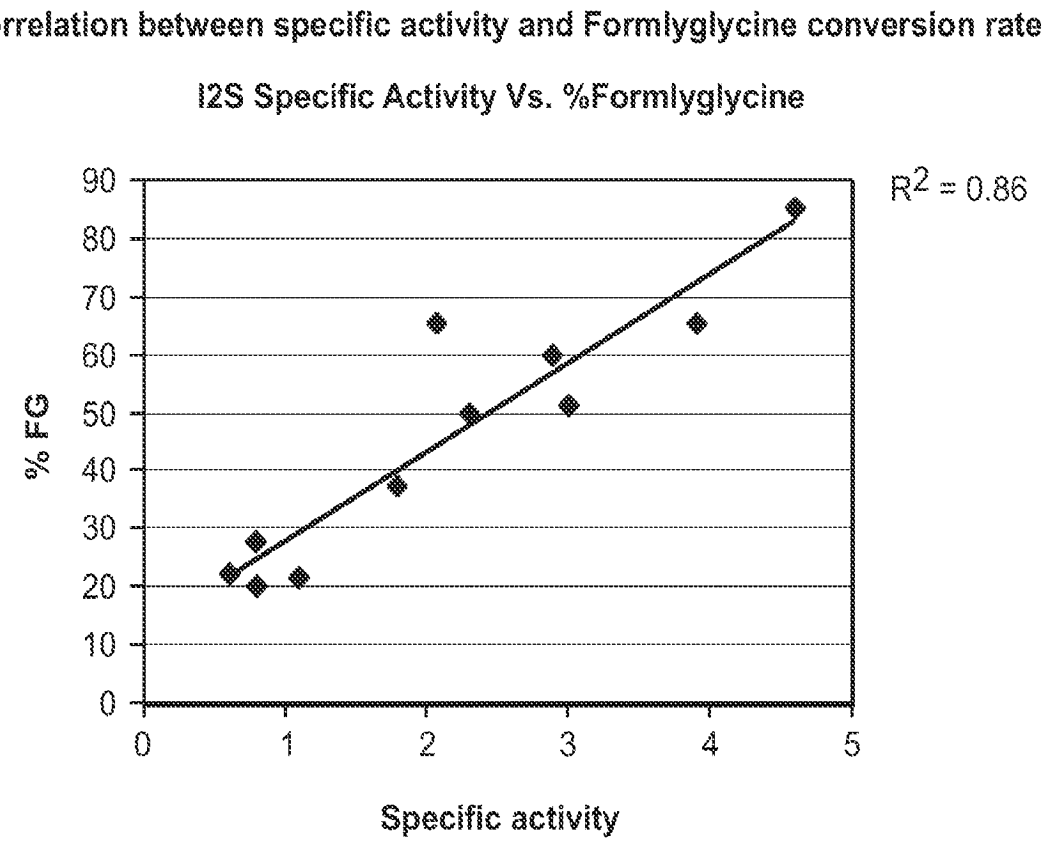

Correlation between specific activity and Formlyglycine conversion rate
I2S Specific Activity Vs. %Formlyglycine
R^2 = 0.86
% FG
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
Specific activity

FIG. 3

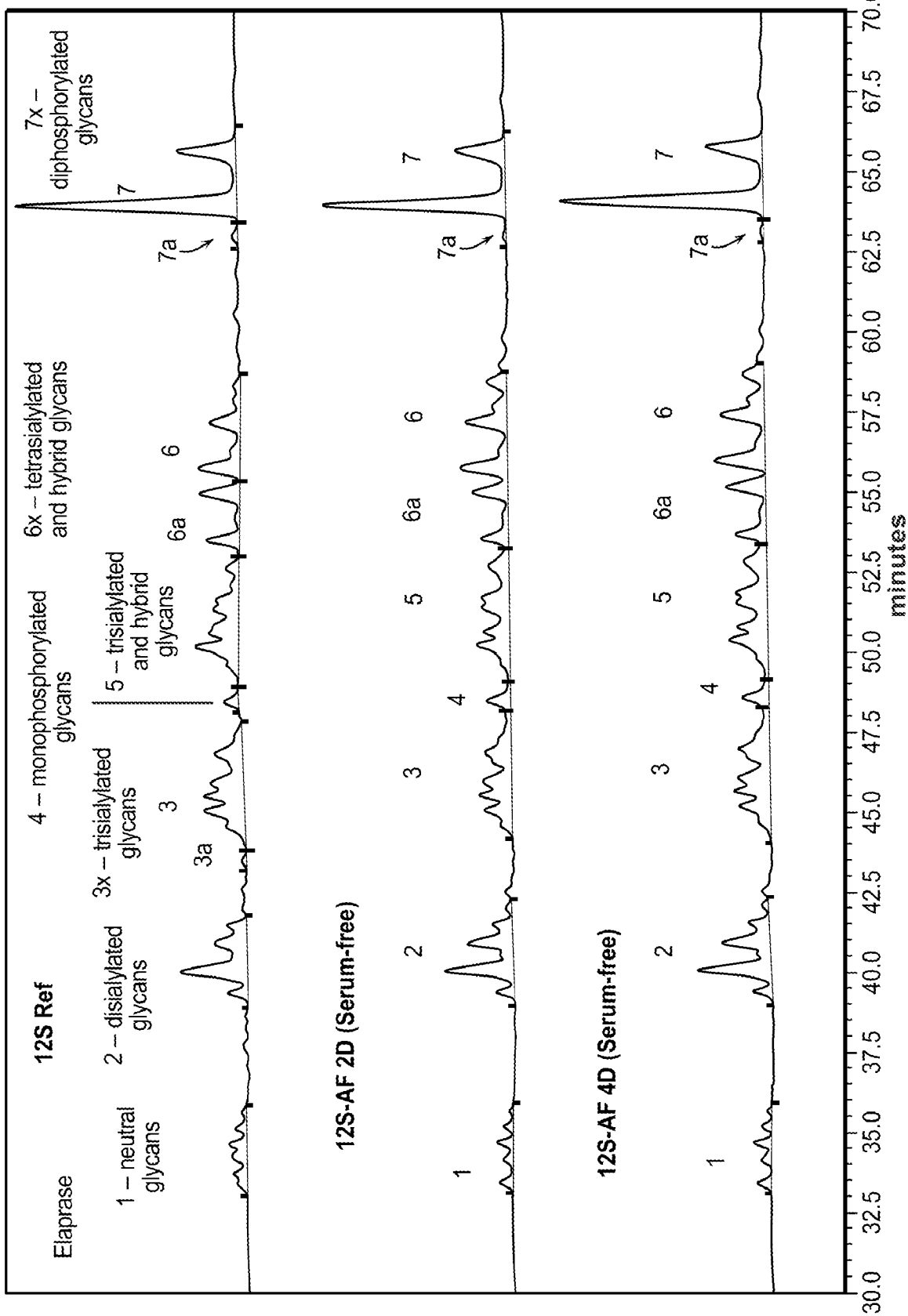

Elaprase
12S Ref
1 – neutral glycans
2 – disialylated glycans
3x – trisialylated glycans
4 – monophosphorylated glycans
5 – trisialylated and hybrid glycans
6x – tetrasialylated and hybrid glycans
7x – diphosphorylated glycans
1
2
3
3a
4
5
6a
6
7
7a
12S-AF 2D (Serum-free)
12S-AF 4D (Serum-free)
30.0
32.5
35.0
37.5
40.0
42.5
45.0
47.5
50.0
52.5
55.0
57.5
60.0
62.5
65.0
67.5
70.0
minutes

FIG. 4

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 January 2014 (03.01.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/005019 A3

(51) **International Patent Classification**:
A61K 38/46 (2006.01) *C12P 21/06* (2006.01)
C07K 1/00 (2006.01)

(21) **International Application Number**:
PCT/US2013/048571

(22) **International Filing Date**:
28 June 2013 (28.06.2013)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/666,719 29 June 2012 (29.06.2012) US

(71) **Applicant**: **SHIRE HUMAN GENETIC THERAPIES, INC.** [US/US]; 300 Shire Way, Lexington, Massachusetts 02421 (US).

(72) **Inventors**: **BOLGOG, Ferenc**; 300 Shire Way, Lexington, Massachusetts 02421 (US). **HEARTLEIN, Mike**; 167 Reed Farm Road, Boxborough, Massachusetts 01719 (US).

(74) **Agent**: **HUDDLESON, Justin P.**; Choate, hall & Stewart LLP, Two International Place, Boston, Massachusetts 02110 (US).

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

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

Published:
— *with international search report (Art. 21(3))*

[Continued on next page]

(54) **Title**: CELLS FOR PRODUCING RECOMBINANT IDURONATE-2-SULFATASE

(57) **Abstract**: The present invention provides, among other things, methods and compositions for production of recombinant I2S protein with improved potency and activity using cells co- express I2S and FGE protein. In some embodiments, cells according to the present invention are engineered to simultaneously over-express recombinant I2S and FGE proteins. Cells according to the invention are adaptable to various cell culture conditions. In some embodiments, cells of the present invention adaptable to a large-scale suspension serum- free culture.

FIG. 2A

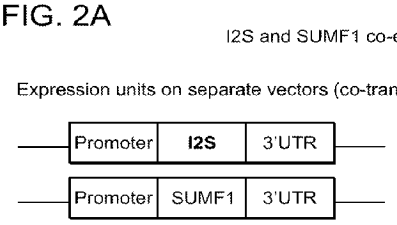
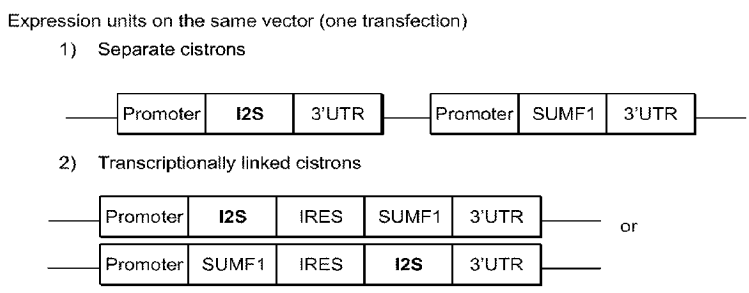


FIG. 2B



WO 2014/005019 A3

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

(88) Date of publication of the international search report:
27 February 2014

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US13/48571

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - A61K 38/46; C07K 1/00; C12P 21/06 (2013.01)
USPC - 435/68.1, 69.1, 325
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8): C12P 21/06; C07K 1/00, 16/00; A61L 33/00; A61K 38/46, 38/17; A61P 31/12; C12Q 1/68; C07H 21/04 (2013.01)
USPC: 435/68.1, 69.1, 325, 350, 41, 196, 320.1; 530/350, 387.1; 522/87; 436/6; 536/2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
MicroPatent (US-G, US-A, EP-A, EP-B, WO, JP-bib, DE-C,B, DE-A, DE-T, DE-U, GB-A, FR-A); Google/Google Scholar; Proquest; ScienceDirect; Search Terms Used: 'iduronate sulfatase', formylglyceine, enzyme, cysteine, recombinant, bioreactor, perfusion, 'roller bottle'

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 2004/072275 A2 (VON FIGURA, K et al.) August 26, 2004; page 4, lines 30-32; page 13, line 29; page 14, lines 3-6; page 54, lines 17-25; page 54, lines 30-33; page 66, lines 29-32; page 71, lines 30-35 to page 72, line 1; page 87, lines 26-30; page 87, lines 31-34 to page 88, lines 1-12; page 91, lines 16-21; Claims 2, 7, 9	1, 3/1 ----- 2, 3/2
Y	WO 2011/044542 A1 (PARDRIDGE, WM et al.) April 14, 2011; paragraphs [0003], [0099], [00161], [00183]; figure 4	1, 2, 3/1, 3/2
Y	WO 2005/113765 A2 (ZANKEL, T et al.) December 1, 2005; page 25, lines 5-11; page 25, lines 26-33 to page 26, lines 1-8	2, 3/2

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 November 2013 (27.11.2013)	12 DEC 2013
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US13/48571

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-33
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

SEQUENCE LISTING

<110> Boldog, Ferenc
Heartlein, Mike

<120> CELLS FOR PRODUCING RECOMBINANT IDURONATE-2-SULFATASE

<130> 2006685-0340

<150> 61/666,719
<151> 2012-06-29

<160> 8

<170> PatentIn version 3.5

<210> 1
<211> 525
<212> PRT
<213> Homo sapiens

<400> 1

Ser Glu Thr Gln Ala Asn Ser Thr Thr Asp Ala Leu Asn Val Leu Leu
1 5 10 15

Ile Ile Val Asp Asp Leu Arg Pro Ser Leu Gly Cys Tyr Gly Asp Lys
20 25 30

Leu Val Arg Ser Pro Asn Ile Asp Gln Leu Ala Ser His Ser Leu Leu
35 40 45

Phe Gln Asn Ala Phe Ala Gln Gln Ala Val Cys Ala Pro Ser Arg Val
50 55 60

Ser Phe Leu Thr Gly Arg Arg Pro Asp Thr Thr Arg Leu Tyr Asp Phe
65 70 75 80

Asn Ser Tyr Trp Arg Val His Ala Gly Asn Phe Ser Thr Ile Pro Gln
85 90 95

Tyr Phe Lys Glu Asn Gly Tyr Val Thr Met Ser Val Gly Lys Val Phe
100 105 110

His Pro Gly Ile Ser Ser Asn His Thr Asp Asp Ser Pro Tyr Ser Trp
115 120 125

Ser Phe Pro Pro Tyr His Pro Ser Ser Glu Lys Tyr Glu Asn Thr Lys
130 135 140

Thr Cys Arg Gly Pro Asp Gly Glu Leu His Ala Asn Leu Leu Cys Pro
145 150 155 160

Val Asp Val Leu Asp Val Pro Glu Gly Thr Leu Pro Asp Lys Gln Ser

165 170 175

Thr Glu Gln Ala Ile Gln Leu Leu Glu Lys Met Lys Thr Ser Ala Ser
180 185 190

Pro Phe Phe Leu Ala Val Gly Tyr His Lys Pro His Ile Pro Phe Arg
195 200 205

Tyr Pro Lys Glu Phe Gln Lys Leu Tyr Pro Leu Glu Asn Ile Thr Leu
210 215 220

Ala Pro Asp Pro Glu Val Pro Asp Gly Leu Pro Pro Val Ala Tyr Asn
225 230 235 240

Pro Trp Met Asp Ile Arg Gln Arg Glu Asp Val Gln Ala Leu Asn Ile
245 250 255

Ser Val Pro Tyr Gly Pro Ile Pro Val Asp Phe Gln Arg Lys Ile Arg
260 265 270

Gln Ser Tyr Phe Ala Ser Val Ser Tyr Leu Asp Thr Gln Val Gly Arg
275 280 285

Leu Leu Ser Ala Leu Asp Asp Leu Gln Leu Ala Asn Ser Thr Ile Ile
290 295 300

Ala Phe Thr Ser Asp His Gly Trp Ala Leu Gly Glu His Gly Glu Trp
305 310 315 320

Ala Lys Tyr Ser Asn Phe Asp Val Ala Thr His Val Pro Leu Ile Phe
325 330 335

Tyr Val Pro Gly Arg Thr Ala Ser Leu Pro Glu Ala Gly Glu Lys Leu
340 345 350

Phe Pro Tyr Leu Asp Pro Phe Asp Ser Ala Ser Gln Leu Met Glu Pro
355 360 365

Gly Arg Gln Ser Met Asp Leu Val Glu Leu Val Ser Leu Phe Pro Thr
370 375 380

Leu Ala Gly Leu Ala Gly Leu Gln Val Pro Pro Arg Cys Pro Val Pro
385 390 395 400

Ser Phe His Val Glu Leu Cys Arg Glu Gly Lys Asn Leu Leu Lys His
405 410 415

Phe Arg Phe Arg Asp Leu Glu Glu Asp Pro Tyr Leu Pro Gly Asn Pro

420 425 430

Arg Glu Leu Ile Ala Tyr Ser Gln Tyr Pro Arg Pro Ser Asp Ile Pro
435 440 445

Gln Trp Asn Ser Asp Lys Pro Ser Leu Lys Asp Ile Lys Ile Met Gly
450 455 460

Tyr Ser Ile Arg Thr Ile Asp Tyr Arg Tyr Thr Val Trp Val Gly Phe
465 470 475 480

Asn Pro Asp Glu Phe Leu Ala Asn Phe Ser Asp Ile His Ala Gly Glu
485 490 495

Leu Tyr Phe Val Asp Ser Asp Pro Leu Gln Asp His Asn Met Tyr Asn
500 505 510

Asp Ser Gln Gly Gly Asp Leu Phe Gln Leu Leu Met Pro
515 520 525

<210> 2
<211> 550
<212> PRT
<213> Homo sapiens

<400> 2

Met Pro Pro Pro Arg Thr Gly Arg Gly Leu Leu Trp Leu Gly Leu Val
1 5 10 15

Leu Ser Ser Val Cys Val Ala Leu Gly Ser Glu Thr Gln Ala Asn Ser
20 25 30

Thr Thr Asp Ala Leu Asn Val Leu Leu Ile Ile Val Asp Asp Leu Arg
35 40 45

Pro Ser Leu Gly Cys Tyr Gly Asp Lys Leu Val Arg Ser Pro Asn Ile
50 55 60

Asp Gln Leu Ala Ser His Ser Leu Leu Phe Gln Asn Ala Phe Ala Gln
65 70 75 80

Gln Ala Val Cys Ala Pro Ser Arg Val Ser Phe Leu Thr Gly Arg Arg
85 90 95

Pro Asp Thr Thr Arg Leu Tyr Asp Phe Asn Ser Tyr Trp Arg Val His
100 105 110

Ala Gly Asn Phe Ser Thr Ile Pro Gln Tyr Phe Lys Glu Asn Gly Tyr
115 120 125

```
Val Thr Met Ser Val Gly Lys Val Phe His Pro Gly Ile Ser Ser Asn
    130                 135                 140

His Thr Asp Asp Ser Pro Tyr Ser Trp Ser Phe Pro Pro Tyr His Pro
145                 150                 155                 160

Ser Ser Glu Lys Tyr Glu Asn Thr Lys Thr Cys Arg Gly Pro Asp Gly
                165                 170                 175

Glu Leu His Ala Asn Leu Leu Cys Pro Val Asp Val Leu Asp Val Pro
            180                 185                 190

Glu Gly Thr Leu Pro Asp Lys Gln Ser Thr Glu Gln Ala Ile Gln Leu
        195                 200                 205

Leu Glu Lys Met Lys Thr Ser Ala Ser Pro Phe Phe Leu Ala Val Gly
    210                 215                 220

Tyr His Lys Pro His Ile Pro Phe Arg Tyr Pro Lys Glu Phe Gln Lys
225                 230                 235                 240

Leu Tyr Pro Leu Glu Asn Ile Thr Leu Ala Pro Asp Pro Glu Val Pro
                245                 250                 255

Asp Gly Leu Pro Pro Val Ala Tyr Asn Pro Trp Met Asp Ile Arg Gln
            260                 265                 270

Arg Glu Asp Val Gln Ala Leu Asn Ile Ser Val Pro Tyr Gly Pro Ile
        275                 280                 285

Pro Val Asp Phe Gln Arg Lys Ile Arg Gln Ser Tyr Phe Ala Ser Val
    290                 295                 300

Ser Tyr Leu Asp Thr Gln Val Gly Arg Leu Leu Ser Ala Leu Asp Asp
305                 310                 315                 320

Leu Gln Leu Ala Asn Ser Thr Ile Ile Ala Phe Thr Ser Asp His Gly
                325                 330                 335

Trp Ala Leu Gly Glu His Gly Glu Trp Ala Lys Tyr Ser Asn Phe Asp
            340                 345                 350

Val Ala Thr His Val Pro Leu Ile Phe Tyr Val Pro Gly Arg Thr Ala
        355                 360                 365

Ser Leu Pro Glu Ala Gly Glu Lys Leu Phe Pro Tyr Leu Asp Pro Phe
    370                 375                 380
```

```
Asp Ser Ala Ser Gln Leu Met Glu Pro Gly Arg Gln Ser Met Asp Leu
385                 390                 395                 400

Val Glu Leu Val Ser Leu Phe Pro Thr Leu Ala Gly Leu Ala Gly Leu
                405                 410                 415

Gln Val Pro Pro Arg Cys Pro Val Pro Ser Phe His Val Glu Leu Cys
            420                 425                 430

Arg Glu Gly Lys Asn Leu Leu Lys His Phe Arg Phe Arg Asp Leu Glu
        435                 440                 445

Glu Asp Pro Tyr Leu Pro Gly Asn Pro Arg Glu Leu Ile Ala Tyr Ser
    450                 455                 460

Gln Tyr Pro Arg Pro Ser Asp Ile Pro Gln Trp Asn Ser Asp Lys Pro
465                 470                 475                 480

Ser Leu Lys Asp Ile Lys Ile Met Gly Tyr Ser Ile Arg Thr Ile Asp
                485                 490                 495

Tyr Arg Tyr Thr Val Trp Val Gly Phe Asn Pro Asp Glu Phe Leu Ala
            500                 505                 510

Asn Phe Ser Asp Ile His Ala Gly Glu Leu Tyr Phe Val Asp Ser Asp
        515                 520                 525

Pro Leu Gln Asp His Asn Met Tyr Asn Asp Ser Gln Gly Gly Asp Leu
    530                 535                 540

Phe Gln Leu Leu Met Pro
545                 550


<210>  3
<211>  312
<212>  PRT
<213>  Homo sapiens

<400>  3

Met Pro Pro Pro Arg Thr Gly Arg Gly Leu Leu Trp Leu Gly Leu Val
1               5                   10                  15

Leu Ser Ser Val Cys Val Ala Leu Gly Ser Glu Thr Gln Ala Asn Ser
            20                  25                  30

Thr Thr Asp Ala Leu Asn Val Leu Leu Ile Ile Val Asp Asp Leu Arg
        35                  40                  45
```

```
Pro Ser Leu Gly Cys Tyr Gly Asp Lys Leu Val Arg Ser Pro Asn Ile
    50                  55                  60

Asp Gln Leu Ala Ser His Ser Leu Leu Phe Gln Asn Ala Phe Ala Gln
65                  70                  75                  80

Gln Ala Val Cys Ala Pro Ser Arg Val Ser Phe Leu Thr Gly Arg Arg
                85                  90                  95

Pro Asp Thr Thr Arg Leu Tyr Asp Phe Asn Ser Tyr Trp Arg Val His
            100                 105                 110

Ala Gly Asn Phe Ser Thr Ile Pro Gln Tyr Phe Lys Glu Asn Gly Tyr
        115                 120                 125

Val Thr Met Ser Val Gly Lys Val Phe His Pro Gly Ile Ser Ser Asn
    130                 135                 140

His Thr Asp Asp Ser Pro Tyr Ser Trp Ser Phe Pro Pro Tyr His Pro
145                 150                 155                 160

Ser Ser Glu Lys Tyr Glu Asn Thr Lys Thr Cys Arg Gly Pro Asp Gly
                165                 170                 175

Glu Leu His Ala Asn Leu Leu Cys Pro Val Asp Val Leu Asp Val Pro
            180                 185                 190

Glu Gly Thr Leu Pro Asp Lys Gln Ser Thr Glu Gln Ala Ile Gln Leu
        195                 200                 205

Leu Glu Lys Met Lys Thr Ser Ala Ser Pro Phe Phe Leu Ala Val Gly
    210                 215                 220

Tyr His Lys Pro His Ile Pro Phe Arg Tyr Pro Lys Glu Phe Gln Lys
225                 230                 235                 240

Leu Tyr Pro Leu Glu Asn Ile Thr Leu Ala Pro Asp Pro Glu Val Pro
                245                 250                 255

Asp Gly Leu Pro Pro Val Ala Tyr Asn Pro Trp Met Asp Ile Arg Gln
            260                 265                 270

Arg Glu Asp Val Gln Ala Leu Asn Ile Ser Val Pro Tyr Gly Pro Ile
        275                 280                 285

Pro Val Asp Phe Gln Glu Asp Gln Ser Ser Thr Gly Phe Arg Leu Lys
    290                 295                 300
```

Thr Ser Ser Thr Arg Lys Tyr Lys
305 310

<210> 4
<211> 343
<212> PRT
<213> Homo sapiens

<400> 4

Met Pro Pro Pro Arg Thr Gly Arg Gly Leu Leu Trp Leu Gly Leu Val
1 5 10 15

Leu Ser Ser Val Cys Val Ala Leu Gly Ser Glu Thr Gln Ala Asn Ser
20 25 30

Thr Thr Asp Ala Leu Asn Val Leu Leu Ile Ile Val Asp Asp Leu Arg
35 40 45

Pro Ser Leu Gly Cys Tyr Gly Asp Lys Leu Val Arg Ser Pro Asn Ile
50 55 60

Asp Gln Leu Ala Ser His Ser Leu Leu Phe Gln Asn Ala Phe Ala Gln
65 70 75 80

Gln Ala Val Cys Ala Pro Ser Arg Val Ser Phe Leu Thr Gly Arg Arg
85 90 95

Pro Asp Thr Thr Arg Leu Tyr Asp Phe Asn Ser Tyr Trp Arg Val His
100 105 110

Ala Gly Asn Phe Ser Thr Ile Pro Gln Tyr Phe Lys Glu Asn Gly Tyr
115 120 125

Val Thr Met Ser Val Gly Lys Val Phe His Pro Gly Ile Ser Ser Asn
130 135 140

His Thr Asp Asp Ser Pro Tyr Ser Trp Ser Phe Pro Pro Tyr His Pro
145 150 155 160

Ser Ser Glu Lys Tyr Glu Asn Thr Lys Thr Cys Arg Gly Pro Asp Gly
165 170 175

Glu Leu His Ala Asn Leu Leu Cys Pro Val Asp Val Leu Asp Val Pro
180 185 190

Glu Gly Thr Leu Pro Asp Lys Gln Ser Thr Glu Gln Ala Ile Gln Leu
195 200 205

```
Leu Glu Lys Met Lys Thr Ser Ala Ser Pro Phe Phe Leu Ala Val Gly
    210                 215                 220

Tyr His Lys Pro His Ile Pro Phe Arg Tyr Pro Lys Glu Phe Gln Lys
225                 230                 235                 240

Leu Tyr Pro Leu Glu Asn Ile Thr Leu Ala Pro Asp Pro Glu Val Pro
                245                 250                 255

Asp Gly Leu Pro Pro Val Ala Tyr Asn Pro Trp Met Asp Ile Arg Gln
            260                 265                 270

Arg Glu Asp Val Gln Ala Leu Asn Ile Ser Val Pro Tyr Gly Pro Ile
        275                 280                 285

Pro Val Asp Phe Gln Arg Lys Ile Arg Gln Ser Tyr Phe Ala Ser Val
    290                 295                 300

Ser Tyr Leu Asp Thr Gln Val Gly Arg Leu Leu Ser Ala Leu Asp Asp
305                 310                 315                 320

Leu Gln Leu Ala Asn Ser Thr Ile Ile Ala Phe Thr Ser Asp His Gly
                325                 330                 335

Phe Leu Met Arg Thr Asn Thr
            340
```

<210> 5
<211> 341
<212> PRT
<213> Homo sapiens

<400> 5

```
Ser Gln Glu Ala Gly Thr Gly Ala Gly Ala Gly Ser Leu Ala Gly Ser
1               5                   10                  15

Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly Ala His Gly Ser Ser Ala
            20                  25                  30

Ala Ala His Arg Tyr Ser Arg Glu Ala Asn Ala Pro Gly Pro Val Pro
        35                  40                  45

Gly Glu Arg Gln Leu Ala His Ser Lys Met Val Pro Ile Pro Ala Gly
    50                  55                  60

Val Phe Thr Met Gly Thr Asp Asp Pro Gln Ile Lys Gln Asp Gly Glu
65                  70                  75                  80

Ala Pro Ala Arg Arg Val Thr Ile Asp Ala Phe Tyr Met Asp Ala Tyr
```

85 90 95

Glu Val Ser Asn Thr Glu Phe Glu Lys Phe Val Asn Ser Thr Gly Tyr
100 105 110

Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser Phe Val Phe Glu Gly Met
115 120 125

Leu Ser Glu Gln Val Lys Thr Asn Ile Gln Gln Ala Val Ala Ala Ala
130 135 140

Pro Trp Trp Leu Pro Val Lys Gly Ala Asn Trp Arg His Pro Glu Gly
145 150 155 160

Pro Asp Ser Thr Ile Leu His Arg Pro Asp His Pro Val Leu His Val
165 170 175

Ser Trp Asn Asp Ala Val Ala Tyr Cys Thr Trp Ala Gly Lys Arg Leu
180 185 190

Pro Thr Glu Ala Glu Trp Glu Tyr Ser Cys Arg Gly Gly Leu His Asn
195 200 205

Arg Leu Phe Pro Trp Gly Asn Lys Leu Gln Pro Lys Gly Gln His Tyr
210 215 220

Ala Asn Ile Trp Gln Gly Glu Phe Pro Val Thr Asn Thr Gly Glu Asp
225 230 235 240

Gly Phe Gln Gly Thr Ala Pro Val Asp Ala Phe Pro Pro Asn Gly Tyr
245 250 255

Gly Leu Tyr Asn Ile Val Gly Asn Ala Trp Glu Trp Thr Ser Asp Trp
260 265 270

Trp Thr Val His His Ser Val Glu Glu Thr Leu Asn Pro Lys Gly Pro
275 280 285

Pro Ser Gly Lys Asp Arg Val Lys Lys Gly Gly Ser Tyr Met Cys His
290 295 300

Arg Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala Ala Arg Ser Gln Asn Thr
305 310 315 320

Pro Asp Ser Ser Ala Ser Asn Leu Gly Phe Arg Cys Ala Ala Asp Arg
325 330 335

Leu Pro Thr Met Asp

340

<210> 6
<211> 374
<212> PRT
<213> Homo sapiens

<400> 6

Met Ala Ala Pro Ala Leu Gly Leu Val Cys Gly Arg Cys Pro Glu Leu
1 5 10 15

Gly Leu Val Leu Leu Leu Leu Leu Leu Ser Leu Leu Cys Gly Ala Ala
20 25 30

Gly Ser Gln Glu Ala Gly Thr Gly Ala Gly Ala Gly Ser Leu Ala Gly
35 40 45

Ser Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly Ala His Gly Ser Ser
50 55 60

Ala Ala Ala His Arg Tyr Ser Arg Glu Ala Asn Ala Pro Gly Pro Val
65 70 75 80

Pro Gly Glu Arg Gln Leu Ala His Ser Lys Met Val Pro Ile Pro Ala
85 90 95

Gly Val Phe Thr Met Gly Thr Asp Asp Pro Gln Ile Lys Gln Asp Gly
100 105 110

Glu Ala Pro Ala Arg Arg Val Thr Ile Asp Ala Phe Tyr Met Asp Ala
115 120 125

Tyr Glu Val Ser Asn Thr Glu Phe Glu Lys Phe Val Asn Ser Thr Gly
130 135 140

Tyr Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser Phe Val Phe Glu Gly
145 150 155 160

Met Leu Ser Glu Gln Val Lys Thr Asn Ile Gln Gln Ala Val Ala Ala
165 170 175

Ala Pro Trp Trp Leu Pro Val Lys Gly Ala Asn Trp Arg His Pro Glu
180 185 190

Gly Pro Asp Ser Thr Ile Leu His Arg Pro Asp His Pro Val Leu His
195 200 205

Val Ser Trp Asn Asp Ala Val Ala Tyr Cys Thr Trp Ala Gly Lys Arg
210 215 220

```
Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ser Cys Arg Gly Gly Leu His
225                 230                 235                 240

Asn Arg Leu Phe Pro Trp Gly Asn Lys Leu Gln Pro Lys Gly Gln His
                245                 250                 255

Tyr Ala Asn Ile Trp Gln Gly Glu Phe Pro Val Thr Asn Thr Gly Glu
            260                 265                 270

Asp Gly Phe Gln Gly Thr Ala Pro Val Asp Ala Phe Pro Pro Asn Gly
        275                 280                 285

Tyr Gly Leu Tyr Asn Ile Val Gly Asn Ala Trp Glu Trp Thr Ser Asp
    290                 295                 300

Trp Trp Thr Val His His Ser Val Glu Glu Thr Leu Asn Pro Lys Gly
305                 310                 315                 320

Pro Pro Ser Gly Lys Asp Arg Val Lys Lys Gly Gly Ser Tyr Met Cys
                325                 330                 335

His Arg Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala Ala Arg Ser Gln Asn
            340                 345                 350

Thr Pro Asp Ser Ser Ala Ser Asn Leu Gly Phe Arg Cys Ala Ala Asp
        355                 360                 365

Arg Leu Pro Thr Met Asp
    370
```

<210> 7
<211> 1653
<212> DNA
<213> Homo sapiens

<400> 7

```
atgcccccgc cccgcaccgg ccgcggcctg ctgtggctgg gcctggtgct gagcagcgtg       60

tgcgtggccc tgggcagcga gacccaggcc aacagcacca ccgacgccct gaacgtgctg      120

ctgatcatcg tggacgacct gcgccccagc ctgggctgct acggcgacaa gctggtgcgc      180

agccccaaca tcgaccagct ggccagccac agcctgctgt tccagaacgc cttcgcccag      240

caggccgtgt gcgcccccag ccgcgtgagc ttcctgaccg gccgccgccc cgacaccacc      300

cgcctgtacg acttcaacag ctactggcgc gtgcacgccg gcaacttcag caccatcccc      360

cagtacttca aggagaacgg ctacgtgacc atgagcgtgg gcaaggtgtt ccaccccggc      420

atcagcagca accacaccga cgacagcccc tacagctgga gcttcccccc ctaccacccc      480
```

agcagcgaga agtacgagaa caccaagacc tgccgcggcc ccgacggcga gctgcacgcc 540

aacctgctgt gccccgtgga cgtgctggac gtgcccgagg gcaccctgcc cgacaagcag 600

agcaccgagc aggccatcca gctgctggag aagatgaaga ccagcgccag ccccttcttc 660

ctggccgtgg gctaccacaa gccccacatc cccttccgct accccaagga gttccagaag 720

ctgtaccccc tggagaacat caccctggcc cccgaccccg aggtgcccga cggcctgccc 780

cccgtggcct acaaccccctg gatggacatc cgccagcgcg aggacgtgca ggccctgaac 840

atcagcgtgc cctacggccc catccccgtg gacttccagc gcaagatccg ccagagctac 900

ttcgccagcg tgagctacct ggacacccag gtgggccgcc tgctgagcgc cctggacgac 960

ctgcagctgg ccaacagcac catcatcgcc ttcaccagcg accacggctg ggccctgggc 1020

gagcacggcg agtgggccaa gtacagcaac ttcgacgtgg ccacccacgt gcccctgatc 1080

ttctacgtgc ccggccgcac cgccagcctg cccgaggccg gcgagaagct gttcccctac 1140

ctggaccccct tcgacagcgc cagccagctg atggagcccg gccgccagag catggacctg 1200

gtggagctgg tgagcctgtt ccccaccctg gccggcctgg ccggcctgca ggtgcccccc 1260

cgctgccccg tgcccagctt ccacgtggag ctgtgccgcg agggcaagaa cctgctgaag 1320

cacttccgct tccgcgacct ggaggaggac ccctacctgc ccggcaaccc ccgcgagctg 1380

atcgcctaca gccagtaccc ccgccccagc gacatccccc agtggaacag cgacaagccc 1440

agcctgaagg acatcaagat catgggctac agcatccgca ccatcgacta ccgctacacc 1500

gtgtgggtgg gcttcaaccc cgacgagttc ctggccaact tcagcgacat ccacgccggc 1560

gagctgtact tcgtggacag cgacccccctg caggaccaca acatgtacaa cgacagccag 1620

ggcggcgacc tgttccagct gctgatgccc tag 1653

<210> 8
<211> 1125
<212> DNA
<213> Homo sapiens

<400> 8
atggctgcgc ccgcactagg gctggtgtgt ggacgttgcc ctgagctggg tctcgtcctc 60

ttgctgctgc tgctctcgct gctgtgtgga gcggcaggga gccaggaggc cgggaccggt 120

gcgggcgcgg ggtcccttgc gggttcttgc ggctgcggca cgccccagcg gcctggcgcc 180

catggcagtt cggcagccgc tcaccgatac tcgcgggagg ctaacgctcc gggccccgta 240

cccggagagc ggcaactcgc gcactcaaag atggtcccca tccctgctgg agtatttaca 300

atgggcacag atgatcctca gataaagcag gatggggaag cacctgcgag gagagttact 360

attgatgcct tttacatgga tgcctatgaa gtcagtaata ctgaatttga gaagtttgtg 420

aactcaactg gctatttgac agaggctgag aagtttggcg actcctttgt ctttgaaggc 480

atgttgagtg agcaagtgaa gaccaatatt caacaggcag ttgcagctgc tccctggtgg 540

ttacctgtga aaggcgctaa ctggagacac ccagaagggc ctgactctac tattctgcac 600

aggccggatc atccagttct ccatgtgtcc tggaatgatg cggttgccta ctgcacttgg 660

gcagggaagc ggctgcccac ggaagctgag tgggaataca gctgtcgagg aggcctgcat 720

aatagacttt tcccctgggg caacaaactg cagcccaaag gccagcatta tgccaacatt 780

tggcagggcg agtttccggt gaccaacact ggtgaggatg gcttccaagg aactgcgcct 840

gttgatgcct tccctcccaa tggttatggc ttatacaaca tagtggggaa cgcatgggaa 900

tggacttcag actggtggac tgttcatcat tctgttgaag aaacgcttaa cccaaaaggt 960

ccccccttctg ggaaagaccg agtgaagaaa ggtggatcct acatgtgcca taggtcttat 1020

tgttacaggt atcgctgtgc tgctcggagc cagaacacac ctgatagctc tgcttcgaat 1080

ctgggattcc gctgtgcagc cgaccgcctg cccaccatgg actga 1125

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 104540517 A
(43)申请公布日 2015.04.22

(21)申请号 201380037937.X

(22)申请日 2013.06.28

(30)优先权数据
61/666,719 2012.06.29 US

(85)PCT国际申请进入国家阶段日
2015.01.16

(86)PCT国际申请的申请数据
PCT/US2013/048571 2013.06.28

(87)PCT国际申请的公布数据
WO2014/005019 EN 2014.01.03

(71)申请人 夏尔人类遗传性治疗公司
地址 美国马萨诸塞州

(72)发明人 F·博尔戈 M·哈特莱因

(74)专利代理机构 北京商专永信知识产权代理事务所(普通合伙) 11400
代理人 邬玥 葛强

(51)Int.Cl.
A61K 38/46(2006.01)
C07K 1/00(2006.01)
C12P 21/06(2006.01)

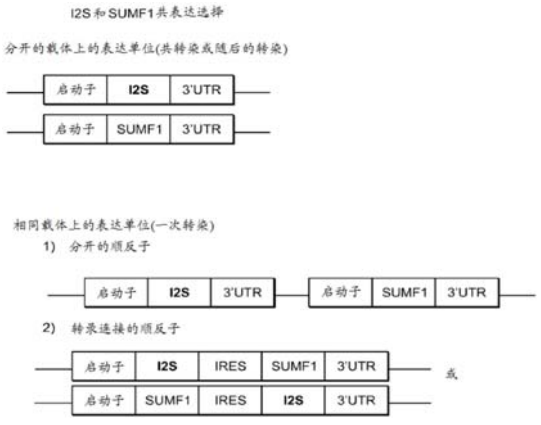
权利要求书2页 说明书30页
序列表15页 附图5页

(54)发明名称

用于产生重组艾杜糖-2-硫酸酯酶的细胞

(57)摘要

本发明除其它事项以外提供了使用共表达I2S和FGE蛋白的细胞产生具有改善的效能和活性的重组I2S蛋白的方法和组合物。在一些实施方案中,将根据本发明的细胞工程化以同时过表达重组I2S和FGE蛋白。根据本发明的细胞适应于各种细胞培养条件。在一些实施方案中,本发明的细胞适应于大规模悬浮无血清培养。



CN 104540517 A

1.一种细胞,其包含

第一核酸,其编码包含与 SEQ ID NO:1 具有至少 70％的同一性的氨基酸序列的艾杜糖 -2- 硫酸酯酶 (I2S) 蛋白 ;和

第二核酸,其编码包含与 SEQ ID NO:5 具有至少 70％的同一性的氨基酸序列的甲酰甘氨酸生成酶 (FGE) 蛋白,

其中所述第一和 / 或所述第二核酸是外源性的,并且其中一旦在细胞培养条件下培养后,所述细胞产生包含至少约 70％的对应于 SEQ ID NO:1 的 Cys59 的半胱氨酸残基至 Cα - 甲酰甘氨酸 (FGly) 的转化的 I2S 蛋白。

2.一种细胞,其包含

第一核酸,其编码包含与 SEQ ID NO:1 具有至少 70％的同一性的氨基酸序列的艾杜糖 -2- 硫酸酯酶 (I2S) 蛋白 ;和

第二核酸,其编码包含与 SEQ ID NO:5 具有至少 70％的同一性的氨基酸序列的甲酰甘氨酸生成酶 (FGE) 蛋白,

其中所述第一和 / 或所述第二核酸是外源性的,并且其中一旦在细胞培养条件下培养后,所述细胞产生包含至少约 60％的对应于 SEQ ID NO:1 的 Cys59 的半胱氨酸残基至 Cα - 甲酰甘氨酸 (FGly) 的转化的 I2S 蛋白并且比生产率为大于约 30 皮克 / 细胞 / 日。

3.根据权利要求 1 或 2 所述的细胞,其中一旦在细胞培养条件下培养后,所述细胞产生包含至少约 80％的对应于 SEQ ID NO:1 的 Cys59 的半胱氨酸残基至 Cα - 甲酰甘氨酸 (FGly) 的转化的 I2S 蛋白。

4.根据前述权利要求中任一项所述的细胞,其中一旦在细胞培养条件下培养后,所述细胞产生包含至少约 90％的对应于 SEQ ID NO:1 的 Cys59 的半胱氨酸残基至 Cα - 甲酰甘氨酸 (FGly) 的转化的 I2S 蛋白。

5.根据前述权利要求中任一项所述的细胞,其中一旦在细胞培养条件下培养后,所述细胞产生包含至少约 95％的对应于 SEQ ID NO:1 的 Cys59 的半胱氨酸残基至 Cα - 甲酰甘氨酸 (FGly) 的转化的 I2S 蛋白。

6.根据前述权利要求中任一项所述的细胞,其中一旦在细胞培养条件下培养后,所述细胞产生包含至少约 97％的对应于 SEQ ID NO:1 的 Cys59 的半胱氨酸残基至 Cα - 甲酰甘氨酸 (FGly) 的转化的 I2S 蛋白。

7.根据前述权利要求中任一项所述的细胞,其中所述第一和 / 或所述第二核酸可操作地连接至 hCMV 启动子。

8.根据前述权利要求中任一项所述的细胞,其中所述第一核酸编码具有与 SEQ ID NO:1 相同的氨基酸序列的 I2S 蛋白。

9.根据前述权利要求中任一项所述的细胞,其中所述第二核酸编码具有与 SEQ ID NO:5 相同的氨基酸序列的 FGE 蛋白。

10.根据前述权利要求中任一项所述的细胞,其中所述第一核酸包含与SEQ ID NO:7具有至少 70％的同一性的序列。

11.根据前述权利要求中任一项所述的细胞,其中所述第一核酸包含SEQ ID NO:7的序列。

12.根据前述权利要求中任一项所述的细胞,其中所述第二核酸包含与SEQ ID NO:8具

有至少 70％的同一性的序列。

13. 根据前述权利要求中任一项所述的细胞，其中所述第二核酸包含与 SEQ ID NO:8 相同的序列。

14. 根据前述权利要求中任一项所述的细胞，其中所述第一和第二核酸均为外源性的。

15. 根据前述权利要求中任一项所述的细胞，其中所述第一和 / 或第二核酸被整合在所述细胞的基因组中。

16. 根据前述权利要求中任一项所述的细胞，其中所述第一和 / 或第二核酸存在于一个或多个染色体外构建体中。

17. 根据前述权利要求中任一项所述的细胞，其中所述细胞是哺乳动物细胞。

18. 根据权利要求 17 所述的细胞，其中所述哺乳动物细胞是人细胞。

19. 根据权利要求 17 所述的细胞，其中所述哺乳动物细胞是 CHO 细胞。

20. 根据前述权利要求中任一项所述的细胞，其中所述细胞适应于悬浮培养。

21. 一种产生重组艾杜糖 -2- 硫酸酯酶 (I2S) 蛋白的方法，其包括在使所述重组 I2S 和 FGE 蛋白在前述权利要求中任一项所述的细胞中共表达的条件下培养所述细胞。

22. 根据权利要求 21 所述的方法，其中大规模培养所述细胞。

23. 根据权利要求 22 所述的方法，其中所述大规模是生物反应器工艺。

24. 根据权利要求 23 所述的方法，其中所述生物反应器工艺是灌注工艺。

25. 根据权利要求 23 或 24 所述的方法，其中所述生物反应器为选自 10L、200L、500L、1000L、1500L 或 2000L 的规模。

26. 根据权利要求 22 所述的方法，其中所述大规模是滚瓶工艺。

27. 根据权利要求 21-26 中任一项所述的方法，其中在无血清培养基中培养所述细胞。

28. 根据权利要求 21-26 中任一项所述的方法，其中在含血清培养基中培养所述细胞。

29. 根据权利要求 21-26 中任一项所述的方法，其中悬浮培养所述细胞。

30. 根据权利要求 21-26 中任一项所述的方法，其中贴壁培养所述细胞。

31. 根据权利要求 21-30 中任一项所述的方法，其中所述方法还包括纯化所述重组 I2S 蛋白的步骤。

32. 一种重组艾杜糖 -2- 硫酸酯酶 (I2S) 蛋白，其通过根据权利要求 21-31 中任一项所述的方法产生。

33. 一种重组艾杜糖 -2- 硫酸酯酶 (I2S) 蛋白，其由根据权利要求 1-20 中任一项所述的细胞产生。

用于产生重组艾杜糖 -2- 硫酸酯酶的细胞

[0001] 相关申请的交叉引用

[0002] 本申请要求 2012 年 6 月 29 日提交的美国临时专利申请系列号 61/666,719 的优先权 ;其完整内容在此通过引用并入。

[0003] 序列表

[0004] 在本说明书提及在 2013 年 6 月 27 日以电子形式提交的命名为“2006685-00340_SEQ_LIST”的 ASCII.txt 文件的序列表。所述 .txt 文件于 2013 年 6 月 25 日产生,并且大小为 25KB。将序列表的全部内容通过引用并入本文。

[0005] 背景

[0006] 粘多糖贮积症 II 型 (MPS II,亨特综合症) 是因酶艾杜糖 -2- 硫酸酯酶 (I2S) 的缺乏而引起的 X 染色体连锁隐性溶酶体贮积症。I2S 从葡糖氨基聚糖 (GAG) 硫酸皮肤素和硫酸乙酰肝素切割末端 2-O- 硫酸酯部分。由于亨特综合症患者中丧失 I2S 酶或存在有缺陷的 I2S 酶,因此 GAG 在多种细胞类型的溶酶体中逐渐累积,从而导致细胞肿胀,器官肿大,组织破坏和器官系统功能障碍。

[0007] 一般而言,具有亨特综合症的人的身体表现包括躯体和神经症状。例如,在亨特综合症的一些病例中,中枢神经系统受累导致发育迟缓和神经系统的问题。而亨特综合症的非神经症状在出生时一般不存在,随着时间的推移,GAG 在体内细胞中的渐进累积可能对身体的外周组织具有巨大影响。GAG 在外周组织中的累积导致患者面部特征的鲜明粗糙度,并且造成前额突出、扁平桥和巨舌,此为亨特综合征患者的最典型特征。类似地,GAG 的累积可不利地影响身体的器官系统。最初表现为心脏、肺和呼吸道的壁的增厚,以及肝、脾和肾的异常肿大,这些深刻的变化最终可导致广泛的灾难性器官衰竭。因此,亨特综合症总是严重的、进行性的和限制寿命的。

[0008] 酶替代疗法 (ERT) 是用于治疗亨特综合症 (MPS II) 的批准的疗法,其包括向亨特综合症患者施用外源替代 I2S 酶。

[0009] 发明概述

[0010] 本发明除其他事项外提供了用于产生重组 I2S 蛋白的改进的方法和组合物,所述重组 I2S 蛋白允许对于亨特综合症的更有效的酶替代疗法。本发明涵盖这样的发现 :可由经工程化以共表达重组 I2S 蛋白和甲酰甘氨酸生成酶 (FGE) 的哺乳动物细胞产生更具效能的重组 I2S 蛋白。出乎意料的是,由这种工程化细胞产生的重组 I2S 蛋白具有异常高水平的 Cα - 甲酰甘氨酸 (FGly) 转化百分比 (例如,大于 70％和高达 100％),从而导致重组 I2S 蛋白的显著增强的酶活性。此外,已使根据本发明的共表达 I2S 和 FGE 蛋白的哺乳动物成功地适应在大规模悬浮培养中生长。因此,本发明使得能够更高效地大规模生产高效能重组 I2S 蛋白。

[0011] 因此,在一个方面,本发明提供了包含第一核酸和第二核酸的细胞,所述第一核酸编码具有与 SEQ ID NO:1 具有至少约 50％ (例如,至少约 55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％或 99％) 的同一性的氨基酸序列的艾杜糖 -2- 硫酸酯酶 (I2S) 蛋白,所述第二核酸编码包含与 SEQ ID NO:5 具有至少约 50％ (例如,至少约

55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％或 99％）的同一性的氨基酸序列的甲酰甘氨酸生成酶 (FGE) 蛋白，其中第一和 / 或第二核酸是外源的，并且其中一旦在细胞培养条件（例如，悬浮或贴壁培养）下培养后，细胞产生包含至少约 70％（例如，至少约 75％、80％、85％、90％、95％、96％、97％、98％、99％或 100％）的对应于 SEQ ID NO:1 的 Cys59 的半胱氨酸残基至 Cα - 甲酰甘氨酸 (FGly) 的转化的 I2S 蛋白。

[0012] 在另一个方面，本发明提供了包含第一核酸和第二核酸的细胞，所述第一核酸编码具有与 SEQ ID NO:1 具有至少约 50％（例如，至少约 55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％或 99％）的同一性的氨基酸序列的艾杜糖 -2- 硫酸酯酶 (I2S)，所述第二核酸编码包含与 SEQ ID NO:5 具有至少约 50％（例如，至少约 55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％或 99％）的同一性的氨基酸序列的甲酰甘氨酸产生酶 (FGE) 蛋白，其中第一和 / 或第二核酸是外源性的，并且其中一旦在细胞培养条件下培养后，细胞产生包含至少约 50％（例如，至少约 55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％或 100％）的对应于 SEQ ID NO:1 的 Cys59 的半胱氨酸残基至 Cα - 甲酰甘氨酸 (FGly) 的转化的 I2S 蛋白，并且比生产率为大于约 10 皮克 / 细胞 / 日（例如，大于约 15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95 或 100 皮克 / 细胞 / 日）。

[0013] 在一些实施方案中，第一核酸编码具有与 SEQ ID NO:1 相同的氨基酸序列的 I2S 蛋白。在一些实施方案中，第二核酸编码具有与 SEQ ID NO:5 相同的氨基酸序列的 FGE 蛋白。

[0014] 在一些实施方案中，第一和 / 或第二核酸可操作地连接至 hCMV 启动子。

[0015] 在一些实施方案中，对第一和 / 或第二核酸进行密码子优化。在一些实施方案中，第一核酸具有与 SEQ ID NO:7 具有至少约 50％（例如，至少约 55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％）的同一性的序列。在具体实施方案中，第一核酸具有 SEQ ID NO:7 的序列。

[0016] 在一些实施方案中，第二核酸包含与 SEQ ID NO:8 具有至少约 50％（例如，至少约 55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％）的同一性的序列。在一些实施方案中，第二核酸具有与 SEQ ID NO:8 相同的序列。

[0017] 在一些实施方案中，第一和第二核酸均为外源性的（也称为重组的）。在一些实施方案中，将第一和 / 或第二核酸整合（例如，稳定地）在细胞的基因组中。在一些实施方案中，第一和 / 或第二核酸存在于一个或多个染色体外构建体中。

[0018] 在一些实施方案中，本发明的细胞为哺乳动物细胞。在某些实施方案中，适合的哺乳动物细胞为人细胞。在某些实施方案中，适合的哺乳动物细胞为 CHO 细胞。

[0019] 在一些实施方案中，根据本发明的细胞适应于悬浮培养。在其它实施方案中，根据本发明的细胞是贴壁的。

[0020] 在其它方面，本发明提供了通过在使得重组 I2S 和 FGE 蛋白在细胞中共表达的条件下，培养本文中不同实施方案中描述的细胞来产生重组艾杜糖 -2- 硫酸酯酶 (I2S) 蛋白的方法。在一些实施方案中，大规模培养细胞。在一些实施方案中，适用于本发明的大规模为生物反应器工艺。在一些实施方案中，适用于本发明的生物反应器为选自 10L、200L、500L、1000L、1500L、2000L 的规模。在一些实施方案中，适用于本发明的大规模（例如，生物

反应器）工艺包括灌注工艺。在一些实施方案中，适用于本发明的大规模（例如，生物反应器）工艺包括分批培养。在一些实施方案中，适用于本发明的大规模工艺为滚瓶工艺。在一些实施方案中，悬浮培养根据本发明的细胞。在其它实施方案中，贴壁培养根据本发明的细胞。

[0021] 在一些实施方案中，在无血清培养基（例如，无动物、化学成分确定的或无蛋白质的培养基）中培养根据本发明的细胞。在其它实施方案中，在含血清培养基中培养根据本发明的细胞。

[0022] 在各个实施方案中，根据本发明的方法还包括纯化重组 I2S 蛋白的步骤。

[0023] 在另一个方面，本发明提供了由本文中各个实施方案中描述的细胞或方法产生的重组艾杜糖 -2- 硫酸酯酶 (I2S) 蛋白。

[0024] 在一些实施方案中，本发明提供了重组艾杜糖 -2- 硫酸酯酶 (I2S) 蛋白的制剂，其中所述重组 I2S 蛋白具有与 SEQ ID NO:1 具有至少约 50％（例如，至少约 55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％）的同一性的氨基酸序列；并且包含至少约 70％（例如，至少约 75％、80％、85％、90％、95％、96％、97％、98％、99％、100％）的对应于 SEQ ID NO:1 的 Cys59 的半胱氨酸残基至 Cα - 甲酰甘氨酸 (FGly) 的转化。在一些实施方案中，重组 I2S 蛋白具有与 SEQ ID NO:1 相同的氨基酸序列。在一些实施方案中，重组 I2S 蛋白具有至少约 20U/mg、30U/mg、40U/mg、50U/mg、60U/mg、70U/mg、80U/mg、90U/mg 或 100U/mg 的比活性，如使用肝素二糖作为底物通过体外硫酸酯释放活性测定所测定的。

[0025] 除其它事项以外，本发明还提供了一种包含本文中各个实施方案中描述的重组 I2S 蛋白和药学上可接受的载体的药物组合物，以及通过向需要治疗的受试者施用本文中描述的重组 I2S 蛋白或包含所述重组 I2S 蛋白的药物组合物来治疗亨特综合症的方法。

[0026] 如本文中所用，除非另有明确所指，否则术语“I2S 蛋白”、“I2S”、“I2S 酶”或语法等同物是指重组 I2S 蛋白分子的制剂。

[0027] 如本申请中所用，术语“约”和“大致”可作为等同物使用。具有或不具有约 / 大致的用于本申请的任何数字旨在涵盖由相关领域的普通技术人员所理解的任何正常波动。

[0028] 本发明的其他特征、目的和优势在下面的详细描述中显而易见。然而，应当理解，所述详细描述，虽然表示本发明的实施方案，但仅通过举例说明的方式给出，而非限制性的。根据所述详细描述，本发明的范围内的各种变化和修改对于本领域技术人员来说是显而易见的。

[0029] 附图简述

[0030] 共同地构成附图的以下描述的图仅用于说明目的而不是为了限制。

[0031] 图 1 描述了编码人艾杜糖 -2- 硫酸酯酶 (I2S) 蛋白的成熟形式的氨基酸序列 (SEQ ID NO:1)，并且指示了蛋白质序列内用于 N 连接糖基化和半胱氨酸转化的潜在位点。

[0032] 图 2 描述了用于共表达 I2S 和 FGE 的示例性构建体（即，SUMF1）设计。(A) 分开的载体上的表达单位（用于共转染或随后的转染）；(B) 相同载体上的表达单位（一次转染）：(1) 分隔的顺反子和 (2) 转录连接的顺反子。

[0033] 图 3 描述了观察到的与甲酰甘氨酸转化百分比相关的 I2S 比活性的示例性水平。

[0034] 图 4 描述了相较于参考重组 I2S 酶，针对使用在无血清细胞培养条件下生长的

I2S-AF 2D 和 4D 细胞系产生的重组 I2S 酶产生的示例性聚糖特征谱。

[0035] 定义

[0036] 为了使本发明更容易被理解，首先定义某些术语。对下列术语和其他术语的附加定义陈述于整个说明书中。

[0037] 氨基酸：如本文中所用，术语“氨基酸”在其最广泛意义上是指可并入多肽链中的任何化合物和/或物质。在一些实施方案中，氨基酸具有一般结构 H2N-C(H)(R)-COOH。在一些实施方案中，氨基酸是天然存在的氨基酸。在一些实施方案中，氨基酸是合成氨基酸；在一些实施方案中，氨基酸是 D-氨基酸；在一些实施方案中，氨基酸是 L-氨基酸。“标准氨基酸”是指通常在天然存在的肽中发现的 20 种标准 L-氨基酸中的任一种。“非标准氨基酸”是指除标准氨基酸外的任何氨基酸，无论其是合成制备的还是从天然来源获得的。如本文中所用，“合成氨基酸”涵盖化学修饰氨基酸，包括但不限于盐，氨基酸衍生物（如酰胺）和/或取代。氨基酸，包括在肽中的羧基和/或氨基末端氨基酸，可通过甲基化、酰胺化、乙酰化、保护基团和/或利用其他化学基团的取代来进行修饰，所述其他化学基团可改变肽的循环半衰期而不会不利地影响它们的活性。氨基酸可参与二硫键。氨基酸可以包含一种或多种翻译后修饰，例如与一种或多种化学实体（例如，甲基、乙酸酯基、乙酰基、磷酸酯基、甲酰基部分、类异戊二烯基、硫酸酯基、聚乙二醇部分、脂质部分、碳水化合物部分、生物素部分等的缔合。在一些实施方案中，可在用于细胞培养的补充培养基中提供本发明的氨基酸，或可将所述氨基酸用于所述补充培养基。在一些实施方案中，提供于或用于补充细胞培养基中的氨基酸可以以盐或水合物的形式提供。

[0038] 大致：如本文中所用，术语“大致”或“约”，如用于一个或多个目标值，意指与所述参照值相似的值。在某些实施方案中，除非另外指出或另外地根据上下文显而易见的（除其中这样的数值超过可能值的 100％外），否则术语“大致”或“约”是指在任一方向（大于或小于）上落在所述参照值的 25％、20％、19％、18％、17％、16％、15％、14％、13％、12％、11％、10％、9％、8％、7％、6％、5％、4％、3％、2％、1％或更少内的值的范围。

[0039] 分批培养：如本文中所用，术语“分批培养”是指培养细胞的方法，在所述方法中在培养过程开始时提供最终将用于培养细胞的所有组分，包括培养基（参见下文中“培养基”的定义）以及细胞本身。通常在某一点停止分批培养，随后收获培养基中细胞和/或组分，并且任选地进行纯化。

[0040] 生物利用度：如本文中所用，术语“生物利用度”通常是指到达受试者的血流的施用剂量的百分比。

[0041] 生物活性：如本文中所用，短语“生物活性”是指在生物系统（例如，细胞培养物、生物体等）中具有活性的任何物质的特征。例如，当向生物体施用时对该生物体具有生物作用的物质被认为是有生物活性的。生物活性还可通过体外测定（例如，体外酶促测定如硫酸酯释放测定法）来确定。在具体实施方案中，当蛋白质或多肽具有生物学活性时，共有蛋白质或多肽的至少一种生物活性的该蛋白质或多肽的一部分通常称为“生物活性”部分。在一些实施方案中，从细胞培养系统产生和/或纯化蛋白质，当向受试者施用时，所述蛋白质显示生物活性。在一些实施方案中，蛋白质需要进一步加工以变得具有生物活性。在一些实施方案中，蛋白质需要翻译后修饰，例如但不限于，糖基化（例如，唾液酸化(sialyation))、法尼基化(farnysylation)、切割、折叠、甲酰甘氨酸转化及其组合，以变得

具有生物活性。在一些实施方案中，作为前体 (proform) 形式（即未成熟的形式）产生的蛋白质可能需要额外的修饰来变得具有生物活性。

[0042] 生物反应器：如本文所用，术语“生物反应器”是指用于宿主细胞培养物生长的容器。生物反应器可具有任何尺寸，只要它对于哺乳动物细胞的培养是有用的。通常情况下，生物反应器将至少为 1 升，并且可以是 10、100、250、500、1000、2500、5000、8000、10,000、12,0000 升或以上或其间的任何容积。生物反应器的内部条件，包括但不限于 pH 值、克分子渗透压浓度、CO_2饱和度、O_2饱和度、温度及其组合，在培养期间通常受到控制。生物反应器可由适合于在本发明的培养条件下在培养基中保持细胞的任何材料（包括玻璃、塑料或金属）组成。在一些实施方案中，生物反应器可用于进行动物细胞培养。在一些实施方案中，生物反应器可用于进行哺乳动物细胞培养。在一些实施方案中，可将生物反应器与来源于这样的生物体的细胞和 / 或细胞系（例如但不限于哺乳动物细胞、昆虫细胞、细菌细胞、酵母细胞和人细胞）一起使用。在一些实施方案中，生物反应器可用于大规模细胞培养生产，并且通常为至少 100 升，并且可以为 200、500、1000、2500、5000、8000、10,000、12,0000 升或以上或其间的任何容。本领域普通技术人员将意识到，并且将能够选择适合的生物反应器用于实施本发明。

[0043] 细胞培养物：如本文中所用，这些术语是指在适合于细胞群体存活和 / 或生长的条件下在培养基中生长的细胞群。如对于本领域普通技术人员将很清楚的，如本文中使用的这些术语可指包含细胞群和所述群于其中生长的培养基的组合。

[0044] 培养：如本文中所用，术语“培养”或语法等同物是指在有利于生长或存活的条件下维持细胞的过程。术语“培养”和“细胞培养”或任何同义词在本申请中可互换使用。

[0045] 培养皿：如本文中所用，术语“培养皿”是指可为培养细胞提供无菌环境的任何容器。示例性培养皿包括但不限于玻璃、塑料或金属容器。

[0046] 酶替代疗法 (ERT)：如本文中所用，术语“酶替代疗法 (ERT)”指的是通过提供缺失的酶校正酶缺乏的任何治疗策略。在一些实施方案中，缺失的酶通过鞘内施用提供。在一些实施方案中，缺失的酶通过向血流内的输注提供。一旦施用，酶被细胞吸收并转运至溶酶体，在溶酶体中酶用于消除因酶缺乏而在溶酶体中积累的物质。通常情况下，为了使溶酶体酶替代疗法有效，将治疗性酶递送至其中表现贮积缺陷的靶组织中适合的细胞中的溶酶体中。

[0047] 表达：如本文中所用，核酸序列的“表达”是指下列事件的一个或多个：(1)RNA 模板从 DNA 序列的产生（例如，通过转录）；(2)RNA 转录物的加工（例如，通过剪接、编辑、5’帽形成和 / 或 3’末端形成）；(3)RNA 至多肽或蛋白质的翻译；和 / 或 (4) 多肽或蛋白质的翻译后修饰。

[0048] 补料分批培养：如本文所用，术语“补料分批培养”是指培养细胞的方法，在所述方法中，在培养过程开始后的某个时间给培养物提供额外的组分。所提供的组分通常包括已在培养过程中被耗尽的细胞的营养补充剂。通常在某个点停止补料分批培养，并且收获培养基中的细胞和 / 或组分和任选地纯化其。

[0049] 片段：如本文中所用，术语“片段”是指多肽，并且被定义为给定的多肽的任何分离的部分，所述部分对于该多肽是独特的，或具有该多肽的特征。如本文中所用，该术语还指给定的多肽的保留了全长多肽的至少一部分活性的任何分离的部分。优选地，保留的活性

的部分为全长多肽的活性的至少 10％。更优选，保留的活性的部分为全长多肽的活性的至少 20％、30％、40％、50％、60％、70％、80％或 90％。更优选，保留的活性的部分为全长多肽的活性的至少 95％、96％、97％、98％或 99％。最优选，保留的活性的部分为全长多肽的活性的 100％。如本文中所用，该术语还指包含至少在全长多肽中发现的已建立的序列元件的给定的多肽的任何部分。优选地，序列元件跨越全长多肽的至少 4-5，更优选至少约 10、15、20、25、30、35、40、45、50 或更多个氨基酸。

[0050] 基因：如本文中所用，术语“基因”是指任何核苷酸序列 DNA 或 RNA，所述核苷酸序列的至少某一部分编码分开的终产物，通常地（但不限于）多肽，所述产物在细胞过程的某一方面起作用。该术语并不意味仅指编码多肽或其它分离的终产物的编码序列，而且还可包括调节基础表达水平的编码序列之前和之后的区域，以及各个编码区段（“外显子”）之间的间插序列（“内含子”）。在一些实施方案中，基因可包括调节序列（例如，启动子、增强子、聚腺苷酸化序列、终止序列、kozac 序列、tata 盒等）和 / 或修饰序列。在一些实施方案中，基因可包括对不编码蛋白质但相反地编码功能性 RNA 分子例如 tRNA 基因、RNAi 诱导剂等的核酸的提及。

[0051] 基因产物或表达产物：如本文中所用，术语“基因产物”或“表达产物”一般是指从基因转录的 RNA（加工前和 / 或加工后）或由从基因转录的 RNA 编码的多肽（修饰前和 / 或修饰后）。

[0052] 基因控制元件：如本文中所用，术语“基因控制元件”是指调节与其可操作地连接的基因的表达的任何序列元件。基因控制元件可通过升高或降低表达水平来起作用，并且可位于编码序列之前、之内或之后。基因控制元件可通过调节例如转录的起始、延伸或终止、mRNA 剪接、mRNA 编辑、mRNA 稳定性、细胞内 mRNA 的定位、翻译的起始、延伸或终止而在基因表达的任何阶段起作用或在基因表达的任何其它阶段起作用。基因控制元件可单独地或彼此组合地起作用。

[0053] 同源性：如本文中所用，术语“同源性”是指聚合分子之间，例如核酸分子（例如，DNA 分子和 / 或 RNA 分子）之间和 / 或多肽分子之间的总体相关性。在一些实施方案中，如果聚合分子的序列具有至少 25％、30％、35％、40％、45％、50％、55％、60％、65％、70％、75％、80％、85％、90％、95％或 99％的同一性，则它们被认为是彼此“同源的”。在一些实施方案中，如果聚合分子的序列具有至少 25％、30％、35％、40％、45％、50％、55％、60％、65％、70％、75％、80％、85％、90％、95％或 99％的相似性，则它们被认为是彼此“同源的”。

[0054] 同一性：如本文中所用，“同一性”是指聚合分子之间，例如核酸分子（例如，DNA 分子和 / 或 RNA 分子）之间和 / 或多肽分子之间的总体相关性。两个核酸序列的百分比同一性的计算，例如，可通过为了最佳比较目的而比对两个序列（例如，可在第一和第二核酸序列的一个或两个中引入缺口来进行最佳比对，并且为了比较目的可忽略不相同的序列）来进行。在某些实施方案中，为了比较目的而比对的序列的长度为参照序列的长度的至少 30％，至少 40％，至少 50％，至少 60％，至少 70％，至少 80％，至少 90％，至少 95％或大体上 100％。随后比较对应核苷酸位置上的核苷酸。当第一序列中的位置被与第二序列中的对应位置相同的核苷酸占据时，则分子在该位置上是相同的。通过考虑两个序列的最佳比对而需要引入的缺口的数目和每个缺口的长度，两个序列之间的同一性百分比是序列共有的相同位置的数目的函数。序列的比较和两个序列之间的同一性百分比的测定可使用数学算

法来实现。例如，例如，两个核苷酸序列之间的同一性百分比可使用 Meyers 和 Miller 的算法 (CABIOS, 1989, 4:11-17)，使用 PAM120 权重残基表、缺口长度罚分 12 和缺口罚分 4 来测定，所述算法已被整合至 ALIGN 程序 (2.0 版）中。或者，可使用 GCG 软件包中的 GAP 程序，使用 NWSgapdna.CMP 矩阵来测定两个核苷酸序列之间的同一性百分比。各种其它序列比对程序是可用的，并且可被用来测定序列同一性，例如，Clustal。

[0055] 提高、增加或减少：如本文中所用，术语“提高”、“增加”或“减少”或语法等同物，指示相对于基线测量，例如在本文中描述的治疗开始之前相同个体中的测量或在本文中描述的治疗不存在的情况下对照个体（或多个对照个体）中的测量的值。“对照个体”是患有与正在治疗的个体相同的溶酶体贮积病形式的个体，所述个体与正在治疗的个体年龄大致相同（以确保被治疗个体与对照个体的疾病分期是可比较的）。

[0056] 鞘内施用：如本文中所用，术语“鞘内施用”或“鞘内注射”是指至椎管内的注射（围绕脊髓的鞘内空间）。可使用各种技术，包括但不限于，通过钻孔的侧脑室注射或池穿刺或腰椎穿刺等。在一些实施方案中，根据本发明的“鞘内施用”或“鞘内递送”是指通过腰部区域或腰部区的 IT 施用或递送，即腰 IT 施用或递送。如本文中所用，术语“腰部区”或“腰部区域”指的是第三和第四腰（下背部）椎骨之间的区域，更包含地，脊椎的 L2-S1 区。

[0057] 分离的：如本文中所用，术语“分离的”是指这样的物质和 / 或实体，所述物质和 / 或实体 (1) 已与当最初产生（无论是在自然界中还是在实验环境中）时与其结合的组分的至少一些分离，和 / 或 (2) 已通过人的手产生、制备和 / 或制造。分离的物质和 / 或实体可与约 10％、约 20％、约 30％、约 40％、约 50％、约 60％、约 70％、约 80％、约 90％、约 91％、约 92％、约 93％、约 94％、约 95％、约 96％、约 97％、约 98％、约 99％或更多的与它们最初结合的其它组分分离。在一些实施方案中，分离物质具有约 80％、约 85％、约 90％、约 91％、约 92％、约 93％、约 94％、约 95％、约 96％、约 97％、约 98％、约 99％或高于约 99％的纯度。如本文中所用，如果物质基本上不含其它组分的话，则它是“纯的”。如本文中所用，分离的物质和 / 或实体的百分比纯度的计算不应包括赋形剂（例如，缓冲液、溶剂、水等）。

[0058] 培养基：如本文中所用，该术语指包含有素滋养生长细胞的营养物的溶液。通常情况下，这些溶液提供了细胞进行最低生长和 / 或存活所需的必需和非必需氨基酸、维生素、能量来源、脂质和微量元素。该溶液还可包含增强高于最小速率的生长和 / 或存活的组分，包括激素和生长因子。在一些实施方案中，将培养基配制为对于细胞存活和增殖是最佳的 pH 和盐浓度。在一些实施方案中，培养基可以是“化学成分确定的培养基”—不含蛋白质、水解产物或未知组合物的组分的无血清培养基。在一些实施方案中，化学成分确定的培养基不含动物来源的组分并且培养基中的所有组分具有已知的化学结构。在一些实施方案中，培养基可以是“基于血清的培养基”—已补充了动物来源的组分例如但不限于胎牛血清、马血清、山羊血清、驴血清和 / 或其组合的培养基。

[0059] 核酸：如本文中所用，术语“核酸”以其最广泛的意义是指为寡核苷酸链或可被并入寡核苷酸链的化合物和 / 或物质。在一些实施方案中，核酸是作为寡核苷酸链或可通过磷酸二酯键并入寡核苷酸链的化合物和 / 或物质。在一些实施方案中，“核酸”是指个别核酸残基（例如，核苷酸和 / 或核苷）。在一些实施方案中，“核酸”是指包含个别核酸残基的寡核苷酸链。如本文所使用的，术语“寡核苷酸”和“多核苷酸”可以互换使用。在一些实施方案中，“核酸”涵盖 RNA 以及单链和 / 或双链 DNA 和 / 或 cDNA。此外，术语“核酸”、“DNA”、

“RNA”和 / 或类似术语包括核酸类似物，即，具有除磷酸二酯骨架外的类似物。例如，所谓的“肽核酸”，其在本领域是已知的并且在骨架中具有肽键而非磷酸二酯键，被认为在本发明的范围之内。术语“编码氨基酸序列的核苷酸序列”包括互为简并形式和 / 或编码相同氨基酸序列的所有核苷酸序列。编码蛋白质和 / 或 RNA 的核苷酸序列可包括内含子。核酸可以从天然来源中纯化，使用重组表达系统产生并任选地纯化，化学合成等。在适当情况下，例如，在化学合成分子的情况下，核酸可包含核苷类似物，例如具有经化学修饰的碱基或糖、骨架修饰等的类似物。除非另有说明，否则核酸序列以 5’至 3’方向显示。术语“核酸区段”在本文中用于指作为更长的核酸序列的一部分的核酸序列。在许多实施例中，核酸区段包含至少 3、4、5、6、7、8、9、10 或更多个残基。在一些实施方案中，核酸是或包含天然核苷（例如，腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧鸟苷和脱氧胞苷）；核苷类似物（例如，2- 氨基腺苷、2- 硫代胸苷、肌苷、吡咯并嘧啶、3- 甲基腺苷、5- 甲基胞苷、C-5 丙炔基 - 胞苷、C-5 丙炔基 - 尿苷、2- 氨基腺苷、C5- 溴尿苷、C5- 氟尿苷、C5- 碘尿苷、C5- 丙炔基 - 尿苷、C5- 丙炔基 - 胞苷、C5- 甲基胞苷、2- 氨基腺苷、7- 脱氮腺苷、7- 脱氮鸟苷、8- 氧代腺苷、8- 氧代鸟苷、O(6)- 甲基鸟嘌呤和 2- 硫代胞苷）；化学修饰碱基；生物修饰的碱基（例如，甲基化碱基）；嵌入碱基；修饰糖（例如，2’- 氟代核糖、核糖、2’- 脱氧核糖、阿拉伯糖和己糖）；和 / 或修饰磷酸基团（例如，硫代磷酸酯和 5’-N- 亚磷酰胺键）。在一些实施方案中，本发明特别针对“未修饰核酸”，意指未经历化学修饰以促进或实现递送的核酸（例如，多核苷酸和残基，包括核苷酸和 / 或核苷）。

[0060] 灌注工艺：如本文中所用，术语“灌注工艺”是指其中在培养过程开始之后将另外的组分连续地或半连续地提供给培养物的培养细胞的方法。所提供的组分通常包括已在培养过程中被耗尽的细胞的营养补充剂。通常在连续或半连续的基础上收获培养基中的一部分细胞和 / 或组分并且任选地对其进行纯化。通常情况下，涉及灌注工艺的细胞培养过程被称为“灌注培养”。通常情况下，在灌注工艺过程中在新鲜培养基中提供营养补充剂。在一些实施方案中，新鲜培养基可与细胞培养过程中使用的基本培养基相同或相似。在一些实施方案中，新鲜的培养基可与基本培养基不同，但包含期望的营养补充剂。在一些实施方案中，新鲜培养基是化学成分确定的培养基。

[0061] 蛋白质：如本文中所用，术语“蛋白质”是指多肽（即，至少 2 个通过肽键彼此连接的氨基酸的串）。蛋白质可包括除氨基酸外的部分（例如，可以是糖蛋白、蛋白聚糖等）和 / 或可另外地被加工或修饰。本领域普通技术人员将理解，“蛋白质”可以是如由细胞产生的完整的多肽链（具有或不具有信号序列），或可为其特征部分。在一些实施方案中，蛋白质有时可包括 1 个以上的多肽链，例如，通过一个或多个二硫键连接或通过其它方式缔合。在一些实施方案中，多肽可含有 L- 氨基酸、D- 氨基酸或两者，并且可以包含多种本领域已知的氨基酸修饰或类似物中的任一种。有用的修饰包括，例如，末端乙酰化、酰胺化、甲基化等。在一些实施方案中，蛋白质可包含天然氨基酸、非天然氨基酸、合成氨基酸及其组合。术语“肽”通常用于指具有长度少于约 100 个氨基酸、少于约 50 个氨基酸、少于 20 个氨基酸或少于 10 个氨基酸的肽。在一些实施方案中，蛋白质是抗体、抗体片段、其生物活性部分和 / 或其特征部分。

[0062] 重组蛋白和重组多肽：如本文中所用，这些术语是指从宿主细胞表达的多肽，所述宿主细胞已经被基因工程化以表达该多肽。在一些实施方案中，重组蛋白可在来源于动物

的宿主细胞中表达。在一些实施方案中，重组蛋白可在来源于昆虫的宿主细胞中表达。在一些实施方案中，重组蛋白可在来源于酵母的宿主细胞中表达。在一些实施方案中，重组蛋白可在来源于原核生物的宿主细胞中表达。在一些实施方案中，重组蛋白可在来源于哺乳动物的宿主细胞中表达。在一些实施方案中，重组蛋白可在来源于人的宿主细胞中表达。在一些实施方案中，所述重组表达的多肽可与在宿主细胞中正常表达的多肽相同或相似。在一些实施方案中，重组表达的多肽对于宿主细胞可以是外来的，即对于在宿主细胞中正常表达的肽是异源的。或者，在一些实施方案中，重组表达的多肽可以是嵌合的，因为多肽的部分包含与在宿主细胞中正常表达的多肽相同或相似的氨基酸序列，而其它部分对于宿主细胞是外来的。

[0063] 替代酶：如本文中所用，术语“替代酶”是指可用于至少部分地替代待治疗的疾病中的有缺陷或缺失的酶的任何酶。在一些实施方案中，术语“替代酶”是指可用于至少部分地替代待治疗的溶酶体贮积症中的有缺陷或缺失的溶酶体酶的任何酶。在一些实施方案中，替代酶能够减少哺乳动物溶酶体中积累的材料或者能够挽救或改善一种或多种溶酶体贮积病症状。适合于本发明的替代酶包括野生型溶酶体酶或经修饰的溶酶体酶并且可使用重组的和合成的方法来产生或从自然来源纯化而来。替代酶可以是重组的、合成的、基因活化的或天然的酶。

[0064] 载体：如本文中所用，“载体”指能够运输与其缔合的另一核酸的核酸分子。在一些实施例中，载体能够在宿主细胞例如真核和/或原核细胞中染色体外复制和/或表达与其连接的核酸。能够指导可操作地连接的基因的表达的载体在本文中称为“表达载体”。

[0065] 发明详述

[0066] 本发明除其它事项以外提供了使用共表达 I2S 和 FGE 蛋白的细胞产生具有提高的效能和活性的重组 I2S 蛋白的方法和组合物。在一些实施方案中，根据本发明的细胞经工程化同时过表达重组 I2S 和 FGE 蛋白。根据本发明的细胞适应于各种细胞培养条件。在一些实施方案中，本发明的细胞适应于大规模无血清悬浮培养。

[0067] 在以下小节中进一步详细地描述了本发明的各个方面。小节的使用并不意味着限制本发明。每一个小节可适用于本发明的任何方面。在本申请中，除非另有说明，否则“或”的使用表示“和/或”。

[0068] 艾杜糖 -2- 硫酸酯酶 (I2S)

[0069] 如本文中所用，I2S 蛋白是可取代天然存在的艾杜糖 -2- 硫酸酯酶 (I2S) 蛋白的至少部分活性或挽救一个或多个与 I2S 缺乏相关的表型或症状的任何蛋白质或蛋白质的部分。如本文中所用，术语“I2S 酶”和“I2S 蛋白”以及语法等同物可互换使用。

[0070] 通常情况下，人 I2S 蛋白作为前体形式产生。人 I2S 的前体形式包含信号肽（全长前体的氨基酸残基 1-25）、原肽（全长前体的氨基酸残基 26-33）和一条链（全长前体的残基 34-550），所述链可被进一步加工成 42kDa 的链（全长前体的残基的 34-455）和 14kDa 的链（全长前体的残基 446-550）。通常情况下，前体形式也称为全长前体或全长 I2S 蛋白，其含有 550 个氨基酸。已除去信号肽的成熟形式的氨基酸序列 (SEQ ID NO：1) 和典型野生型或天然存在的人 I2S 蛋白的全长前体的氨基酸序列 (SEQ ID NO:2) 示于表 1 中。信号肽加以下划线。此外，人 I2S 蛋白同种型 a 和 b 前体的氨基酸序列也分别提供于表 1 中的 SEQ ID NO：3 和 4。

[0071] 表 1. 人艾杜糖 -2- 硫酸酯酶

[0072]

成熟形式	SETQANSTTDALNVLLIIVDDLRPSLGCYGDKLVRSPNIDQLASHSLLFQNAFA QQAVCAPSRVSFLTGRRPDTTRLYDFNSYWRVHAGNFSTIPQYFKENGYVTMSV GKVFHPGISSNHTDDSPYSWSFPPYHPSSEKYENTKTCRGPDGELHANLLCPVD VLDVPEGTLPDKQSTEQAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPKEFQKL YPLENITLAPDPEVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVDFQRK IRQSYFASVSYLDTQVGRLLSALDDLQLANSTIIAFTSDHGWALGEHGEWAKYS NFDVATHVPLIFYVPGRTASLPEAGEKLFPYLDPFDSASQLMEPGRQSMDLVEL VSLFPTLAGLAGLQVPPRCPVPSFHVELCREGKNLLKHFRFRDLEEDPYLPGNP RELIAYSQYPRPSDIPQWNSDKPSLKDIKIMGYSIRTIDYRYTVWVGFNPDEFL ANFSDIHAGELYFVDSDPLQDHNMYNDSQGGDLFQLLMP(SEQ ID NO:1)
全长前体	MPPPRTGRGLLWLGLVLSSVCVALGSETQANSTTDALNVLLIIVDDLRPSLGCY GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYDFNSY

[0073]

(同种型a)	WRVHAGNFSTIPQYFKENGYVTMSVGKVFHPGISSNHTDDSPYSWSFPPYHPSS EKYENTKTCRGPDGELHANLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSA SPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDPEVPDGLPPVAYNPWMDI RQREDVQALNISVPYGPIPVDFQRKIRQSYFASVSYLDTQVGRLLSALDDLQLA NSTIIAFTSDHGWALGEHGEWAKYSNFDVATHVPLIFYVPGRTASLPEAGEKLF PYLDPFDSASQLMEPGRQSMDLVELVSLFPTLAGLAGLQVPPRCPVPSFHVELC REGKNLLKHFRFRDLEEDPYLPGNPRELIAYSQYPRPSDIPQWNSDKPSLKDIK IMGYSIRTIDYRYTVWVGFNPDEFLANFSDIHAGELYFVDSDPLQDHNMYNDSQ GGDLFQLLMP(SEQ ID NO:2)
同种型b前体	MPPPRTGRGLLWLGLVLSSVCVALGSETQANSTTDALNVLLIIVDDLRPSLGCY GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYDFNSY WRVHAGNFSTIPQYFKENGYVTMSVGKVFHPGISSNHTDDSPYSWSFPPYHPSS EKYENTKTCRGPDGELHANLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSA SPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDPEVPDGLPPVAYNPWMDI RQREDVQALNISVPYGPIPVDFQEDQSSTGFRLKTSSTRKYK (SEQ ID NO:3)
同种型c前体	MPPPRTGRGLLWLGLVLSSVCVALGSETQANSTTDALNVLLIIVDDLRPSLGCY GDKLVRSPNIDQLASHSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYDFNSY WRVHAGNFSTIPQYFKENGYVTMSVGKVFHPGISSNHTDDSPYSWSFPPYHPSS EKYENTKTCRGPDGELHANLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSA SPFFLAVGYHKPHIPFRYPKEFQKLYPLENITLAPDPEVPDGLPPVAYNPWMDI RQREDVQALNISVPYGPIPVDFQRKIRQSYFASVSYLDTQVGRLLSALDDLQLA NSTIIAFTSDHGFLMRTNT(SEQ ID No:4)

[0074] 因此,在一些实施方案中,I2S 酶为成熟人 I2S 蛋白 (SEQ ID NO:1)。如本文中公开的,SEQ ID NO:1 代表人 I2S 蛋白的典型氨基酸序列。在一些实施方案中,I2S 蛋白可以是因在 I2S 基因的 5’UTR 内的可选择起始位点上的转录产生的 SEQ ID NO:1 的剪接同种型和 / 或变体。在一些实施方案中,适合的替代酶可以是成熟人 I2S 蛋白的同源物或类似物。例如,成熟人 I2S 蛋白的同源物或类似物可以是相较于野生型或天然存在的 I2S 蛋白 (例如, SEQ ID NO:1) 包含一个或多个氨基酸取代、缺失和 / 或插入,同时保留大部分 I2S 蛋白活性的修饰成熟人 I2S 蛋白。因此,在一些实施方案中,适合于本发明的替代酶与成熟人 I2S 蛋白 (SEQ ID NO:1) 大体上同源。在一些实施方案中,适合于本发明的替代酶具有与 SEQ ID NO:1 具有至少 50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更大的同源性的氨基酸序列。在一些实施方案中,适合于本发明的替代酶与成熟人 I2S 蛋白 (SEQ ID NO:1) 大体上同一。在一些实施方案中,适合于本发明的替代酶具有与 SEQ ID NO:1 具有至少 50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更大同一性

的氨基酸序列。在一些实施方案中，适合于本发明的替代酶包含成熟人 I2S 蛋白的片段或部分。

[0075] 或者，I2S 酶是全长 I2S 蛋白。在一些实施方案中，I2S 酶可以是全长人 I2S 蛋白的同源物或类似物。例如，全长人 I2S 蛋白的同源物或类似物可以是相较于野生型或天然存在的全长 I2S 蛋白（例如，SEQ ID NO:2）包含一个或多个氨基酸取代、缺失和 / 或插入，同时保留大部分 I2S 蛋白活性的修饰全长人 I2S 蛋白。因此，在一些实施方案中，I2S 酶与全长人 I2S 蛋白（SEQ ID NO:2）大体上同源。在一些实施方案中，适合于本发明的 I2S 酶具有与 SEQ ID NO:2 具有至少 50％、55％、60％、65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更大同源性的氨基酸序列。在一些实施方案中，适合于本发明的 I2S 酶与 SEQ ID NO:2 大体上同一。在一些实施方案中，适合于本发明的 I2S 酶具有与 SEQ ID NO:2 具有至少 50％、55％、60％、65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更高同一性的氨基酸序列。在一些实施方案中，适合于本发明的 I2S 酶包含全长人 I2S 蛋白的片段或部分。如本文中所用，全长 I2S 蛋白通常包含信号肽序列。

[0076] 在一些实施方案中，适合于本发明的 I2S 酶为人 I2S 同种型 a 蛋白。在一些实施方案中，适合的 I2S 酶可以是人 I2S 同种型 a 蛋白的同源物或类似物。例如，人 I2S 同种型 a 蛋白的同源物或类似物可以是相较于野生型或天然存在的人 I2S 同种型 a 蛋白（例如，SEQ ID NO:3）包含一个或多个氨基酸取代、缺失和 / 或插入，同时保留大部分 I2S 蛋白活性的修饰人 I2S 同种型 a 蛋白。因此，在一些实施方案中，I2S 酶与人 I2S 同种型 a 蛋白（SEQ ID NO:3）大体上同源。在一些实施方案中，I2S 酶具有与 SEQ ID NO:3 具有至少 50％、55％、60％、65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更高同源性的氨基酸序列。在一些实施方案中，I2S 酶与 SEQ ID NO:3 大体上同一。在一些实施方案中，适用于本发明的 I2S 酶具有与 SEQ ID NO:3 具有至少 50％、55％、60％、65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更高同一性的氨基酸序列。在一些实施方案中，适用于本发明的 I2S 酶包含人 I2S 同种型 a 蛋白的片段或部分。如本文中所用，人 I2S 同种型 a 蛋白通常包含信号肽序列。

[0077] 在一些实施方案中，I2S 酶为人 I2S 同种型 b 蛋白。在一些实施方案中，I2S 酶可以是人 I2S 同种型 b 蛋白的同源物或类似物。例如，人 I2S 同种型 b 蛋白的同源物或类似物可以是相较于野生型或天然存在的人 I2S 同种型 b 蛋白（例如，SEQ ID NO:4）包含一个或多个氨基酸取代、缺失和 / 或插入，同时保留大部分 I2S 蛋白活性的修饰人 I2S 同种型 b 蛋白。因此，在一些实施方案中，I2S 酶与人 I2S 同种型 b 蛋白（SEQ ID NO:4）大体上同源。在一些实施方案中，I2S 酶具有与 SEQ ID NO:4 具有至少 50％、55％、60％、65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更高同源性的氨基酸序列。在一些实施方案中，I2S 酶与 SEQ ID NO:4 大体上同一。在一些实施方案中，I2S 酶具有与 SEQ ID NO:4 具有至少 50％、55％、60％、65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更高同一性的氨基酸序列。在一些实施方案中，适合于本发明的 I2S 酶包含人 I2S 同种型 b 蛋白的片段或部分。如本文中所用，人 I2S 同种型 b 蛋白通常包含信号肽序列。

[0078] 人 I2S 蛋白的同源物或类似物可按照对于本领域普通技术人员来说是已知的用

于改变多肽序列的方法，例如见于汇编此类方法的参考资料中的方法来制备。在一些实施方案中，氨基酸的保守取代包括在下列组内的氨基酸间的置换：(a)M、I、L、V；(b)F、Y、W；(c)K、R、H；(d)A、G；(e)S、T；(f)Q、N；和 (g)E、D。在一些实施方案中，“保守氨基酸取代”是指不改变在其中进行氨基酸取代的蛋白质的相关电荷或大小特征的氨基酸取代。

[0079] 在一些实施方案中，I2S 酶包含结合于细胞表面上的受体以促进细胞摄取和 / 或溶酶体靶向的部分。例如，这样的受体可以是不依赖于阳离子的甘露糖 -6- 磷酸受体 (CI-MPR)，其结合甘露糖 -6- 磷酸 (M6P) 残基。此外，CI-MPR 也结合其它蛋白质，包括 IGF- Ⅱ。适合的溶酶体靶向部分可以是 IGF-I、IGF-II、RAP、p97 及其变体、同源物或片段 (例如，包括那些具有与野生型成熟人 IGF-I、IGF-II、RAP、p97 肽序列具有至少 70％、75％、80％、85％、90％或 95％同一性的序列的肽)。在一些实施方案中，M6P 残基结合的受体可以是阳离子依赖性的。

[0080] 甲酰甘氨酸生成酶 (FGE)

[0081] 通常情况下，I2S 的酶活性受保守半胱氨酸 (例如，对应于成熟人 I2S(SEQ ID NO:1) 的氨基酸 59) 至甲酰甘氨酸的翻译后修饰影响，所述甲酰甘氨酸也被称为 2- 氨基 -3- 氧代丙酸或氧代 - 丙氨酸。这种翻译后修饰一般在蛋白质合成过程中发生于内质网中，并且由甲酰甘氨酸生成酶 (FGE) 催化。I2S 的酶比活性通常与 I2S 具有甲酰甘氨酸修饰所达到的程度正相关。例如，具有相对高的甲酰甘氨酸修饰量的 I2S 蛋白制剂通常具有相对高的酶比活性；然而具有相对低的甲酰甘氨酸修饰量的 I2S 蛋白制剂通常具有相对低的酶比活性。

[0082] 因此，适合于产生根据本发明的重组 I2S 蛋白的细胞通常还表达 FGE 蛋白。在一些实施方案中，适合的细胞表达内源性蛋白 FGE。在一些实施方案中，适合的细胞经工程化以表达与重组 I2S 组合的外源 (也被称为重组的) 甲酰甘氨酸生成酶 (FGE)。在一些实施方案中，适合的细胞经改造以激活内源 FGE 基因以便 FGE 蛋白的表达水平或活性升高。

[0083] 通常情况下，人 FGE 蛋白产生为前体形式。人 FGE 的前体形式包含信号肽 (全长前体的氨基酸残基 1-33) 和链 (全长前体的残基 34-374)。通常情况下，前体形式也称为为全长前体或全长 FGE 蛋白，其含有 374 个氨基酸。除去信号肽的成熟形式的氨基酸序列 (SEQ ID NO:5) 和典型野生型或天然存在的人 FGE 蛋白的全长前体的氨基酸序列 (SEQ ID NO:6) 示于表 2 中。

[0084] 表 2. 人甲酰甘氨酸生成酶 (FGE)

[0085]

成熟形式	SQEAGTGAGAGSLAGSCGCGTPQRPGAHGSSAAAHRYSREANAPGPVPGERQLA HSKMVPIPAGVFTMGTDDPQIKQDGEAPARRVTIDAFYMDAYEVSNTEFEKFVN STGYLTEAEKFGDSFVFEGMLSEQVKTNIQQAVAAAPWWLPVKGANWRHPEGPD STILHRPDHPVLHVSWNDAVAYCTWAGKRLPTEAEWEYSCRGGLHNRLFPWGNK LQPKGQHYANIWQGEFPVTNTGEDGFQGTAPVDAFPPNGYGLYNIVGNAWEWTS DWWTVHHSVEETLNPKGPPSGKDRVKKGGSYMCHRSYCYRYRCAARSQNTPDSS ASNLGFRCAADRLPTMD (SEQ ID NO:5)
全长前体	MAAPALGLVCGRCPELGLVLLLLLLSLLCGAAGSQEAGTGAGAGSLAGSCGCGT PQRPGAHGSSAAAHRYSREANAPGPVPGERQLAHSKMVPIPAGVFTMGTDDPQI KQDGEAPARRVTIDAFYMDAYEVSNTEFEKFVNSTGYLTEAEKFGDSFVFEGML SEQVKTNIQQAVAAAPWWLPVKGANWRHPEGPDSTILHRPDHPVLHVSWNDAVA YCTWAGKRLPTEAEWEYSCRGGLHNRLFPWGNKLQPKGQHYANIWQGEFPVTNT GEDGFQGTAPVDAFPPNGYGLYNIVGNAWEWTSDWWTVHHSVEETLNPKGPPSG KDRVKKGGSYMCHRSYCYRYRCAARSQNTPDSSASNLGFRCAADRLPTMD (SEQ ID NO:6)

[0086] 因此，在一些实施方案中，适合本发明的FGE酶为成熟人FGE蛋白(SEQ ID NO:5)。在一些实施方案中，适合的FGE酶可以是成熟人FGE蛋白的同源物或类似物。例如，成熟人FGE蛋白的同源物或类似物可以是相较于野生型或天然存在的FGE蛋白(例如，SEQ ID NO:5)包含一个或多个氨基酸取代、缺失和/或插入，同时保留大部分FGE蛋白活性的修饰成熟人FGE蛋白。因此，在一些实施方案中，适合于本发明的FGE酶与成熟人FGE蛋白(SEQ ID NO:5)大体上同源。在一些实施方案中，适合于本发明的FGE酶具有与SEQ ID NO:5具有至少50％、55％、60％、65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更高同源性的氨基酸序列。在一些实施方案中，适合于本发明的FGE酶与成熟人FGE蛋白(SEQ ID NO:5)大体上同一。在一些实施方案中，适合于本发明的FGE酶具有与SEQ ID NO:5具有至少50％、55％、60％、65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更高同一性的氨基酸序列。在一些实施方案中，适合于本发明的FGE酶包含成熟人FGE蛋白的片段或部分。

[0087] 或者，适合本发明的FGE酶为全长FGE蛋白。在一些实施方案中，FGE酶可以是全长人FGE蛋白的同源物或类似物。例如，全长人FGE蛋白的同源物或类似物可以是相较于野生型或天然存在的全长FGE蛋白(例如，SEQ ID NO:6)包含一个或多个氨基酸取代、缺失和/或插入，同时保留大部分FGE蛋白活性的修饰全长人FGE蛋白。因此，在一些实施方案中，适合于本发明的FGE酶与全长人FGE蛋白(SEQ ID NO:6)大体上同源。在一些实施方案中，适合于本发明的FGE酶具有与SEQ ID NO:4具有至少50％、55％、60％、65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更高同源性的氨基酸序列。在一些实施方案中，适合于本发明的FGE酶与SEQ ID NO:6大体上同一。在一些实施方案中，适合于本发明的FGE酶具有与SEQ ID NO:6具有至少50％、55％、60％、65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更高同一性的氨基酸序列。在一些实施方案中，适合于本发明的FGE酶包含全长人FGE蛋白的片段或部分。如本文中所用，全长FGE蛋白通常包含信号肽序列。

[0088] 编码示例性FGE蛋白的示例性核酸序列和氨基酸序列公开于US公开号20040229250中，其全部内容通过引用并入本文。

[0089] 共表达I2S和FGE的细胞

[0090] 本发明认识到对使用细胞培养系统进行生物活性I2S的高水平商业生产的需要。因为许多生产因素可影响特定宿主细胞的选择，所以本说明书中所公开的核酸分子针对广

泛的原核和真核细胞和 / 或细胞系，包括但不限于，源自细菌菌株、酵母菌株、昆虫细胞、动物细胞、哺乳动物细胞和人细胞的细胞系。本发明的方面还提供了表达构建体以及用于表达本说明书中公开的天然存在的以及修饰的 I2S 和 / 或 FGE 蛋白的重组稳定细胞系的产生。此外，本发明的方面还提供了用于使用本说明书的公开核酸序列产生表达 I2S 和 FGE 的细胞系的方法。

[0091] 编码 I2S 和 / 或 FGE 蛋白的核酸

[0092] 在一些实施方案中，提供了核酸分子，其包含编码本文中各个实施方案中描述的目标重组基因（在本文中称为转基因）例如 I2S 和 / 或 FGE 蛋白的核酸序列。在一些实施方案中，编码转基因的核酸可被修饰来提供编码的 I2S 和 / 或 FGE 蛋白的增加的表达，这也称为密码子优化。例如，编码转基因的核酸可通过改变编码序列的开放阅读框来进行修饰。如本文中所用，术语“开放阅读框”与“ORF”同义，并且意指潜在地能够编码蛋白质或蛋白质的部分的任何核苷酸序列。开放阅读框通常始于起始密码子（表示为，例如 AUG（对于 RNA 分子）和以标准代码表示的 DNA 分子中的 ATG），并且按密码子三联体阅读直至读框终于终止密码子（表示为，例如 UAA、UGA 或 UAG（对于 RNA 分子）和以标准代码表示的 DNA 分子中的 TAA、TGA 或 TAG）。如本文中所用，术语“密码子”是指核酸分子中 3 个核苷酸的序列，其在蛋白质合成过程中指定特定的氨基酸序列；也称为三联体或密码子三联体。例如，在标准遗传密码中的 64 个可能的密码子中，两个密码子 GAA 和 GAG 编码氨基酸谷氨酰胺，而密码子 AAA 和 AAG 指定氨基酸赖氨酸。在标准遗传密码中，3 个密码子是终止密码子，其不指定氨基酸。如本文中所用的，术语“同义密码子”是指编码单个氨基酸的密码子中的任一个和全部。除甲硫氨酸和色氨酸外，氨基酸由 2 至 6 个同义密码子编码。例如，在标准遗传密码中，编码氨基酸丙氨酸的 4 个同义密码子是 GCA、GCC、GCG 和 GCU，指定谷氨酰胺的两个同义密码子是 GAA 和 GAG，并且编码赖氨酸的两个同义密码子是 AAA 及 AAG。

[0093] 在一些实施方案中，编码的 I2S 和 / 或 FGE 蛋白的开放阅读框的核酸可使用标准密码子优化方法来进行修饰。用于密码子优化的各种商业算法是可用的，并且可用于实施本发明。通常情况下，密码子优化并不改变编码的氨基酸序列。在一些实施方案中，密码子优化可导致氨基酸改变如取代、缺失或插入。通常情况下，这样的氨基酸改变大体上不改变蛋白质活性。

[0094] 编码 I2S 和 FGE 蛋白的示例性核酸序列分别示于下面的 SEQ ID NO :7 和 8 中。

[0095] SEQ ID NO :7 编码艾杜糖 -2- 硫酸酯酶 (I2S) 的示例性核酸序列

[0096] ATGCCCCCGCCCCGCACCGGCCGCGGCCTGCTGTGGCTGGGCCTGGTGCTGAGCAGCGTGTGCGTGGCCCTGGGCAGCGAGACCCAGGCCAACAGCACCACCGACGCCCTGAACGTGCTGCTGATCATCGTGGACGACCTGCGCCCCAGCCTGGGCTGCTACGGCGACAAGCTGGTGCGCAGCCCCAACATCGACCAGCTGGCCAGCCACAGCCTGCTGTTCCAGAACGCCTTCGCCCAGCAGGCCGTGTGCGCCCCCAGCCGCGTGAGCTTCCTGACCGGCCGCCGCCCCGACACCACCCGCCTGTACGACTTCAACAGCTACTGGCGCGTGCACGCCGGCAACTTCAGCACCATCCCCCAGTACTTCAAGGAGAACGGCTACGTGACCATGAGCGTGGGCAAGGTGTTCCACCCCGGCATCAGCAGCAACCACACCGACGACAGCCCCTACAGCTGGAGCTTCCCCCCCTACCACCCCAGCAGCGAGAAGTACGAGAACACCAAGACCTGCCGCGGCCCCGACGGCGAGCTGCACGCCAACCTGCTGTGCCCCGTGGACGTGCTGGACGTGCCCGAGGGCACCCTGCCCGACAAGCAGAGCACCGAGCAGGCCATCCAGCTGCTGGAGAAGATGAAGACCAGCGCCAGCCCCTTCTTCCTGGCCGTGGGCTACCACAAGCCCCACATCCCCTTCCGCTACCCCAAGGAGTTCCAGAAGCTGTACCCCCTGGAGAACATCACCCTGGCCCCCGACCCCGAG

GTGCCCGACGGCCTGCCCCCCGTGGCCTACAACCCCTGGATGGACATCCGCCAGCGCGAGGACGTGCAGGCCCTGAA
CATCAGCGTGCCCTACGGCCCCATCCCCGTGGACTTCCAGCGCAAGATCCGCCAGAGCTACTTCGCCAGCGTGAGCT
ACCTGGACACCCAGGTGGGCCGCCTGCTGAGCGCCCTGGACGACCTGCAGCTGGCCAACAGCACCATCATCGCCTTC
ACCAGCGACCACGGCTGGGCCCTGGGCGAGCACGGCGAGTGGGCCAAGTACAGCAACTTCGACGTGGCCACCCACGT
GCCCCTGATCTTCTACGTCCCCGGCCGCACCGCCAGCCTGCCCGAGGCCGGCGAGAAGCTGTTCCCCTACCTGGACC
CCTTCGACAGCGCCAGCCAGCTGATGGAGCCCGGCCGCCAGAGCATGGACCTGGTGGAGCTGGTGAGCCTGTTCCCC
ACCCTGGCCGGCCTGGCCGGCCTGCAGGTGCCCCCCCGCTGCCCCGTGCCCAGCTTCCACGTGGAGCTGTGCCGCGA
GGGCAAGAACCTGCTGAAGCACTTCCGCTTCCGCGACCTGGAGGAGGACCCCTACCTGCCCGGCAACCCCCGCGAGC
TGATCGCCTACAGCCAGTACCCCCGCCCCAGCGACATCCCCCAGTGGAACAGCGACAAGCCCAGCCTGAAGGACATC
AAGATCATGGGCTACAGCATCCGCACCATCGACTACCGCTACACCGTGTGGGTGGGCTTCAACCCCGACGAGTTCCT
GGCCAACTTCAGCGACATCCACGCCGGCGAGCTGTACTTCGTGGACAGCGACCCCCTGCAGGACCACAACATGTACA
ACGACAGCCAGGGCGGCGACCTGTTCCAGCTGCTGATGCCCTAG

[0097] SEQ ID NO:8 编码全长前体甲酰甘氨酸生成酶 (FGE) 的示例性核酸序列

[0098] ATGGCTGCGCCCGCACTAGGGCTGGTGTGTGGACGTTGCCCTGAGCTGGGTCTCGTCCTCTTGCTGCTG
CTGCTCTCGCTGCTGTGTGGAGCGGCAGGGAGCCAGGAGGCCGGGACCGGTGCGGGCGCGGGGTCCCTTGCGGGTTC
TTGCGGCTGCGGCACGCCCCAGCGGCCTGGCGCCCATGGCAGTTCGGCAGCCGCTCACCGATACTCGCGGGAGGCTA
ACGCTCCGGGCCCCGTACCCGGAGAGCGGCAACTCGCGC

[0099] ACTCAAAGATGGTCCCCATCCCTGCTGGAGTATTTACAATGGGCACAGATGATCCTCAGATAAAGCAGG
ATGGGGAAGCACCTGCGAGGAGAGTTACTATTGATGCCTTTTACATGGATGCCTATGAAGTCAGTAATACTGAATTT
GAGAAGTTTGTGAACTCAACTGGCTATTTGACAGAGGCTGAGAAGTTTGGCGACTCCTTTGTCTTTGAAGGCATGTT
GAGTGAGCAAGTGAAGACCAATATTCAACAGGCAGTTGCAGCTGCTCCCTGGTGGTTACCTGTGAAAGGCGCTAACT
GGAGACACCCAGAAGGGCCTGACTCTACTATTCTGCACAGGCCGGATCATCCAGTTCTCCATGTGTCCTGGAATGAT
GCGGTTGCCTACTGCACTTGGGCAGGGAAGCGGCTGCCCACGGAAGCTGAGTGGGAATACAGCTGTCGAGGAGGCCT
GCATAATAGACTTTTCCCCTGGGGCAACAAACTGCAGCCCAAAGGCCAGCATTATGCCAACATTTGGCAGGGCGAGT
TTCCGGTGACCAACACTGGTGAGGATGGCTTCCAAGGAACTGCGCCTGTTGATGCCTTCCCTCCCAATGGTTATGGC
TTATACAACATAGTGGGGAACGCATGGGAATGGACTTCAGACTGGTGGACTGTTCATCATTCTGTTGAAGAAACGCT
TAACCCAAAAGGTCCCCCTTCTGGGAAAGACCGAGTGAAGAAAGGTGGATCCTACATGTGCCATAGGTCTTATTGTT
ACAGGTATCGCTGTGCTGCTCGGAGCCAGAACACACCTGATAGCTCTGCTTCGAATCTGGGATTCCGCTGTGCAGCC
GACCGCCTGCCCACCATGGACTGA

[0100] 在一些实施方案中,核苷酸变化可改变开放阅读框内的同义密码子,以与在被选择来表达 I2S 和 / 或 FGE 的特定异源细胞中发现的内源密码子使用一致。或者或另外地,核苷酸变化可能改变开放阅读框内的 G+C 含量,以更好地匹配在存在于异源宿主细胞中的内源核酸序列中发现的开放阅读框的平均 G+C 含量。核苷酸变化还可改变在 I2S 或 FGE 序列中发现的多聚单核苷酸区域或内部调控或结构位点。因而,设想了多种修饰或优化的核苷酸序列,包括但不限于,在原核细胞、酵母细胞、昆虫细胞以及哺乳动物细胞中提供增加的 I2S 和 / 或 FGE 蛋白的表达的核酸序列。

[0101] 因此,在一些实施方案中,适合于本发明的编码 I2S 蛋白的核酸具有与 SEQ ID NO:7 具有至少 50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高同一性的核苷酸序列。在一些实施方案中,适合于本发

明的编码FGE蛋白的核酸具有与SEQ ID NO:8具有至少50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高同一性的核苷酸序列。通常情况下,修饰核酸编码具有或不具有氨基酸序列改变的I2S和/或FGE蛋白。如果存在氨基酸改变,则这样的改变通常不显著改变I2S或FGE蛋白活性。

[0102] 表达载体

[0103] 可将编码如在本申请中描述的I2S和/或FGE蛋白的核酸序列分子克隆(插入)入合适的载体以用于在宿主细胞中扩增或表达。多种表达载体可被用于实施本发明,包括但不限于,原核表达载体;酵母表达载体;昆虫表达载体和哺乳动物表达载体。适合于本发明示例性载体包括但不限于基于病毒的载体(例如,基于AAV的载体、基于逆转录病毒的载体、基于质粒的载体)。在一些实施方案中,可将编码I2S和FGE蛋白的核酸序列分别插入分开的载体。在一些实施方案中,可将编码I2S和FGE蛋白的核酸序列分别插入同一载体中。通常情况下,编码I2S或FGE蛋白的核酸可操作地连接于各种调控序列或元件。

[0104] 调控序列或元件

[0105] 可将各种调控序列或元件整合在适合于本发明的表达载体中。示例性调控序列或元件包括但不限于启动子、增强子、阻遏子或抑制序列、5'非翻译(或非编码)序列、内含子、3'非翻译(或非编码)序列。

[0106] 如本文中所用,"启动子"或"启动子序列"是能够结合细胞中的RNA聚合酶(例如,直接地或通过其他启动子结合蛋白或物质)并起始编码序列的的转录的DNA调控区。启动子序列通常通过转录起始位点结合于其3'末端,并向上游(5'方向)延伸以包含以任何水平启动转录所必需的最小数目的碱基或元件。可将启动子与表达控制序列(包括增强子和阻遏序列)或与待表达的核酸可操作地连接,或可操作地连接于所述表达控制序列。在一些实施方案中,启动子可以是诱导型的。在一些实施方案中,诱导型启动子可以是单向的或生物定向的。在一些实施方案中,启动子可以是组成型启动子。在一些实施方案中,启动子可以是杂交启动子,其中包含转录调控区的序列获自一个来源并且包含转录起始区的序列获自第二来源。用于将控制元件连接于转基因内的编码序列的系统在本领域是公知的(一般的分子生物学和重组DNA技术描述于在Sambrook,Fritsch和Maniatis,Molecular Cloning:A Laboratory Manual,第2版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,NY,1989,其通过引用并入本文中)。适合于插入转基因以在多种生长和诱导条件于各种宿主细胞中表达的商业载体本领域中也是公知的。

[0107] 在一些实施方案中,可将特定启动子用于控制转基因在哺乳动物宿主细胞中的表达,例如但不限于,SRα-启动子(Takebe等人,Molec.and Cell.Bio.8:466-472(1988))、人CMV立即早期启动子(Boshart等人,Cell41:521-530(1985);Foecking等人,Gene 45:101-105(1986))、人CMV启动子、人CMV5启动子、鼠CMV立即早期启动子、EF1-α-启动子、用于肝特异性表达的杂交CMV启动子(例如,通过将CMV立即早期启动子与人α-1-抗胰蛋白酶(HAT)或白蛋白(HAL)启动子的转录启动子元件缀合产生)或用于肝细胞瘤特异性表达的启动子(例如,其中将人白蛋白(HAL;约1000bp)或人α-1-抗胰蛋白酶(HAT,约2000bp)的转录启动子元件与人α-1-微球蛋白和bikunin前体基因(AMBP)的145长的增强子元件组合;HAL-AMBP和HAT-AMBP)、SV40早期启动子区域(Benoist等人,Nature 290:304-310(1981))、黄杉毒蛾(Orgyia pseudotsugata)立即早期启动子、疱疹胸苷激酶

启动子 (Wagner 等人，Proc.Natl.Acad.Sci.USA 78:1441-1445(1981))；或金属硫蛋白基因的调控序列 (Brinster 等人，Nature 296:39-42(1982))。在一些实施方案中，哺乳动物启动子是组成型启动子，例如，但不限于，次黄嘌呤磷酸核糖转移酶 (HPTR) 启动子、腺苷脱氨酶启动子、丙酮酸激酶启动子、β - 肌动蛋白启动子以及对于本领域普通技术人员来说是已知的其他组成型启动子。

[0108] 在一些实施方案中，特定启动子可用于控制转基因在原核宿主细胞的表达，例如但不限于 β 内酰胺酶启动子 (Villa-Komaroff 等人，Proc.Natl.Acad.Sci.USA 75:3727-3731(1978))；tac 启动子 (DeBoer 等人，Proc.Natl.Acad.Sci.USA80:21-25(1983))；T7 启动子、T3 启动子、M13 启动子或 M16 启动子；在酵母宿主细胞中，例如但不限于 GAL1、GAL4 或 GAL10 启动子、ADH(醇脱氢酶) 启动子、PGK(磷酸甘油激酶) 启动子、碱性磷酸酶启动子、甘油醛 -3- 磷酸脱氢酶III (TDH3) 启动子、甘油醛 -3- 磷酸脱氢酶 II (TDH2) 启动子、甘油醛 -3- 磷酸脱氢酶 I (TDH1) 启动子、丙酮酸激酶 (PYK)、烯醇酶 (ENO) 或丙糖磷酸异构酶 (TPI)。

[0109] 在一些实施方案中，启动子可以是病毒启动子，其中有许多启动子能够在几个宿主细胞类型 (包括哺乳动物细胞) 中调节转基因的表达。已显示在真核细胞中驱动编码序列的组成型表达的病毒启动子包括例如猿病毒启动子、单纯疱疹病毒启动子、乳头状瘤病毒启动子、腺病毒启动子、人类免疫缺陷病毒 (HIV) 启动子、劳斯肉瘤病毒启动子、巨细胞病毒 (CMV) 启动子、莫洛尼鼠白血病病毒 (Moloney murine leukemia virus) 和其它逆转录病毒的长末端重复 (LTR)、单纯疱疹病毒的胸苷激酶启动子以及对于本领域普通技术人员来说是已知的其它病毒启动子。

[0110] 在一些实施方案中，表达载体的基因控制元件还可包括分别牵涉转录和翻译的起始的 5' 非转录和 5' 非翻译序列，例如 TATA 盒、加帽序列、CAAT 序列、Kozak 序列等。增强子元件可任选地用于增加待表达的多肽或蛋白质的表达水平。已显示在哺乳动物细胞中发挥作用的增强子元件的实例包括 SV40 早期基因增强子 (如在 Dijkema 等人，EMBO J.(1985)4:761 中描述的) 和来源于劳斯肉瘤病毒 (RSV)(如在 Gorman 等人，Proc.Natl.Acad.Sci.USA(1982b)79:6777 中描述的) 和人巨细胞病毒 (如 Boshart 等人，Cell(1985)41:521 中描述的) 的长末端重复 (LTR) 的增强子 / 启动子。表达载体的基因控制元件将还包括涉及转录和翻译终止的 3' 非转录和 3' 非翻译序列。分别地，例如多聚腺苷酸化 (polyA) 信号用于从启动子转录的 mRNA 的 3' 末端的稳定和加工。PolyA 信号包括，例如，兔 β 球蛋白 polyA 信号、牛生长激素 polyA 信号、鸡 β 球蛋白终止子 /polyA 信号或 SV40 晚期 polyA 区。

[0111] 可选择标记

[0112] 表达载体将优选地，但任选地包含至少一种可选择标记。在一些实施方案中，可选择标记是可操作地连接于一个或多个基因调控元件的编码抗性基因的核酸序列，当在细胞毒性化学药品和 / 或药物存在的情况下生长时，该核酸序列赋予宿主细胞维持活力的能力。在一些实施方案中，可选择试剂可用于维持表达载体在宿主细胞内的保留。在一些实施方案中，可以使用可选择试剂以防止表达载体内的转基因序列的修饰 (即甲基化) 和 / 或沉默。在一些实施方案中，可选择试剂用于维持载体在宿主细胞内的附加型表达。在一些实施方案中，可选择试剂用于促进转基因序列至宿主细胞基因组中的稳定整合。在一些实施

方案中，试剂和/或抗性基因可以包括，但不限于，氨甲蝶呤(MTX)、二氢叶酸还原酶(DHFR，美国专利号4,399,216；4,634,665；4,656,134；4,956,288；5,149,636；5,179,017)、氨苄青霉素、新霉素(G418)、zeomycin、霉酚酸或谷氨酰胺合成酶(GS，美国专利号5,122,464；5,770,359；5,827,739)(对于真核宿主细胞)；四环素、氨苄青霉素、卡那霉素或氯霉素(对于原核宿主细胞)；和URA3、LEU2、HIS3、LYS2、HIS4、ADE8、CUP1或TRP1(对于酵母宿主细胞)。

[0113] 可将表达载体转染、转化或转导入宿主细胞。如本文中所用，术语“转染”、“转化”和“转导”均指外源核酸序列至宿主细胞中的引入。在一些实施方案中，将包含编码I2S和/或FGE的核酸序列的表达载体同时转染、转化或转导入宿主细胞。在一些实施方案中，将包含编码I2S和/或FGE的核酸序列的表达载体依次转染、转化或转导入宿主细胞。例如，首先可将编码I2S蛋白的载体转染、转化或转导入宿主细胞，随后转染、转化或转导编码FGE蛋白的载体，反之亦然。本领域中公知的转化、转染和转导方法的实例包括脂质体递送，即Hawley-Nelson,Focus 15:73(1193)的lipofectamine™(Gibco BRL)法、电穿孔、Graham和van der Erb,Virology,52:456-457(1978)的$CaPO_4$递送法、DEAE-葡聚糖介导的递送、显微注射、生物轰击颗粒递送、聚凝胺介导的递送、阳离子介导的脂质递送、转导和病毒感染，例如，逆转录病毒、慢病毒、腺病毒腺相关病毒和杆状病毒(Baculovirus)(昆虫细胞)。细胞宿主转化的一般方面在本领域中例如由Axel在美国专利号4,399,216；Sambrook，同上，第1-4和16-18章；Ausubel，同上，第1,9,13,15和16章中进行了描述。关于用于转化哺乳动物细胞的各种技术，参见Keown等人，Methods in Enzymology(1989)，Keown等人，Methods in Enzymology,185:527-537(1990)和Mansour等人，Nature,336:348-352(1988)。

[0114] 在被引入细胞后，表达载体可被稳定地整合在基因组中，或作为染色体外构建体存在。载体还可被扩增，并且多个拷贝可存在于基因组中或被整合在基因组。在一些实施方案中，本发明的细胞可包含1、2、3、4、5、6、7、8、9、10、15、20或更多个拷贝的编码I2S蛋白的核酸。在一些实施方案中，本发明的细胞可包含1、2、3、4、5、6、7、8、9、10、15、20或更多个拷贝的编码FGE蛋白的核酸。在一些实施方案中，本发明的细胞可包含多个拷贝(例如，2、3、4、5、6、7、8、9、10、15、20或更多个)的编码I2S和FGE蛋白的核酸。

[0115] 宿主细胞

[0116] 如本文中所用，术语“宿主细胞”指可用于产生重组I2S酶的细胞。具体地，宿主细胞适合于大规模产生重组I2S酶。适合的宿主细胞可来源于多种生物体，包括但不限于哺乳动物、植物、鸟类(例如，禽系统)、昆虫、酵母和细菌。在一些实施方案中，宿主细胞是哺乳动物细胞。在一些实施方案中，适合的宿主细胞不是内体酸化缺陷型细胞。

[0117] 哺乳动物细胞系

[0118] 可按照本发明将易于进行细胞培养和多肽表达的任何哺乳动物细胞或细胞类型用作宿主细胞。可按照本发明使用的哺乳动物细胞的非限制性实例包括人胚胎肾293细胞(HEK293)、HeLa细胞、BALB/c小鼠骨髓瘤细胞系(NSO/1，ECACC号：85110503)、人成视网膜细胞(PER.C6(CruCell,Leiden,The Netherlands))、由SV40转化的猴肾CV1细胞系(COS-7，ATCC CRL 1651)、人胚肾系(经亚克隆以在悬浮培养物中生长的293或293细胞，Graham等人，J.Gen Virol.,36:59(1977))、幼仓鼠肾细胞(BHK，ATCC CCL10)、中国仓鼠卵巢细胞+/-DHFR(CHO,Urlaub和Chasin,Proc.Natl.Acad.Sci.USA,77:4216(1980))、小

鼠支持细胞 (sertoli cell)(TM4,Mather,Biol.Reprod.,23:243-251(1980))、猴肾细胞 (CV1ATCC CCL70)、非洲绿猴肾细胞 (VERO-76,ATCC CRL-1587)、人宫颈癌细胞 (HeLa,ATCC CCL 2)、犬肾细胞 (MDCK,ATCC CCL 34)、buffalo大鼠肝细胞 (BRL3A,ATCC CRL 1442)、人肺细胞 (W138,ATCC CCL 75)、人肝细胞病毒 (Hep G2,HB 8065)、小鼠乳腺肿瘤 (MMT 060562,ATCC CCL51)、TRI细胞 (Mather等人,Annals N.Y.Acad.Sci.,383:44-68(1982))、MRC 5细胞、FS4细胞和人肝细胞瘤细胞系 (Hep G2)。在一些实施方案中,适合的哺乳动物细胞不是内体酸化缺陷型细胞。

[0119] 此外,可按照本发明使用任何数目的表达多肽或蛋白质的商业可得和非商购可得的杂交瘤细胞系。本领域技术人员将理解,杂交瘤细胞系可能具有不同的营养需要和/或可能需要不同的培养条件来进行最佳生长和多肽或蛋白质表达,并将能够根据需要来修改条件。

[0120] 非哺乳动物细胞系

[0121] 可按照本发明将易于进行细胞培养和多肽表达的任何非哺乳动物来源的细胞或细胞类型用作宿主细胞。可按照本发明使用的非哺乳动物宿主细胞和细胞系的非限制性实例包括来源于巴斯德毕赤酵母 (Pichia pastoris)、甲醇毕赤酵母 (Pichia methanolica)、安格斯毕赤酵母 (Pichia angusta)、粟酒裂殖酵母 (Schizosacccharomyces pombe)、酿酒酵母 (Saccharomyces cerevisiae) 和解脂耶氏酵母 (Yarrowia lipolytica)(对于酵母);草地夜蛾 (Sodoptera frugiperda)、粉纹夜蛾 (Trichoplusis ni)、黑腹果蝇 (Drosophila melangoster) 和烟草天蛾 (Manduca sexta)(对于昆虫);和大肠杆菌 (Escherichia coli)、鼠伤寒沙门氏菌 (Salmonella typhimurium)、枯草芽孢杆菌 (Bacillus subtilis)、地衣形芽孢杆菌 (Bacillus lichenifonnis)、脆弱拟杆菌 (Bacteroides fragilis)、产气荚膜梭状芽胞杆菌 (Clostridia perfringens)、难辨梭状芽孢杆菌 (Clostridia difficile)(对于细菌);和来自两栖动物的非洲爪蟾 (Xenopus Laevis) 的细胞和细胞系。

[0122] 适应贴壁生长对悬浮生长

[0123] 在某些实施方案中,宿主细胞基于某些优选属性被选择用于产生细胞系,或在被选择用于培养细胞的特定条件下生长。本领域技术人员将理解,这样的属性可以基于已建立的细胞系(即表征的商购可得细胞系)的已知特征和/或性状或通过经验评价来确定。在一些实施方案中,可针对其在饲养层细胞上生长的能力选择细胞系。在一些实施方案中,可针对其在悬浮液中生长的能力选择细胞系。在一些实施方案中,可针对作为贴壁单层细胞生长的能力选择细胞系。在一些实施方案中,可将这样的细胞与任何组织培养皿或用适当的粘附基质处理的任何器皿一起使用。在一些实施方案中,适当的粘附基质选自胶原蛋白(例如胶原蛋白 I、II、II 或 IV)、明胶、纤连蛋白、层粘连蛋白、玻连蛋白、纤维蛋白原、BD Matrigel™、基底膜基质、硫酸皮肤素蛋白聚糖、多聚-D-赖氨酸和/或其组合。在一些实施方案中,可在特定生长条件下选择和修饰贴壁宿主细胞以在悬浮液中生长。这样的修饰贴壁细胞以在悬浮液中生长的方法在本领域中是已知的。例如,可通过随时间逐渐从生长培养基除去动物血清来对细胞进行调节,以在悬浮培养中生长。

[0124] 细胞系选择和评价

[0125] 根据本发明,针对其以商业可行性规模产生重组 I2S 蛋白的能力选择经工程化以表达重组 I2S 蛋白的细胞。具体地,根据本发明的工程化细胞能够以高水平和/或以高酶

活性产生重组 I2S。在一些实施方案中，一旦在培养条件（例如，标准大规模悬浮或贴壁培养条件）下培养后，期望的细胞可以以如下量产生 I2S 酶：5 皮克 / 细胞 / 日或大于 5 皮克 / 细胞 / 日（例如，大于约 10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95 或 100 皮克 / 细胞 / 日）。在一些实施方案中，一旦在细胞培养条件（例如，标准大规模悬浮或贴壁培养条件）下培养后，期望的细胞能够以在如下范围内的量产生 I2S 酶：约 5-100 皮克 / 细胞 / 日（例如，约 5-90 皮克 / 细胞 / 日、约 5-80 皮克 / 细胞 / 日、约 5-70 皮克 / 细胞 / 日、约 5-60 皮克 / 细胞 / 日、约 5-50 皮克 / 细胞 / 日、约 5-40 皮克 / 细胞 / 日、约 5-30 皮克 / 细胞 / 日、约 10-90 皮克 / 细胞 / 日、约 10-80 皮克 / 细胞 / 日、约 10-70 皮克 / 细胞 / 日、约 10-60 皮克 / 细胞 / 日、约 10-50 皮克 / 细胞 / 日、约 10-40 皮克 / 细胞 / 日、约 10-30 皮克 / 细胞 / 日、约 20-90 皮克 / 细胞 / 日、约 20-80 皮克 / 细胞 / 日、约 20-70 皮克 / 细胞 / 日、约 20-60 皮克 / 细胞 / 日、约 20-50 皮克 / 细胞 / 日、约 20-40 皮克 / 细胞 / 日、约 20-30 皮克 / 细胞 / 日）。

[0126] 如上所讨论的，通常情况下，I2S 的酶活性受保守半胱氨酸（例如，在氨基酸 59 上）至甲酰甘氨酸的翻译后修饰影响。这种翻译后修饰一般在蛋白质合成过程中发生于内质网中，并且由 FGE 催化。I2S 的酶活性通常与 I2S 具有甲酰甘氨酸修饰所达到的程度正相关。例如，具有相对高的甲酰甘氨酸修饰量的 I2S 制剂通常具有相对高的酶比活性；然而具有相对低的甲酰甘氨酸修饰量的 I2S 制剂通常具有相对低的酶比活性。

[0127] 还可以预期，I2S 与 FGE 蛋白或 mRNA 之间的比率还可以影响所产生的重组 I2S 蛋白上的甲酰甘氨酸修饰。在一些实施方案中，在期望的细胞中表达的 I2S 和 FGE 具有不同的蛋白质和 / 或 mRNA 表达水平。在一些实施方案中，I2S 蛋白或 mRNA 表达水平为 FGE 的蛋白质或 mRNA 水平的至少 0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1.0、1.5、2.0、2.5、3.0、3.5、4.0、4.5、5.0、5.5、6.0、6.5、7.0、7.5、8、9 或 10 倍。在一些实施方案中，重组 FGE 蛋白或 mRNA 表达水平为 I2S 的蛋白质或 mRNA 水平的至少 0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1.0、1.5、2.0、2.5、3.0、3.5、4.0、4.5、5.0、5.5、6.0、6.5、7.0、7.5、8、9 或 10 倍。

[0128] 在一些实施方案中，一旦在细胞培养条件（例如，标准大规模悬浮或贴壁细胞培养条件）下培养后，期望的细胞可产生 I2S 蛋白，该蛋白包含至少约 50％（例如，至少约 55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％或 100％）的对应于 SEQ ID NO:1 的 Cys59 的半胱氨酸残基至 Cα - 甲酰甘氨酸 (FGly) 的转化。在一些实施方案中，一旦在细胞培养条件（例如，标准大规模悬浮或贴壁细胞培养条件）下培养后，期望的细胞可以以如下量产生 I2S 酶：约 5 皮克 / 细胞 / 日的量或大于约 5 皮克 / 细胞 / 日（例如，大于约 10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95 或 100 皮克 / 细胞 / 日），该酶包含至少约 50％（例如，至少约 55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％或 100％）的对应于 SEQ ID NO:1 的 Cys59 的半胱氨酸残基至 Cα - 甲酰甘氨酸 (FGly) 的转化。

[0129] FGly 转化百分比

[0130] 各种方法是已知的，并且可以用于测定 FGly 转化百分比。通常，甲酰甘氨酸转化率的百分比 (％ FG) 可使用下列公式来计算：

[0131]

$$\%FG\ (DS的) = \frac{活性I2S分子的数目}{总的(活性+无活性)I2S分子的数目} \times 100$$

[0132] 例如50％FG意味着一半的纯化的重组I2S是无酶活性的（无任何治疗效果）。各种方法可用于计算％FG。例如，可使用肽图谱。简言之，可使用蛋白酶（例如，胰蛋白酶或糜蛋白酶）将I2S蛋白消化成短肽。可使用色谱法（例如，HPLC）分离和表征短肽，以便可相较于对照（例如，无FGly转化的I2S蛋白或具有100％FGly转化的I2S蛋白）测定每一个肽（特别地包含对应于成熟人I2S的位置59的位置的肽）的性质和量。可测定包含FGly的肽的量（对应于活性I2S分子的数目）和具有FGly和Cys的肽的总量（对应于总的I2S分子的数目），和计算反映％FG的比率。

[0133] 比活性

[0134] 如上所讨论的，通常情况下，I2S的酶活性受保守半胱氨酸（例如，在氨基酸59）至甲酰甘氨酸的翻译后修饰影响。因此，I2S的酶活性通常与I2S具有甲酰甘氨酸修饰所达到的程度正相关。例如，具有相对高的甲酰甘氨酸修饰量的I2S制剂通常具有相对高的酶比活性；然而具有相对低的甲酰甘氨酸修饰量的I2S制剂通常具有相对低的酶比活性。

[0135] 如本领域技术人员可理解的，由本发明的细胞产生的重组I2S蛋白的酶活性可通过各种体外和体内测定来测量。在一些实施方案中，如使用肝素二糖作为底物通过体外硫酸酯释放活性测定测量的，产生的重组I2S蛋白的期望的酶活性为至少约20U/mg、30U/mg、40U/mg、50U/mg、60U/mg、70U/mg、80U/mg、90U/mg或100U/mg。在一些实施方案中，如使用肝素二糖作为底物通过体外硫酸酯释放活性测定测量的，产生的重组I2S蛋白的期望的酶活性在约20-100U/mg（例如，约20-90U/mg、约20-80U/mg、约20-70U/mg、约20-60U/mg、约20-50U/mg、约20-40U/mg、约20-30U/mg、约30-100U/mg、约30-90U/mg、约30-80U/mg、约30-70U/mg、约30-60U/mg、约30-50U/mg、约30-40U/mg、约40-100U/mg、约40-90U/mg、约40-80U/mg、约40-70U/mg、约40-60U/mg、约40-50U/mg）的范围内。下文中提供了用于使用肝素二糖作为底物进行体外硫酸酯释放活性测定的示例性条件。通常情况下，该测定测量I2S从天然来源的底物肝素二糖释放硫酸根离子的能力。所释放的硫酸酯可通过离子色谱法来进行定量。在一些情况下，离子色谱法配备有电导率检测器。作为一个非限制性实例，首先将样品缓冲交换至10mM乙酸钠（pH 6），以解除配制缓冲液中磷酸根离子产生的抑制作用。随后用反应缓冲液（10mM乙酸钠，pH 4.4）将样品稀释至0.075mg/ml，并在37℃于30μL反应体积中以0.3μg I2S/100μg底物的酶与底物比与肝素二糖温育2小时。然后通过在100℃加热3分钟来终止反应。使用具有IonPac AG18保护柱的Dionex IonPac AS18分析柱进行分析。以1.0mL/分钟将等度方法与30mM氢氧化钾一起使用，持续15分钟。从在1.7至16.0纳摩尔的范围内的硫酸酯标准的线性回归分析计算由I2S样品释放的硫酸酯的量。该报告值表达为单位每mg蛋白质，其中1个单位定义为每小时释放的1微摩尔的硫酸酯，并且蛋白质浓度通过A280测量来测定。

[0136] 在一些实施方案中，还可使用本领域已知的各种其它方法，例如测量4-甲基伞形酮基－硫酸酯至硫酸酯和天然发荧光的4-甲基伞形酮（4-MUF）的水解的4-MUF测定来测定由本发明的细胞产生的重组I2S蛋白的酶活性。在一些实施方案中，如通过体外4-MUF

测定测量的，产生的重组 I2S 蛋白的期望的酶活性为至少约 2U/mg、4U/mg、6U/mg、8U/mg、10U/mg、12U/mg、14U/mg、16U/mg、18U/mg 或 20U/mg。在一些实施方案中，如通过体外 4-MUF 测定测量的，产生的重组 I2S 蛋白的期望的酶活性在约 0-50U/mg（例如，约 0-40U/mg、约 0-30U/mg、约 0-20U/mg、约 0-10U/mg、约 2-50U/mg、约 2-40U/mg、约 2-30U/mg、约 2-20U/mg、约 2-10U/mg、约 4-50U/mg、约 4-40U/mg、约 4-30U/mg、约 4-20U/mg、约 4-10U/mg、约 6-50U/mg、约 6-40U/mg、约 6-30U/mg、约 6-20U/mg、约 6-10U/mg）的范围内。下文中提供了用于进行体外 4-MUF 测定的示例性条件。通常情况下，4-MUF 测定测量 I2S 蛋白将 4- 甲基伞形酮基硫酸酯（4-MUF-SO_4）水解成硫酸酯和天然发荧光的 4- 甲基伞形酮（4-MUF）的能力。1 毫单位的活性被定义为在 37℃于 1 分钟内将 1 纳摩尔的 4-MUF-SO_4转化成 4-MUF 所需的酶的量。通常情况下，通过具有已知活性的 I2S 测试样品产生的平均荧光单位（MFU）可用于产生标准曲线，该曲线可被用来计算目标样品的酶活性。

[0137] 细胞培养基和条件

[0138] 各种细胞培养基和条件可用于使用根据本发明的工程化的细胞产生重组 I2S 蛋白。例如，可在含血清或无血清的培养基中产生重组 I2S 蛋白。在一些实施方案中，在无血清培养基中产生重组 I2S 蛋白。在一些实施方案中，在无动物的培养基，即缺乏动物来源组分的培养基中产生重组 I2S 蛋白。在一些实施方案中，在化学成分确定的培养基中产生重组 I2S 蛋白。如本文中所用，术语“化学成分确定的营养培养基”是指基本上所有的化学成分是已知的培养基。在一些实施方案中，化学成分确定的培养基不含动物来源的组分如血清、血清来源的蛋白质（例如，白蛋白或胎球蛋白），以及其他组分。在一些情况下，化学成分确定的培养基包含一种或多种蛋白质（例如，蛋白质生长因子或细胞因子）。在一些情况下，化学成分确定的培养基包含一种或多种蛋白质水解产物。在其他情况下，化学成分确定的营养培养基是不含蛋白质的培养基，即，不包含蛋白质、水解产物或未知组合物的组分的无血清培养基。

[0139] 在一些实施方案中，化学成分确定的培养基可补充一种或多种动物来源组分。这样的动物来源的组分包括但不限于胎牛血清、马血清、山羊血清、驴血清、人血清和血清来源的蛋白质例如白蛋白（例如，牛血清白蛋白或人血清白蛋白）。

[0140] 各种细胞培养条件可以用于大规模产生重组 I2S 蛋白，包括但不限于滚瓶培养、生物反应器分批培养和生物反应器补料分批培养。在一些实施方案中，重组 I2S 蛋白是由悬浮培养的细胞产生的。在一些实施方案中，重组 I2S 蛋白是通过贴壁细胞产生的。

[0141] 示例性细胞培养基和培养条件在实施例部分中进行了描述。用于产生重组 I2S 蛋白的另外的示例性方法和组合物在与其同一日提交的标题为“用于产生重组艾杜糖 -2- 硫酸酯酶的方法和组合物（Methods and Compositions for Producing Recombinant Iduronate-2-Sulfatase）”的临时申请中进行了描述，所述临时申请的全部公开内容通过引用并入本文。

[0142] 表达的 I2S 蛋白的纯化

[0143] 各种方法可用于纯化或分离根据本文所述的各种方法产生的 I2S 蛋白。在一些实施方案中，表达的 I2S 蛋白被分泌至培养基中，从而可以例如通过离心或过滤（例如，作为纯化过程的第一步骤）除去细胞和其他固体。或者或另外地，将表达的 I2S 蛋白结合在宿主细胞的表面上。在本实施方案中，将表达多肽或蛋白质的宿主细胞裂解以进行纯化。哺

乳动物宿主细胞的裂解可以通过本领域普通技术人员公知的许多方法，包括利用玻璃珠进行的物理破碎和对高 pH 条件的暴露来实现。

[0144] 可通过标准方法（包括但不限于色谱法（例如，离子交换色谱、亲和色谱、尺寸排阻色谱和羟基磷灰石色谱）、凝胶过滤、离心或差异溶解度、乙醇沉淀）或通过任何其它可获得的用于蛋白质纯化的技术（参见，例如 Scopes, Protein Purification Principles and Practice 第 2 版，Springer-Verlag, New York, 1987；Higgins, S. J. 和 Hames, B.D.（编辑），Protein Expression: A Practical Approach, Oxford Univ Press, 1999；和 Deutscher, M.P., Simon, M.I., Abelson, J.N.（编辑），Guide to Protein Purification: Methods in Enzymology (Methods in Enzymology Series, 第 182 卷), Academic Press, 1997，将所有参考资料通过引用并入本文）分离和纯化 I2S 蛋白。对于具体地免疫亲和色谱法，可通过将蛋白质结合于包含抗体的亲和柱来使其分离，所述抗体是针对该蛋白质产生的并且被附着至固定载体。或者，可通过标准重组技术将亲和标签例如流感病毒衣壳蛋白序列、多聚组氨酸或谷胱甘肽 -S- 转移酶连接于蛋白质，以使其通过在适当的亲和柱上通过来容易地纯化。可以在任何或所有阶段添加蛋白酶抑制剂，例如苯基甲基磺酰氟 (PMSF)、亮肽素、胃酶抑素或抑肽酶以在纯化过程中减少或消除所述多肽或蛋白质的降解。当必需裂解细胞以分离和纯化表达的多肽或蛋白质时，特别想要蛋白酶抑制剂。

[0145] 示例性纯化方法描述于下面的实施例部分中。另外的纯化方法描述于与其同一日提交的标题为“重组 I2S 蛋白的纯化 (Purification of Recombinant I2S Protein)”的临时申请中，所述临时申请的全部公开内容通过引用并入本文。

[0146] 药物组合物和施用

[0147] 可按照已知的方法向亨特综合症患者施用纯化的重组 I2S 蛋白。例如，可静脉内、皮下、肌内、肠胃外、透皮或透粘膜（例如，经口或经鼻））递送纯化的重组 I2S 蛋白。

[0148] 在一些实施方案中，可通过静脉内施用向受试者施用重组 I2S 或包含其的药物组合物。

[0149] 在一些实施方案中，通过鞘内施用向受试者施用重组 I2S 或包含其的药物组合物。如本文中所用的，术语“鞘内施用”或“鞘内注射”是指至椎管（围绕脊髓的鞘内空间）内的注射。可使用各种技术，包括但不限于，通过钻孔的侧脑室注射或池穿刺或腰椎穿刺等。在一些实施方案中，根据本发明的“鞘内施用”或“鞘内递送”是指通过腰部区或区域的 IT 施用或递送，即腰 IT 施用或递送。如本文中所用，术语“腰部区域”或“腰部区”指的是第三和第四腰（下背部）椎骨之间的区，更包含地，脊椎的 L2-S1 区域。

[0150] 在一些实施方案中，通过皮下（即皮肤下方）向受试者施用重组 I2S 或包含其的药物组合物。为此目的，可使用注射器注射制剂。然而，用于该制剂的施用的其它装置是可获得的，例如注射装置（例如，Inject-ease™和 Genject ™装置）；注射器笔（例如 GenPen™）；无针装置（例如，MediJector™和 BioJector ™）；和皮下贴片递送系统。

[0151] 在一些实施方案中，可将鞘内施用与其他施用途径（例如，静脉内、皮下、肌内、肠胃外、透皮或透粘膜（例如，经口或经鼻））一起使用。

[0152] 本发明设想了治疗有效量的本文中描述的重组 I2S 或包含所述重组 I2S 的药物组合物的单次以及多次施用。可定期施用重组 I2S 或包含其的药物组合物，这取决于受试者

的病况（例如，溶酶体贮积病）的性质、严重性和程度。在一些实施方案中，可定期（例如，每年一次，每六个月一次，每五个月一次，每三个月一次，每两月一次（每两个月一次），每月一次（每个月一次），每两周一次（每隔两周一次），每周一次，每日一次或连续地）周期性施用治疗有效的量的重组 I2S 或包含其的药物组合物。

[0153] 可用生理上可接受的载体或赋形剂配制重组 I2S 或包含其的药物组合物以制备药物组合物。载体和治疗剂可以是无菌的。所述制剂应该适合施用模式。

[0154] 适合的药学上可接受的载体包括但不限于：水、盐溶液（例如，NaCl）、盐水、缓冲盐水、醇、甘油、乙醇、阿拉伯胶、植物油、苄醇、聚乙二醇、明胶、碳水化合物如乳糖、直链淀粉或淀粉、糖例如甘露醇、蔗糖或其他糖、葡萄糖、硬脂酸镁、滑石、硅酸、粘性石蜡、芳香油、脂肪酸酯、羟甲基纤维素、聚乙烯吡咯烷酮等及其组合。必要时，可将药物制剂与助剂（例如，润滑剂、防腐剂、稳定剂、润湿剂、乳化剂、用于影响渗透压的盐、缓冲剂、着色剂、调味剂和 / 或芳香物质等）混合，所述助剂其不与活性化合物产生不利反应或干扰它们的活性。在一些实施方案中，使用适合于静脉内施用的水溶性载体。

[0155] 必要时，组合物或药剂还可包含少量润湿剂或乳化剂或 pH 缓冲剂。组合物可以是液体溶液、悬浮液、乳液、片剂、丸剂、胶囊剂、持续释放制剂或粉剂。还可利用常规粘合剂和载体如甘油三酯将组合物配制成栓剂。口服制剂可以包含标准载体如药物级的甘露醇、乳糖、淀粉、硬脂酸镁、聚乙烯吡咯烷酮、糖精钠、纤维素、碳酸镁等。

[0156] 可按照常规程序将组合物或药剂配制为适于向人类施用的药物组合物。例如，在一些实施方案中，用于静脉内施用的组合物通常是在无菌等渗含水缓冲液中的溶液。必要时，组合物还可以包含增溶剂和局部麻醉剂以减轻注射部位的疼痛。通常，单独地提供成分或以单位剂型将成分混合在一起，例如，作为标明活性剂的量的密闭容器例如安瓿或小袋中的干燥的冻干粉剂或无水浓缩物。当将通过输注施用组合物时，可用含有无菌药用级水、盐水或右旋糖 / 水的输液瓶分配所述组合物。当通过注射施用组合物时，可提供一安瓿的注射用无菌水或盐水以便可在施用前混合成分。

[0157] 如本文中所用，主要基于本发明的药物组合物中包含的治疗剂的总量确定术语“治疗有效量”。通常，治疗有效量足以对受试者实现有意义的益处（例如，治疗、调节、治愈、预防和 / 或改善潜在的疾病或病症）。例如，治疗有效量可以是足以实现期望的治疗和 / 或预防效果的量，例如足以调节溶酶体酶受体或它们的活性，从而治疗这样的溶酶体贮积病或其症状（例如，在向受试者施用本发明的组合物后减少或消除“斑状体”或细胞空泡形成的存在或发生率）的量。通常地，向需要其的受试者施用的治疗剂（例如，重组溶酶体酶）的量将取决于受试者的特征。这样的特征包括受试者的状态、疾病的严重度、一般健康状况、年龄、性别和体重。取决于这些和其他相关因素，本领域普通技术人员将能够容易地确定适当的剂量。此外，可任选地将客观和主观测定用于鉴定最佳剂量范围。

[0158] 通常在可包括多个单位剂量的给药方案中施用治疗有效量。对于任何特定的治疗性蛋白，治疗有效量（和 / 或有效给药方案内的适当单位剂量）可以例如取决于施用途径、取决于与其它药剂的组合而变化。同样地，对于任何特定患者的具体的治疗有效量（和 / 或单位剂量）还可取决于多种因素，包括被治疗的病症和病症的严重程度；所使用的具体药剂的活性；使用的具体组合物；患者的年龄、体重、一般健康状况、性别和饮食；施用时间、施用途径和 / 或使用的特定融合蛋白的排泄或代谢速率；治疗的持续时间；以及在医学

领域中公知的类似因素。

[0159] 另外的示例性药物组合物和施用方法描述于标题为“用于艾杜糖 -2- 硫酸酯酶的 CNS 递送的方法和组合物 (Methods and Compositions for CNS Deliv ery of Iduronate-2-Sulfatase)”的 PCT 公布 WO2011/163649 ;和 2012 年 3 月 30 日提交的标题为“艾杜糖醛酸 2 硫酸酯酶的皮下施用 (Subcutaneous administ ration of iduronate 2sulfatase)”的临时申请序列号 61/618,638,将这两者的全部公开内容通过引用并入本文。

[0160] 应当进一步理解,对于任何特定的受试者,应当根据个体需要和施用或监督酶替代疗法的施用的人的专业判断来随时间调整具体剂量方案,并且本文中所陈述的剂量范围仅是示例性的并且无意限定本发明的范围或实践。

实施例

[0161] 实施例 1. 共表达重组 I2S 和 FGE 的最优化的细胞系的产生

[0162] 本实施例举例说明可用于产生重组 I2S 蛋白的共表达重组 I2S 和 FGE 的示例性最优化的细胞系。对于本领域技术人员将很明确的是许多替代方法、表达载体和克隆技术是可获得的。

[0163] 人艾杜糖的 2- 硫酸酯酶 (I2S) 的典型成熟形式是 525 个氨基酸的糖蛋白,其经历大量的加工和翻译后修饰以进行酶的激活,如糖基化和半胱氨酸至甲酰甘氨酸的转化 (图 1)。在哺乳动物细胞中,I2S 酶内的保守半胱氨酸残基 (即,在氨基酸 59 上) 通过甲酰甘氨酸生成酶 (FGE) 被转化为甲酰甘氨酸。I2S 酶的活性位点内的半胱氨酸至甲酰甘氨酸的转化是产生人硫酸酯酶的活性形式中的一个重要步骤。本实验的目的是工程化共表达 I2S 和 FGE 的最优化人细胞系以用于产生活性重组 I2S。

[0164] 图 2 举例说明了许多用于共表达 I2S 和 FGE 的示例性构建体设计。例如,I2S 和 FGE 的表达单位可以位于单独的载体,并且可共转染或单独转染单独的载体 (图 2A)。或者,I2S 和 FGE 的表达单位可以位于同一载体 (图 2B)。在一种配置中,I2S 和 FGE 可以在相同的载体上,但在单独的启动子控制下,也被称为单独的顺反子 (图 2B(1))。或者,I2S 和 FGE 可以设计为转录连接的顺反子,即,I2S 和 FGE 被设计为在相同启动子控制下的一个开放阅读框 (图 2B(2))。通常情况下,内部核糖体进入位点 (IRES) 被设计来允许信使 RNA 的不依赖于帽的翻译起始 (图 2B(2))。

[0165] 将人细胞系工程化以共表达具有 SEQ ID NO :2 中显示的氨基酸序列的人 I2S 蛋白和具有 SEQ ID NO :6 中显示的氨基酸序列的人甲酰甘氨酸生成酶 (FGE)。

[0166] SEQ ID NO :2

[0167] ＞全长前体艾杜糖 -2- 硫酸酯酶

[0168] MPPPRTGRGLLWLGLVLSSVCVALGSETQANSTTDALNVLLIIVDDLRPSLGCYGDKLVRSPNIDQLAS
HSLLFQNAFAQQAVCAPSRVSFLTGRRPDTTRLYDFNSYWRVHAGNFSTIPQYFKENGYVTMSVGKVFHPGISSNHT
DDSPYSWSFPPYHPSSEKYENTKTCRGPDGELHANLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSASPFFLAV
GYHKPHIPFRYPKEFQKLYPLENITLAPDPEVPDGLPPVAYNPWMDIRQREDVQALNISVPYGPIPVDFQRKIRQSY
FASVSYLDTQVGRLLSALDDLQLANSTIIAFTSDHGWALGEHGEWAKYSNFDVATHVPLIFYVPGRTASLPEAGEKL
FPYLDPFDSASQLMEPGRQSMDLVELVSLFPTLAGLAGLQVPPRCPVPSFHVELCREGKNLLKHFRFRDLEEDPYLP

GNPRELIAYSQYPRPSDIPQWNSDKPSLKDIKIMGYSIRTIDYRYTVWVGFNPDEFLANFSDIHAGELYFVDSDPLQ
DHNMYNDSQGGDLFQLLMP

[0169] SEQ ID NO:6

[0170] 全长人 FGE 前体：

[0171] MAAPALGLVCGRCPELGLVLLLLLLSLLCGAAGSQEAGTGAGAGSLAGSCGCGTPQRPGAHGSSAAAHR
YSREANAPGPVPGERQLAHSKMVPIPAGVFTMGTDDPQIKQDGEAPARRVTIDAFYMDAYEVSNTEFEKFVNSTGYL
TEAEKFGDSFVFEGMLSEQVKTNIQQAVAAAPWWLPVKGANWRHPEGPDSTILHRPDHPVLHVSWNDAVAYCTWAGK
RLPTEAEWEYSCRGGLHNRLFPWGNKLQPKGQHYANIWQGEFPVTNTGEDGFQGTAPVDAFPPNGYGLYNIVGNAWE
WTSDWWTVHHSVEETLNPKGPPSGKDRVKKGGSYMCHRSYCYRYRCAARSQNTPDSSASNLGFRCAADRLPTMD

[0172] 为了产生表达 I2S 的细胞系，用编码具有 SEQ ID NO:2 中显示的氨基酸序列的 I2S 蛋白的密码子优化的核酸序列 (SEQ ID NO.7) 和编码 SEQ ID NO.6 中显示的人 FGE 酶的核酸序列 (SEQ ID NO.8) 稳定地转染细胞。

[0173] SEQ ID NO:7

[0174] ＞智人密码子优化的艾杜糖 -2- 硫酸酯酶 (IDS)、转录变体 1、mRNA

[0175] ATGCCCCCGCCCCGCACCGGCCGCGGCCTGCTGTGGCTGGGCCTGGTGCTGAGCAGCGTGTGCGTGGCC
CTGGGCAGCGAGACCCAGGCCAACAGCACCACCGACGCCCTGAACGTGCTGCTGATCATCGTGGACGACCTGCGCCC
CAGCCTGGGCTGCTACGGCGACAAGCTGGTGCGCAGCCCCAACATCGACCAGCTGGCCAGCCACAGCCTGCTGTTCC
AGAACGCCTTCGCCCAGCAGGCCGTGTGCGCCCCCAGCCGCGTGAGCTTCCTGACCGGCCGCCGCCCCGACACCACC
CGCCTGTACGACTTCAACAGCTACTGGCGCGTGCACGCCGGCAACTTCAGCACCATCCCCCAGTACTTCAAGGAGAA
CGGCTACGTGACCATGAGCGTGGGCAAGGTGTTCCACCCCGGCATCAGCAGCAACCACACCGACGACAGCCCCTACA
GCTGGAGCTTCCCCCCCTACCACCCCAGCAGCGAGAAGTACGAGAACACCAAGACCTGCCGCGGCCCCGACGGCGAG
CTGCACGCCAACCTGCTGTGCCCCGTGGACGTGCTGGACGTGCCCGAGGGCACCCTGCCCGACAAGCAGAGCACCGA
GCAGGCCATCCAGCTGCTGGAGAAGATGAAGACCAGCGCCAGCCCCTTCTTCCTGGCCGTGGGCTACCACAAGCCCC
ACATCCCCTTCCGCTACCCCAAGGAGTTCCAGAAGCTGTACCCCCTGGAGAACATCACCCTGGCCCCCGACCCCGAG
GTGCCCGACGGCCTGCCCCCCGTGGCCTACAACCCCTGGATGGACATCCGCCAGCGCGAGGACGTGCAGGCCCTGAA
CATCAGCGTGCCCTACGGCCCCATCCCCGTGGACTTCCAGCGCAAGATCCGCCAGAGCTACTTCGCCAGCGTGAGCT
ACCTGGACACCCAGGTGGGCCGCCTGCTGAGCGCCCTGGACGACCTGCAGCTGGCCAACAGCACCATCATCGCCTTC
ACCAGCGACCACGGCTGGGCCCTGGGCGAGCACGGCGAGTGGGCCAAGTACAGCAACTTCGACGTGGCCACCCACGT
GCCCCTGATCTTCTACGTGCCCGGCCGCACCGCCAGCCTGCCCGAGGCCGGCGAGAAGCTGTTCCCCTACCTGGACC
CCTTCGACAGCGCCAGCCAGCTGATGGAGCCCGGCCGCCAGAGCATGGACCTGGTGGAGCTGGTGAGCCTGTTCCCC
ACCCTGGCCGGCCTGGCCGGCCTGCAGGTGCCCCCCCGCTGCCCCGTGCCCAGCTTCCACGTGGAGCTGTGCCGCGA
GGGCAAGAACCTGCTGAAGCACTTCCGCTTCCGCGACCTGGAGGAGGACCCCTACCTGCCCGGCAACCCCCGCGAGC
TGATCGCCTACAGCCAGTACCCCCGCCCCAGCGACATCCCCCAGTGGAACAGCGACAAGCCCAGCCTGAAGGACATC
AAGATCATGGGCTACAGCATCCGCACCATCGACTACCGCTACACCGTGTGGGTGGGCTTCAACCCCGACGAGTTCCT
GGCCAACTTCAGCGACATCCACGCCGGCGAGCTGTACTTCGTGGACAGCGACCCCCTGCAGGACCACAACATGTACA
ACGACAGCCAGGGCGGCGACCTGTTCCAGCTGCTGATGCCCTAG

[0176] SEQ ID NO:8

[0177] ＞智人全长前体甲酰甘氨酸生成酶 (FGE)，mRNA

[0178] ATGGCTGCGCCCGCACTAGGGCTGGTGTGTGGACGTTGCCCTGAGCTGGGTCTCGTCCTCTTGCTGCTG

CTGCTCTCGCTGCTGTGTGGAGCGGCAGGGAGCCAGGAGGCCGGGACCGGTGCGGGCGCGGGGTCCCTTGCGGGTTC
TTGCGGCTGCGGCACGCCCCAGCGGCCTGGCGCCCATGGCAGTTCGGCAGCCGCTCACCGATACTCGCGGGAGGCTA
ACGCTCCGGGCCCCGTACCCGGAGAGCGGCAACTCGCGCACTCAAAGATGGTCCCCATCCCTGCTGGAGTATTTACA
ATGGGCACAGATGATCCTCAGATAAAGCAGGATGGGGAAGCACCTGCGAGGAGAGTTACTATTGATGCCTTTTACAT
GGATGCCTATGAAGTCAGTAATACTGAATTTGAGAAGTTTGTGAACTCAACTGGCTATTTGACAGAGGCTGAGAAGT
TTGGCGACTCCTTTGTCTTTGAAGGCATGTTGAGTGAGCAAGTGAAGACCAATATTCAACAGGCAGTTGCAGCTGCT
CCCTGGTGGTTACCTGTGAAAGGCGCTAACTGGAGACACCCAGAAGGGCCTGACTCTACTATTCTGCACAGGCCGGA
TCATCCAGTTCTCCATGTGTCCTGGAATGATGCGGTTGCCTACTGCACTTGGGCAGGGAAGCGGCTGCCCACGGAAG
CTGAGTGGGAATACAGCTGTCGAGGAGGC CTGCATAATAGACTTTTCCCCTGGGGCAACAAACTGCAGCCCAAAGG
CCAGCATTATGCCAACATTTGGCAGGGCGAGTTTCCGGTGACCAACACTGGTGAGGATGGCTTCCAAGGAACTGCGC
CTGTTGATGCCTTCCCTCCCAATGGTTATGGCTTATACAACATAGTGGGGAACGCATGGGAATGGACTTCAGACTGG
TGGACTGTTCATCATTCTGTTGAAGAAACGCTTAACCCAAAAGGTCCCCCTTCTGGGAAAGACCGAGTGAAGAAAGG
TGGATCCTACATGTGCCATAGGTCTTATTGTTACAGGTATCGCTGTGCTGCTCGGAGCCAGAACACACCTGATAGCT
CTGCTTCGAATCTGGGATTCCGCTGTGCAGCCGACCGCCTGCCCACCATGGACTGA

[0179] 通过人 CMV 启动子控制 I2S 和 FGE 编码核酸序列。I2S mRNA 的翻译导致 550 个氨基酸的全长 I2S 蛋白 (SEQ ID NO :2) 的合成，所述蛋白质包含 25 个氨基酸的信号肽。除去信号肽，并且从细胞分泌可溶性酶。

[0180] 将细菌新霉素磷酸转移酶 (neo) 编码序列和 / 或杀稻瘟菌素 S(Blasticidin S) 脱氨酶 (BSD) 基因用于使得能够分别使用新霉素类似物 G418 和 / 或杀稻瘟菌素选择转染的细胞。此外，在 I2S 和 / 或 FGE 编码载体上使用小鼠二氢叶酸还原酶 (DHFR) 基因以允许通过甲氨蝶呤 (MTX) 选择来分离包含增加的拷贝的 I2S 和 / 或 FGE 编码序列的细胞系。

[0181] 分离产生 I2S 的细胞，并使其经历适当的药物选择来分离具有增加的拷贝数的转染的 I2S 和 / 或 FGE 基因的细胞。通过 ELISA 进行 I2S 的定量。

[0182] 也使细胞群在甲氨蝶呤 (MTX) 中经历分步选择以分离具有增加的 I2S 生产率的细胞。在 MTX 选择期间通过 ELISA 监控 I2S 的生产率。

[0183] 经过多轮扩增后，随后通过从包含 10％小牛血清的 DMEM 逐步减少至无血清的化学成分确定的培养基，使几个产生 I2S 的克隆在无血清培养基中经历悬浮培养适应。通过有限稀释克隆建立几个单独的克隆群体。通过 I2S 酶活性测定和 ELISA 筛选菌落。两个稳定的细胞系 2D 和 4D 显示高百分比的活力和 I2S 的强劲表达，并且被选择用于进一步开发。

[0184] 实施例 2. 共表达 I2S 和 FGE 的稳定细胞系的评价

[0185] 进行额外的实验以表征共表达 I2S 和 FGE 的两个细胞系 2D 和 4D。

[0186] 比活性

[0187] 评价 I2S 酶的第一比活性。使用基于荧光的 4-MUF 测定分析从 2D 和 4D 细胞系产生的 I2S 酶的比活性。简言之，该测定测量了 I2S 底物 4- 甲基伞形酮基硫酸酯 (4-MUF-SO_4) 的水解。在 4-MUF-SO_4底物被 I2S 切割后，该分子被转化成硫酸酯和天然发荧光的 4- 甲基伞形酮 (4-MUF)。作为结果，I2S 酶活性可以通过评价随时间的荧光信号的整体变化来测定。对于本实验，使从 I2S-AF 2D 和 4D 人细胞系产生的纯化的 I2S 酶与 4- 甲基伞形酮基硫酸酯 (4-MUF-SO_4) 钾盐 (Sigma 目录号 M-7133) 温育。使用以 1:100、1:200 和 1:20,000 的原液稀释的商购可得的 I2S 酶，利用一系列对照参照样品进行测定的校准。在 37℃运行

酶测定，并使用校准的荧光计进行测定。使用对于每个参考标准获得的荧光值，使用下列公式确定变异系数百分比：

[0188]

$$\%CV = \frac{\text{原始荧光值的标准差}(N=3)}{\text{平均荧光值}} X\,100\%$$

[0189] 随后使用下列公式将百分比 CV 值用于计算每一个样品的修正平均荧光，以确定以 mU/mL 表示的报告酶的活性：

[0190]

$$mU/mL = (CFU)\left(\frac{1\text{纳摩尔}/L}{10\,FU}\right)\left(\frac{1L}{10^3 mL}\right)\left(\frac{2.11mL}{0.01mL}\right)\left(\frac{1\text{小时}}{60\text{分钟}}\right)\left(\frac{1mU}{\text{纳摩尔}}\right)(DF)$$

[0191] CFU ＝负修正平均荧光

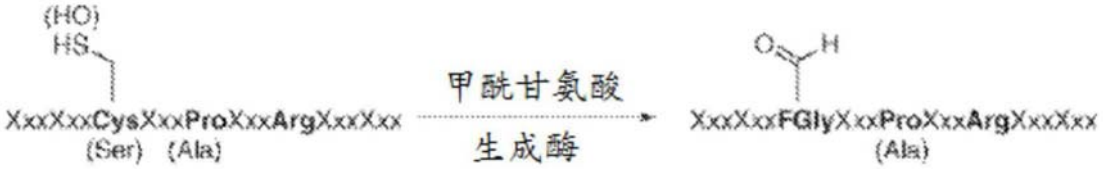
[0192] DF - 稀释因子

[0193] 1 毫单位的活性是在 37℃于 1 分钟内将 1 纳摩尔的 4- 甲基伞形酮基硫酸酯转化为 4- 甲基伞形酮所需的酶的量。

[0194] 甲酰甘氨酸转化百分比

[0195] 肽图谱可用于确定 FGly 转化百分比。I2S 活化需要通过甲酰甘氨酸生成酶 (FGE) 进行的半胱氨酸（对应于成熟人 I2S 的位置 59）至甲酰甘氨酸的转化，如下文中显示的：

[0196]



[0197] 因此，甲酰甘氨酸转化的百分比（% FG）可以使用下列公式计算：

[0198]

$$\%FG\,(DS的) = \frac{\text{活性I2S分子的数目}}{\text{总的(活性+无活性)I2S分子的数目}} \times 100$$

[0199] 例如，50％ FG 意指一半纯化的重组 I2S 无酶活性而没有任何治疗效果。

[0200] 肽图谱用于计算％ FG。简言之，使用蛋白酶（例如，胰蛋白酶或糜蛋白酶）将重组 I2S 蛋白消化成短肽。使用 HPLC 分离和表征短肽。表征包含对应于成熟人 I2S 的位置 59 的位置的肽以确定相较于对照（例如，无 FGly 转化的 I2S 蛋白或具有 100％ FGly 转化的 I2S 蛋白）位置 59 上的 Cys 是否被转化成 FGly。可基于对应的峰面积测定包含 FGly 的肽的量（对应于活性 I2S 分子的数目）和具有 FGly 和 Cys 的肽的总量（对应于总的 I2S 分子的数目），并计算反映％ FG 的比率。

[0201] FGly 转化百分比与比活性之间的相关性

[0202] FGly 转化百分比与比活性之间的示例性相关性示于图 3 中。如可以看到的，该数据表明更高百分比的甲酰甘氨酸转化导致更高的 I2S 酶活性。

[0203] 聚糖图谱

[0204] 测定由细胞系 2D 和 4D 产生的重组 I2S 蛋白的聚糖组成。使用阴离子交换色谱法进行聚糖组成的定量。如下所述，在这些条件下产生的重组 I2S 的聚糖图谱由七个峰群组成，根据递增量的负电荷（至少部分地因酶消化而从唾液酸和甘露糖 -6- 磷酸糖型产生）洗脱。简言之，使用如下方式处理利用无血清细胞培养方法（I2S-AF 2D 无血清和 I2S-AF 4D 无血清）获得的纯化的重组 I2S 和产生的参照重组 I2S：(1) 使用纯化的神经氨酸酶（分离自产脲节杆菌 (Arthrobacter Ureafaciens)(10mU/μL)，Roche Biochemical(Indianapolis,IN)，目录号 269611(1U/100μL)）以除去唾液酸残留物，(2) 在 37±1℃用碱性磷酸酶处理 2 小时以完全释放甘露糖 -6- 磷酸残留物，(3) 碱性磷酸酶 + 神经氨酸酶或 (4) 无处理。使用配备有 Dionex CarboPac PA1 保护柱的 CarboPac PA1 分析柱，通过利用脉冲安培检测 (HPAE-PAD) 的高效阴离子交换色谱法分析每一种酶消化。对于每一种测定，运行一系列在 0.4 至 2.0 纳摩尔的范围内的唾液酸和甘露糖 -6- 磷酸标准。在环境柱温度下将使用在 100mM 氢氧化钠中的 48mM 乙酸钠的等度方法以 1.0mL/ 分钟的流速运行至少 15 分钟，以洗脱每个峰。将从每个单独的运行产生的 I2S-AF 和参照 I2S 样品的数据分别组合成单个色谱来表示每个相应的重组蛋白的聚糖图谱。如图 4 中显示的，通过细胞系 2D 和 4D 产生的 I2S 的示例性聚糖图谱显示构成中性、单、双唾液酸化、单磷酸化、三唾液酸化和混合（单唾液酸化和封端的甘露糖 -6- 磷酸）、四唾液酸化和混合（双唾液酸化和封端的甘露糖 -6- 磷酸）以及二磷酸化聚糖的代表性洗脱峰（按洗脱顺序）。

[0205] 实施例 3. 无血清悬浮细胞培养

[0206] 本实施例显示大规模无血清悬浮培养可被开发来培养最优化的细胞系以产生重组 I2S。

[0207] 无血清悬浮细胞培养系统

[0208] 简言之，使用实施例 1 的 2D 或 4D 细胞系建立种子培养物。将细胞转移至含有补充了 MTX（以用于选择）的无血清化学成分确定的扩增培养基的 250mL 组织培养摇瓶中，用碳酸氢钠调节至 pH7.3 并在 37℃、5% CO_2下生长数天。在培养物达到足够的细胞密度和活力后，将初始种子培养物用于将一系列逐步细胞培养扩增的第一扩增接种在 500mL 组织培养摇瓶中，随后接种于 1L 组织培养摇瓶中。

[0209] 通过将每一个 1L 的培养物转移到 10L Cellbag bioreactor® (Wave Europe)，并添加扩增培养基来进行分批培养扩增。在达到足够的细胞密度后，添加新的扩增培养基，并使细胞生长至足够的密度。将 10L Cellbag 转移到 Wave bioreactor® 系统 (Wave Europe)，并改变培养条件以允许在连续培养基灌注的条件下生长。递送扩增生长培养基，并收集样品以进行 pH、谷氨酰胺、谷氨酸、葡萄糖、铵、乳酸盐、pCO_2和克分子渗透压浓度的离线代谢物分析。

[0210] 在达到足够的细胞密度后，将整个 10L 的细胞培养物转移至包含新鲜的扩增培养基的 50L Wave Cellbag bioreactor®，并使用 Wave bioreactor® 系统使其生长至足够的细胞密度。

[0211] 随后使用 200L 的一次性生物反应器和离心灌注装置（Centritech® CELL II 单位，Pneumatic Scale Corporation) 进行细胞扩增，所述装置被设计来浓缩细胞和澄清培养基，以在灌注介导的细胞培养过程中再循环。用 50L 培养物的一部分接种扩增培养基（调

整至 pH 7.10)，并使其生长至足够的细胞密度。

[0212] 随后，将 200L 培养物的一部分用于接种在生产培养基（被调整至 pH 7.20）中的 2000L 一次性生物反应器和离心机灌注装置（Centritech® CELL II 单位，Pneumatic Scale Corporation）中。使细胞在分批生长条件下生长。在为期 2 天的生长后，调整条件以进行连续灌注，直至到达过渡期。使细胞在灌注生长条件下生长持续 24 小时的过渡期。

[0213] 对于生产期，使用 2 个 Centritech CELL II 单位。在过渡期开始后约 24 小时开始生产期，并且通过调节泄放速率来维持期望的时间。

[0214] 虽然本文描述的某些化合物、组合物和方法已经根据某些实施方案进行特定的描述，但是下列实施例仅用于说明本发明的化合物，并且无意对其进行限制。

[0215] 如本文在说明书中和权利要求书中使用的冠词“一个 (a)”和“一种 (an)”，除非清楚地表明与之相反，否则应该被理解为包括其复数个指示物。在某组的一个或多个成员之间包括“或者”的权利要求或者描述，被认为：如果一个、多于一个或所有组成员存在、被采用或者以其它方式与给定的产品或过程相关，那么就是符合的；除非指示相反或者以其它方式从上下文中明显可见。本发明包括这样的实施方案：其中所述组的只有一个成员存在、被采用或者以其它方式与给定的产品或过程相关。本发明也包括这样的实施方案：其中多于一个或者整个组成员存在、被采用或者以其他方式与给定的产品或过程相关。而且，应该理解，本发明涵盖来自一条或多条所列出的权利要求所有的变型、组合以及排列，其中一种或多种限制、元素、条款、描述性术语等根据所述相同基础权利要求（或者相关、任意其它权利要求）被引入到另一权利要求，除非以其它方式指出或者除非其对于本领域技术人员将明显地出现矛盾或者不一致。当元素以列表形式呈现（例如，在 Markush 组中或者类似的形式）时，应该理解，所述元素的每个亚组也被公开，并且任何元素可以从所述组删除。应该理解，一般而言，当本发明、或者本发明的某些方面，被指出包括特定的元素、特征等时，那么本发明或者本发明的方面的某些实施方案由这些元素、特征等组成或基本由其组成。为了鉴定起见，本文没有在每种情形中用过多的词汇明确地陈述那些实施方案。还应该理解，本发明的任何实施方案或者方面可以被明确地从所述权利要求排除，无论所述具体的排除是否在本说明书中叙述。本文所引用的出版物、网址和其它参考材料，为了描述本发明的背景以及提供关于其实践的其它细节，都通过引用并入本文。

[0001]

序列表

<110> F•博尔戈
M•哈特莱因

<120> 用于产生重组艾杜糖-2-硫酸酯酶的细胞

<130> 2006685-0340

<150> 61/666,719
<151> 2012-06-29

<160> 8

<170> PatentIn 3.5版

<210> 1
<211> 525
<212> PRT
<213> 智人

<400> 1

Ser Glu Thr Gln Ala Asn Ser Thr Thr Asp Ala Leu Asn Val Leu Leu
1 5 10 15

Ile Ile Val Asp Asp Leu Arg Pro Ser Leu Gly Cys Tyr Gly Asp Lys
20 25 30

Leu Val Arg Ser Pro Asn Ile Asp Gln Leu Ala Ser His Ser Leu Leu
35 40 45

Phe Gln Asn Ala Phe Ala Gln Gln Ala Val Cys Ala Pro Ser Arg Val
50 55 60

Ser Phe Leu Thr Gly Arg Arg Pro Asp Thr Thr Arg Leu Tyr Asp Phe
65 70 75 80

Asn Ser Tyr Trp Arg Val His Ala Gly Asn Phe Ser Thr Ile Pro Gln
85 90 95

Tyr Phe Lys Glu Asn Gly Tyr Val Thr Met Ser Val Gly Lys Val Phe
100 105 110

His Pro Gly Ile Ser Ser Asn His Thr Asp Asp Ser Pro Tyr Ser Trp
115 120 125

Ser Phe Pro Pro Tyr His Pro Ser Ser Glu Lys Tyr Glu Asn Thr Lys

[0002]

130 135 140

Thr Cys Arg Gly Pro Asp Gly Glu Leu His Ala Asn Leu Leu Cys Pro
145 150 155 160

Val Asp Val Leu Asp Val Pro Glu Gly Thr Leu Pro Asp Lys Gln Ser
165 170 175

Thr Glu Gln Ala Ile Gln Leu Leu Glu Lys Met Lys Thr Ser Ala Ser
180 185 190

Pro Phe Phe Leu Ala Val Gly Tyr His Lys Pro His Ile Pro Phe Arg
195 200 205

Tyr Pro Lys Glu Phe Gln Lys Leu Tyr Pro Leu Glu Asn Ile Thr Leu
210 215 220

Ala Pro Asp Pro Glu Val Pro Asp Gly Leu Pro Pro Val Ala Tyr Asn
225 230 235 240

Pro Trp Met Asp Ile Arg Gln Arg Glu Asp Val Gln Ala Leu Asn Ile
245 250 255

Ser Val Pro Tyr Gly Pro Ile Pro Val Asp Phe Gln Arg Lys Ile Arg
260 265 270

Gln Ser Tyr Phe Ala Ser Val Ser Tyr Leu Asp Thr Gln Val Gly Arg
275 280 285

Leu Leu Ser Ala Leu Asp Asp Leu Gln Leu Ala Asn Ser Thr Ile Ile
290 295 300

Ala Phe Thr Ser Asp His Gly Trp Ala Leu Gly Glu His Gly Glu Trp
305 310 315 320

Ala Lys Tyr Ser Asn Phe Asp Val Ala Thr His Val Pro Leu Ile Phe
325 330 335

Tyr Val Pro Gly Arg Thr Ala Ser Leu Pro Glu Ala Gly Glu Lys Leu
340 345 350

Phe Pro Tyr Leu Asp Pro Phe Asp Ser Ala Ser Gln Leu Met Glu Pro

[0003]

355 360 365

Gly Arg Gln Ser Met Asp Leu Val Glu Leu Val Ser Leu Phe Pro Thr
370 375 380

Leu Ala Gly Leu Ala Gly Leu Gln Val Pro Pro Arg Cys Pro Val Pro
385 390 395 400

Ser Phe His Val Glu Leu Cys Arg Glu Gly Lys Asn Leu Leu Lys His
405 410 415

Phe Arg Phe Arg Asp Leu Glu Glu Asp Pro Tyr Leu Pro Gly Asn Pro
420 425 430

Arg Glu Leu Ile Ala Tyr Ser Gln Tyr Pro Arg Pro Ser Asp Ile Pro
435 440 445

Gln Trp Asn Ser Asp Lys Pro Ser Leu Lys Asp Ile Lys Ile Met Gly
450 455 460

Tyr Ser Ile Arg Thr Ile Asp Tyr Arg Tyr Thr Val Trp Val Gly Phe
465 470 475 480

Asn Pro Asp Glu Phe Leu Ala Asn Phe Ser Asp Ile His Ala Gly Glu
485 490 495

Leu Tyr Phe Val Asp Ser Asp Pro Leu Gln Asp His Asn Met Tyr Asn
500 505 510

Asp Ser Gln Gly Gly Asp Leu Phe Gln Leu Leu Met Pro
515 520 525

<210> 2
<211> 550
<212> PRT
<213> 智人

<400> 2

Met Pro Pro Pro Arg Thr Gly Arg Gly Leu Leu Trp Leu Gly Leu Val
1 5 10 15

Leu Ser Ser Val Cys Val Ala Leu Gly Ser Glu Thr Gln Ala Asn Ser
20 25 30

[0004]

```
Thr Thr Asp Ala Leu Asn Val Leu Leu Ile Ile Val Asp Asp Leu Arg
        35                  40                  45

Pro Ser Leu Gly Cys Tyr Gly Asp Lys Leu Val Arg Ser Pro Asn Ile
    50                  55                  60

Asp Gln Leu Ala Ser His Ser Leu Leu Phe Gln Asn Ala Phe Ala Gln
65                  70                  75                  80

Gln Ala Val Cys Ala Pro Ser Arg Val Ser Phe Leu Thr Gly Arg Arg
                85                  90                  95

Pro Asp Thr Thr Arg Leu Tyr Asp Phe Asn Ser Tyr Trp Arg Val His
            100                 105                 110

Ala Gly Asn Phe Ser Thr Ile Pro Gln Tyr Phe Lys Glu Asn Gly Tyr
        115                 120                 125

Val Thr Met Ser Val Gly Lys Val Phe His Pro Gly Ile Ser Ser Asn
    130                 135                 140

His Thr Asp Asp Ser Pro Tyr Ser Trp Ser Phe Pro Pro Tyr His Pro
145                 150                 155                 160

Ser Ser Glu Lys Tyr Glu Asn Thr Lys Thr Cys Arg Gly Pro Asp Gly
                165                 170                 175

Glu Leu His Ala Asn Leu Leu Cys Pro Val Asp Val Leu Asp Val Pro
            180                 185                 190

Glu Gly Thr Leu Pro Asp Lys Gln Ser Thr Glu Gln Ala Ile Gln Leu
        195                 200                 205

Leu Glu Lys Met Lys Thr Ser Ala Ser Pro Phe Phe Leu Ala Val Gly
    210                 215                 220

Tyr His Lys Pro His Ile Pro Phe Arg Tyr Pro Lys Glu Phe Gln Lys
225                 230                 235                 240

Leu Tyr Pro Leu Glu Asn Ile Thr Leu Ala Pro Asp Pro Glu Val Pro
                245                 250                 255
```

[0005]

Asp Gly Leu Pro Pro Val Ala Tyr Asn Pro Trp Met Asp Ile Arg Gln
260 265 270

Arg Glu Asp Val Gln Ala Leu Asn Ile Ser Val Pro Tyr Gly Pro Ile
275 280 285

Pro Val Asp Phe Gln Arg Lys Ile Arg Gln Ser Tyr Phe Ala Ser Val
290 295 300

Ser Tyr Leu Asp Thr Gln Val Gly Arg Leu Leu Ser Ala Leu Asp Asp
305 310 315 320

Leu Gln Leu Ala Asn Ser Thr Ile Ile Ala Phe Thr Ser Asp His Gly
325 330 335

Trp Ala Leu Gly Glu His Gly Glu Trp Ala Lys Tyr Ser Asn Phe Asp
340 345 350

Val Ala Thr His Val Pro Leu Ile Phe Tyr Val Pro Gly Arg Thr Ala
355 360 365

Ser Leu Pro Glu Ala Gly Glu Lys Leu Phe Pro Tyr Leu Asp Pro Phe
370 375 380

Asp Ser Ala Ser Gln Leu Met Glu Pro Gly Arg Gln Ser Met Asp Leu
385 390 395 400

Val Glu Leu Val Ser Leu Phe Pro Thr Leu Ala Gly Leu Ala Gly Leu
405 410 415

Gln Val Pro Pro Arg Cys Pro Val Pro Ser Phe His Val Glu Leu Cys
420 425 430

Arg Glu Gly Lys Asn Leu Leu Lys His Phe Arg Phe Arg Asp Leu Glu
435 440 445

Glu Asp Pro Tyr Leu Pro Gly Asn Pro Arg Glu Leu Ile Ala Tyr Ser
450 455 460

Gln Tyr Pro Arg Pro Ser Asp Ile Pro Gln Trp Asn Ser Asp Lys Pro
465 470 475 480

[0006]

Ser Leu Lys Asp Ile Lys Ile Met Gly Tyr Ser Ile Arg Thr Ile Asp
485 490 495

Tyr Arg Tyr Thr Val Trp Val Gly Phe Asn Pro Asp Glu Phe Leu Ala
500 505 510

Asn Phe Ser Asp Ile His Ala Gly Glu Leu Tyr Phe Val Asp Ser Asp
515 520 525

Pro Leu Gln Asp His Asn Met Tyr Asn Asp Ser Gln Gly Gly Asp Leu
530 535 540

Phe Gln Leu Leu Met Pro
545 550

<210> 3
<211> 312
<212> PRT
<213> 智人

<400> 3

Met Pro Pro Pro Arg Thr Gly Arg Gly Leu Leu Trp Leu Gly Leu Val
1 5 10 15

Leu Ser Ser Val Cys Val Ala Leu Gly Ser Glu Thr Gln Ala Asn Ser
20 25 30

Thr Thr Asp Ala Leu Asn Val Leu Leu Ile Ile Val Asp Asp Leu Arg
35 40 45

Pro Ser Leu Gly Cys Tyr Gly Asp Lys Leu Val Arg Ser Pro Asn Ile
50 55 60

Asp Gln Leu Ala Ser His Ser Leu Leu Phe Gln Asn Ala Phe Ala Gln
65 70 75 80

Gln Ala Val Cys Ala Pro Ser Arg Val Ser Phe Leu Thr Gly Arg Arg
85 90 95

Pro Asp Thr Thr Arg Leu Tyr Asp Phe Asn Ser Tyr Trp Arg Val His
100 105 110

[0007]

```
Ala Gly Asn Phe Ser Thr Ile Pro Gln Tyr Phe Lys Glu Asn Gly Tyr
        115                 120                 125

Val Thr Met Ser Val Gly Lys Val Phe His Pro Gly Ile Ser Ser Asn
    130                 135                 140

His Thr Asp Asp Ser Pro Tyr Ser Trp Ser Phe Pro Pro Tyr His Pro
145                 150                 155                 160

Ser Ser Glu Lys Tyr Glu Asn Thr Lys Thr Cys Arg Gly Pro Asp Gly
                165                 170                 175

Glu Leu His Ala Asn Leu Leu Cys Pro Val Asp Val Leu Asp Val Pro
            180                 185                 190

Glu Gly Thr Leu Pro Asp Lys Gln Ser Thr Glu Gln Ala Ile Gln Leu
        195                 200                 205

Leu Glu Lys Met Lys Thr Ser Ala Ser Pro Phe Phe Leu Ala Val Gly
    210                 215                 220

Tyr His Lys Pro His Ile Pro Phe Arg Tyr Pro Lys Glu Phe Gln Lys
225                 230                 235                 240

Leu Tyr Pro Leu Glu Asn Ile Thr Leu Ala Pro Asp Pro Glu Val Pro
                245                 250                 255

Asp Gly Leu Pro Pro Val Ala Tyr Asn Pro Trp Met Asp Ile Arg Gln
            260                 265                 270

Arg Glu Asp Val Gln Ala Leu Asn Ile Ser Val Pro Tyr Gly Pro Ile
        275                 280                 285

Pro Val Asp Phe Gln Glu Asp Gln Ser Ser Thr Gly Phe Arg Leu Lys
    290                 295                 300

Thr Ser Ser Thr Arg Lys Tyr Lys
305                 310
```

<210> 4
<211> 343
<212> PRT

[0008]

<213> 智人

<400> 4

Met Pro Pro Pro Arg Thr Gly Arg Gly Leu Leu Trp Leu Gly Leu Val
1 5 10 15

Leu Ser Ser Val Cys Val Ala Leu Gly Ser Glu Thr Gln Ala Asn Ser
20 25 30

Thr Thr Asp Ala Leu Asn Val Leu Leu Ile Ile Val Asp Asp Leu Arg
35 40 45

Pro Ser Leu Gly Cys Tyr Gly Asp Lys Leu Val Arg Ser Pro Asn Ile
50 55 60

Asp Gln Leu Ala Ser His Ser Leu Leu Phe Gln Asn Ala Phe Ala Gln
65 70 75 80

Gln Ala Val Cys Ala Pro Ser Arg Val Ser Phe Leu Thr Gly Arg Arg
85 90 95

Pro Asp Thr Thr Arg Leu Tyr Asp Phe Asn Ser Tyr Trp Arg Val His
100 105 110

Ala Gly Asn Phe Ser Thr Ile Pro Gln Tyr Phe Lys Glu Asn Gly Tyr
115 120 125

Val Thr Met Ser Val Gly Lys Val Phe His Pro Gly Ile Ser Ser Asn
130 135 140

His Thr Asp Asp Ser Pro Tyr Ser Trp Ser Phe Pro Pro Tyr His Pro
145 150 155 160

Ser Ser Glu Lys Tyr Glu Asn Thr Lys Thr Cys Arg Gly Pro Asp Gly
165 170 175

Glu Leu His Ala Asn Leu Leu Cys Pro Val Asp Val Leu Asp Val Pro
180 185 190

Glu Gly Thr Leu Pro Asp Lys Gln Ser Thr Glu Gln Ala Ile Gln Leu
195 200 205

[0009]

```
Leu Glu Lys Met Lys Thr Ser Ala Ser Pro Phe Phe Leu Ala Val Gly
    210                 215                 220

Tyr His Lys Pro His Ile Pro Phe Arg Tyr Pro Lys Glu Phe Gln Lys
225                 230                 235                 240

Leu Tyr Pro Leu Glu Asn Ile Thr Leu Ala Pro Asp Pro Glu Val Pro
                245                 250                 255

Asp Gly Leu Pro Pro Val Ala Tyr Asn Pro Trp Met Asp Ile Arg Gln
            260                 265                 270

Arg Glu Asp Val Gln Ala Leu Asn Ile Ser Val Pro Tyr Gly Pro Ile
        275                 280                 285

Pro Val Asp Phe Gln Arg Lys Ile Arg Gln Ser Tyr Phe Ala Ser Val
    290                 295                 300

Ser Tyr Leu Asp Thr Gln Val Gly Arg Leu Leu Ser Ala Leu Asp Asp
305                 310                 315                 320

Leu Gln Leu Ala Asn Ser Thr Ile Ile Ala Phe Thr Ser Asp His Gly
                325                 330                 335

Phe Leu Met Arg Thr Asn Thr
            340
```

<210> 5
<211> 341
<212> PRT
<213> 智人

<400> 5

```
Ser Gln Glu Ala Gly Thr Gly Ala Gly Ala Gly Ser Leu Ala Gly Ser
1               5                   10                  15

Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly Ala His Gly Ser Ser Ala
            20                  25                  30

Ala Ala His Arg Tyr Ser Arg Glu Ala Asn Ala Pro Gly Pro Val Pro
        35                  40                  45

Gly Glu Arg Gln Leu Ala His Ser Lys Met Val Pro Ile Pro Ala Gly
```

[0010]

50 55 60

Val Phe Thr Met Gly Thr Asp Asp Pro Gln Ile Lys Gln Asp Gly Glu
65 70 75 80

Ala Pro Ala Arg Arg Val Thr Ile Asp Ala Phe Tyr Met Asp Ala Tyr
85 90 95

Glu Val Ser Asn Thr Glu Phe Glu Lys Phe Val Asn Ser Thr Gly Tyr
100 105 110

Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser Phe Val Phe Glu Gly Met
115 120 125

Leu Ser Glu Gln Val Lys Thr Asn Ile Gln Gln Ala Val Ala Ala Ala
130 135 140

Pro Trp Trp Leu Pro Val Lys Gly Ala Asn Trp Arg His Pro Glu Gly
145 150 155 160

Pro Asp Ser Thr Ile Leu His Arg Pro Asp His Pro Val Leu His Val
165 170 175

Ser Trp Asn Asp Ala Val Ala Tyr Cys Thr Trp Ala Gly Lys Arg Leu
180 185 190

Pro Thr Glu Ala Glu Trp Glu Tyr Ser Cys Arg Gly Gly Leu His Asn
195 200 205

Arg Leu Phe Pro Trp Gly Asn Lys Leu Gln Pro Lys Gly Gln His Tyr
210 215 220

Ala Asn Ile Trp Gln Gly Glu Phe Pro Val Thr Asn Thr Gly Glu Asp
225 230 235 240

Gly Phe Gln Gly Thr Ala Pro Val Asp Ala Phe Pro Pro Asn Gly Tyr
245 250 255

Gly Leu Tyr Asn Ile Val Gly Asn Ala Trp Glu Trp Thr Ser Asp Trp
260 265 270

Trp Thr Val His His Ser Val Glu Glu Thr Leu Asn Pro Lys Gly Pro

[0011]

275 280 285

Pro Ser Gly Lys Asp Arg Val Lys Lys Gly Gly Ser Tyr Met Cys His
290 295 300

Arg Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala Ala Arg Ser Gln Asn Thr
305 310 315 320

Pro Asp Ser Ser Ala Ser Asn Leu Gly Phe Arg Cys Ala Ala Asp Arg
325 330 335

Leu Pro Thr Met Asp
340

<210> 6
<211> 374
<212> PRT
<213> 智人

<400> 6

Met Ala Ala Pro Ala Leu Gly Leu Val Cys Gly Arg Cys Pro Glu Leu
1 5 10 15

Gly Leu Val Leu Leu Leu Leu Leu Leu Ser Leu Leu Cys Gly Ala Ala
20 25 30

Gly Ser Gln Glu Ala Gly Thr Gly Ala Gly Ala Gly Ser Leu Ala Gly
35 40 45

Ser Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly Ala His Gly Ser Ser
50 55 60

Ala Ala Ala His Arg Tyr Ser Arg Glu Ala Asn Ala Pro Gly Pro Val
65 70 75 80

Pro Gly Glu Arg Gln Leu Ala His Ser Lys Met Val Pro Ile Pro Ala
85 90 95

Gly Val Phe Thr Met Gly Thr Asp Asp Pro Gln Ile Lys Gln Asp Gly
100 105 110

Glu Ala Pro Ala Arg Arg Val Thr Ile Asp Ala Phe Tyr Met Asp Ala
115 120 125

[0012]

```
Tyr Glu Val Ser Asn Thr Glu Phe Glu Lys Phe Val Asn Ser Thr Gly
    130                 135                 140

Tyr Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser Phe Val Phe Glu Gly
145                 150                 155                 160

Met Leu Ser Glu Gln Val Lys Thr Asn Ile Gln Gln Ala Val Ala Ala
                165                 170                 175

Ala Pro Trp Trp Leu Pro Val Lys Gly Ala Asn Trp Arg His Pro Glu
            180                 185                 190

Gly Pro Asp Ser Thr Ile Leu His Arg Pro Asp His Pro Val Leu His
        195                 200                 205

Val Ser Trp Asn Asp Ala Val Ala Tyr Cys Thr Trp Ala Gly Lys Arg
    210                 215                 220

Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ser Cys Arg Gly Gly Leu His
225                 230                 235                 240

Asn Arg Leu Phe Pro Trp Gly Asn Lys Leu Gln Pro Lys Gly Gln His
                245                 250                 255

Tyr Ala Asn Ile Trp Gln Gly Glu Phe Pro Val Thr Asn Thr Gly Glu
            260                 265                 270

Asp Gly Phe Gln Gly Thr Ala Pro Val Asp Ala Phe Pro Pro Asn Gly
        275                 280                 285

Tyr Gly Leu Tyr Asn Ile Val Gly Asn Ala Trp Glu Trp Thr Ser Asp
    290                 295                 300

Trp Trp Thr Val His His Ser Val Glu Glu Thr Leu Asn Pro Lys Gly
305                 310                 315                 320

Pro Pro Ser Gly Lys Asp Arg Val Lys Lys Gly Gly Ser Tyr Met Cys
                325                 330                 335

His Arg Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala Ala Arg Ser Gln Asn
            340                 345                 350
```

[0013]

Thr Pro Asp Ser Ser Ala Ser Asn Leu Gly Phe Arg Cys Ala Ala Asp
355 360 365

Arg Leu Pro Thr Met Asp
370

<210> 7
<211> 1653
<212> DNA
<213> 智人

<400> 7
atgcccccgc cccgcaccgg ccgcggcctg ctgtggctgg gcctggtgct gagcagcgtg 60

tgcgtggccc tgggcagcga gacccaggcc aacagcacca ccgacgccct gaacgtgctg 120

ctgatcatcg tggacgacct gcgccccagc ctgggctgct acggcgacaa gctggtgcgc 180

agccccaaca tcgaccagct ggccagccac agcctgctgt tccagaacgc cttcgcccag 240

caggccgtgt gcgcccccag ccgcgtgagc ttcctgaccg gccgccgccc cgacaccacc 300

cgcctgtacg acttcaacag ctactggcgc gtgcacgccg gcaacttcag caccatcccc 360

cagtacttca aggagaacgg ctacgtgacc atgagcgtgg gcaaggtgtt ccaccccggc 420

atcagcagca accacaccga cgacagcccc tacagctgga gcttcccccc ctaccacccc 480

agcagcgaga agtacgagaa caccaagacc tgccgcggcc ccgacggcga gctgcacgcc 540

aacctgctgt gccccgtgga cgtgctggac gtgcccgagg gcaccctgcc cgacaagcag 600

agcaccgagc aggccatcca gctgctggag aagatgaaga ccagcgccag ccccttcttc 660

ctggccgtgg gctaccacaa gccccacatc cccttccgct accccaagga gttccagaag 720

ctgtaccccc tggagaacat caccctggcc cccgaccccg aggtgcccga cggcctgccc 780

cccgtggcct acaaccccctg gatggacatc cgccagcgcg aggacgtgca ggccctgaac 840

atcagcgtgc cctacggccc catccccgtg gacttccagc gcaagatccg ccagagctac 900

ttcgccagcg tgagctacct ggacacccag gtgggccgcc tgctgagcgc cctggacgac 960

ctgcagctgg ccaacagcac catcatcgcc ttcaccagcg accacggctg ggccctgggc 1020

gagcacggcg agtgggccaa gtacagcaac ttcgacgtgg ccacccacgt gcccctgatc 1080

ttctacgtgc ccggccgcac cgccagcctg cccgaggccg gcgagaagct gttccctac 1140

ctggacccct tcgacagcgc cagccagctg atggagcccg gccgccagag catggacctg 1200

[0014]

gtggagctgg tgagcctgtt ccccaccctg gccggcctgg ccggcctgca ggtgcccccc 1260

cgctgccccg tgcccagctt ccacgtggag ctgtgccgcg agggcaagaa cctgctgaag 1320

cacttccgct tccgcgacct ggaggaggac ccctacctgc ccggcaaccc ccgcgagctg 1380

atcgcctaca gccagtaccc ccgccccagc gacatccccc agtggaacag cgacaagccc 1440

agcctgaagg acatcaagat catgggctac agcatccgca ccatcgacta ccgctacacc 1500

gtgtgggtgg gcttcaaccc cgacgagttc ctggccaact tcagcgacat ccacgccggc 1560

gagctgtact tcgtggacag cgacccccctg caggaccaca acatgtacaa cgacagccag 1620

ggcggcgacc tgttccagct gctgatgccc tag 1653

<210> 8
<211> 1125
<212> DNA
<213> 智人

<400> 8
atggctgcgc ccgcactagg gctggtgtgt ggacgttgcc ctgagctggg tctcgtcctc 60

ttgctgctgc tgctctcgct gctgtgtgga gcggcaggga gccaggaggc cgggaccggt 120

gcgggcgcgg ggtcccttgc gggttcttgc ggctgcggca cgccccagcg gcctggcgcc 180

catggcagtt cggcagccgc tcaccgatac tcgcgggagg ctaacgctcc gggccccgta 240

cccggagagc ggcaactcgc gcactcaaag atggtcccca tccctgctgg agtatttaca 300

atgggcacag atgatcctca gataaagcag gatggggaag cacctgcgag gagagttact 360

attgatgcct tttacatgga tgcctatgaa gtcagtaata ctgaatttga gaagtttgtg 420

aactcaactg gctatttgac agaggctgag aagtttggcg actcctttgt ctttgaaggc 480

atgttgagtg agcaagtgaa gaccaatatt caacaggcag ttgcagctgc tccctggtgg 540

ttacctgtga aaggcgctaa ctggagacac ccagaagggc ctgactctac tattctgcac 600

aggccggatc atccagttct ccatgtgtcc tggaatgatg cggttgccta ctgcacttgg 660

gcagggaagc ggctgcccac ggaagctgag tgggaataca gctgtcgagg aggcctgcat 720

aatagacttt tcccctgggg caacaaactg cagcccaaag gccagcatta tgccaacatt 780

tggcagggcg agtttccggt gaccaacact ggtgaggatg gcttccaagg aactgcgcct 840

gttgatgcct tccctcccaa tggttatggc ttatacaaca tagtggggaa cgcatgggaa 900

tggacttcag actggtggac tgttcatcat tctgttgaag aaacgcttaa cccaaaaggt 960

ccccccttctg ggaaagaccg agtgaagaaa ggtggatcct acatgtgcca taggtcttat 1020

[0015]

tgttacaggt atcgctgtgc tgctcggagc cagaacacac ctgatagctc tgcttcgaat 1080

ctgggattcc gctgtgcagc cgaccgcctg cccaccatgg actga 1125

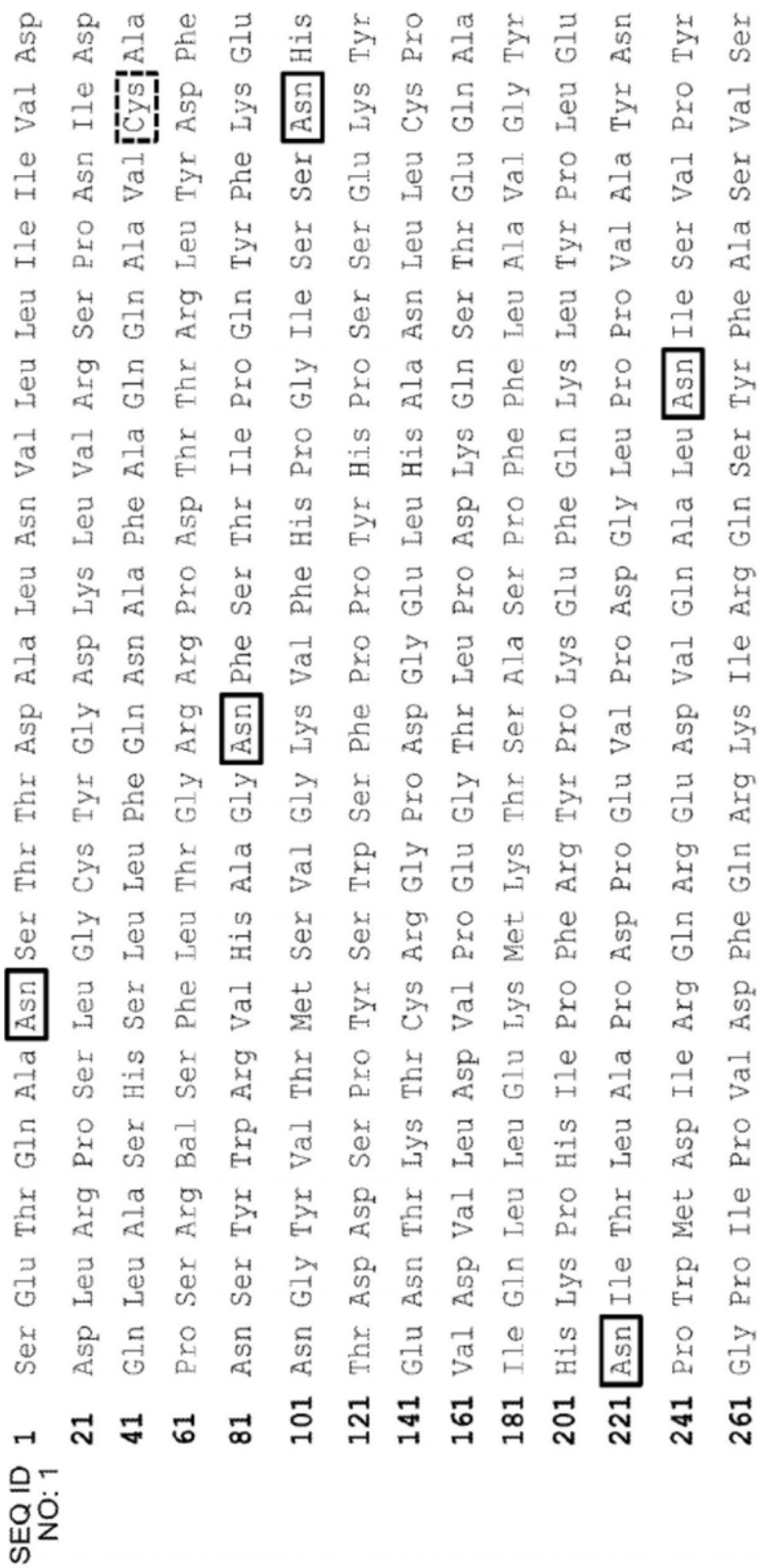

SEQ ID NO: 1
1 Ser Glu Thr Gln Ala Asn Ser Thr Thr Asp Ala Leu Asn Val Leu Leu Ile Ile Val Asp
21 Asp Leu Arg Pro Ser Leu Gly Cys Tyr Gly Asp Lys Leu Val Arg Ser Pro Asn Ile Asp
41 Gln Leu Ala Ser His Ser Leu Leu Phe Gln Asn Ala Phe Ala Gln Gln Ala Val Cys Ala
61 Pro Ser Arg Bal Ser Phe Leu Thr Gly Arg Arg Pro Asp Thr Thr Arg Leu Tyr Asp Phe
81 Asn Ser Tyr Trp Arg Val His Ala Gly Asn Phe Ser Thr Ile Pro Gln Tyr Phe Lys Glu
101 Asn Gly Tyr Val Thr Met Ser Val Gly Lys Val Phe His Pro Gly Ile Ser Ser Asn His
121 Thr Asp Asp Ser Pro Tyr Ser Trp Ser Phe Pro Pro Tyr His Pro Ser Ser Glu Lys Tyr
141 Glu Asn Thr Lys Thr Cys Arg Gly Pro Asp Gly Glu Leu His Ala Asn Leu Leu Cys Pro
161 Val Asp Val Leu Asp Val Pro Glu Gly Thr Leu Pro Asp Lys Gln Ser Thr Glu Gln Ala
181 Ile Gln Leu Leu Glu Lys Met Lys Thr Ser Ala Ser Pro Phe Phe Leu Ala Val Gly Tyr
201 His Lys Pro His Ile Pro Phe Arg Tyr Pro Lys Glu Phe Gln Lys Leu Tyr Pro Leu Glu
221 Asn Ile Thr Leu Ala Pro Asp Pro Glu Val Pro Asp Gly Leu Pro Pro Val Ala Tyr Asn
241 Pro Trp Met Asp Ile Arg Gln Arg Glu Asp Val Gln Ala Leu Asn Ile Ser Val Pro Tyr
261 Gly Pro Ile Pro Val Asp Phe Gln Arg Lys Ile Arg Gln Ser Tyr Phe Ala Ser Val Ser

图 1A

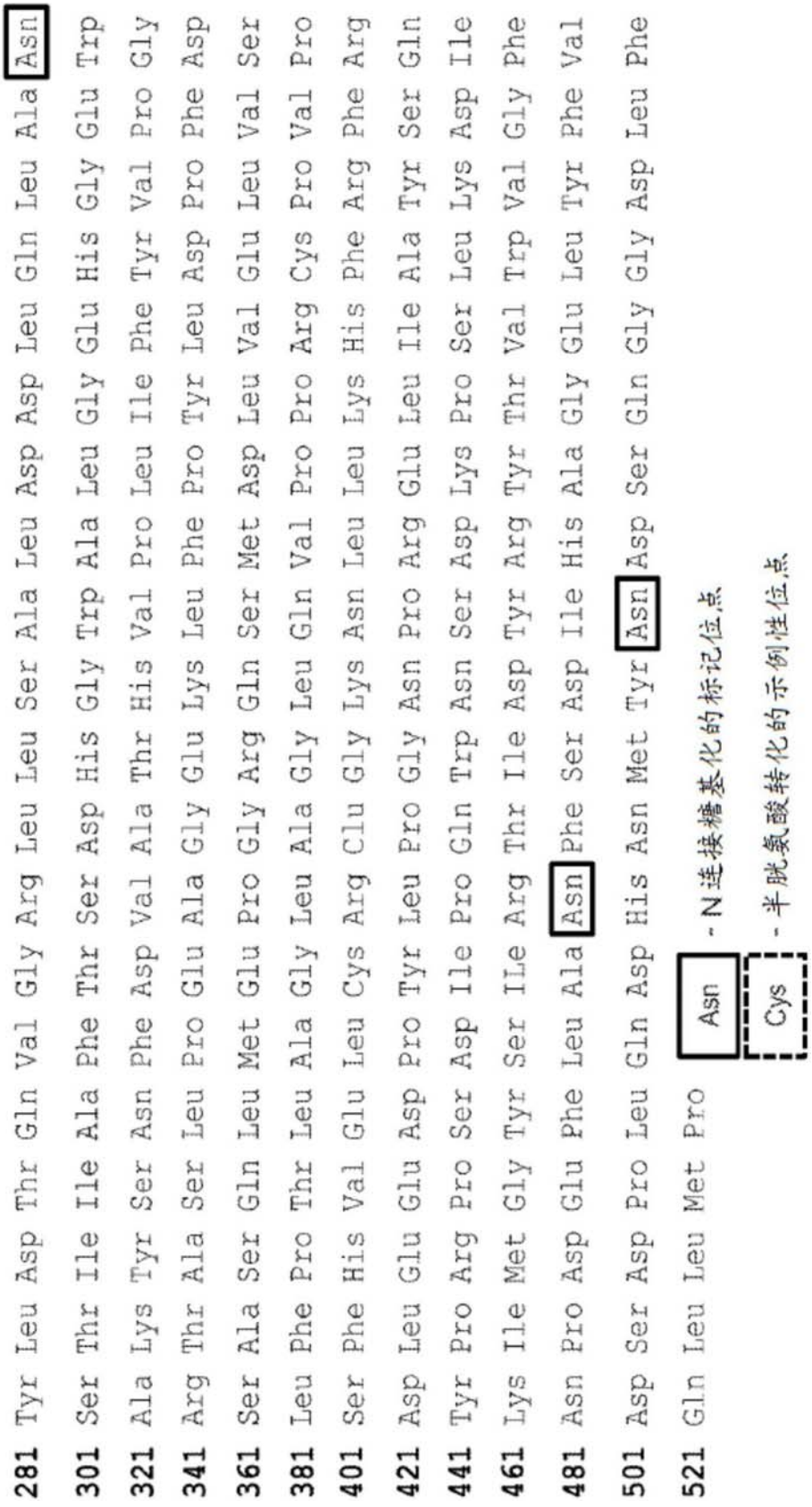

281 Tyr Leu Asp Thr Gln Val Gly Arg Leu Leu Ser Ala Leu Asp Asp Leu Gln Leu Ala Asn
301 Ser Thr Ile Ile Ala Phe Thr Ser Asp His Gly Trp Ala Leu Gly Glu His Gly Glu Trp
321 Ala Lys Tyr Ser Asn Phe Asp Val Ala Thr His Val Pro Leu Ile Phe Tyr Val Pro Gly
341 Arg Thr Ala Ser Leu Pro Glu Ala Gly Glu Lys Leu Phe Pro Tyr Leu Asp Pro Phe Asp
361 Ser Ala Ser Gln Leu Met Glu Pro Gly Arg Gln Ser Met Asp Leu Val Glu Leu Val Ser
381 Leu Phe Pro Thr Leu Ala Leu Ala Gly Leu Gln Val Pro Pro Arg Cys Pro Val Pro
401 Ser Phe His Val Glu Leu Cys Arg Clu Gly Lys Asn Leu Leu Lys His Phe Arg Phe Arg
421 Asp Leu Glu Glu Asp Pro Tyr Leu Pro Gly Asn Pro Arg Glu Leu Ile Ala Tyr Ser Gln
441 Tyr Pro Arg Pro Ser Asp Ile Pro Gln Trp Asn Ser Asp Lys Pro Ser Leu Lys Asp Ile
461 Lys Ile Met Gly Tyr Ser ILe Arg Thr Ile Asp Tyr Arg Tyr Thr Val Trp Val Gly Phe
481 Asn Pro Asp Glu Phe Leu Ala Asn Phe Ser Asp Ile His Ala Gly Glu Leu Tyr Phe Val
501 Asp Ser Asp Pro Leu Gln Asp His Asn Met Tyr Asn Asp Ser Gln Gly Gly Asp Leu Phe
521 Gln Leu Leu Met Pro
Asn
Cys
- N连接糖基化的标记位点
- 半胱氨酸转化的示例性位点

图 1B

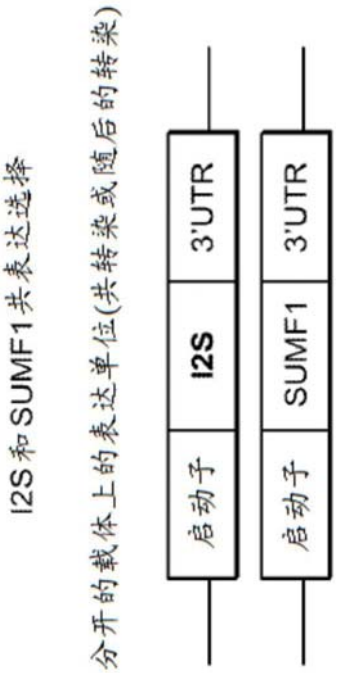

I2S和SUMF1共表达选择
分开的载体上的表达单位(共转染或随后的转染)
启动子
I2S
3'UTR
启动子
SUMF1
3'UTR

图 2A

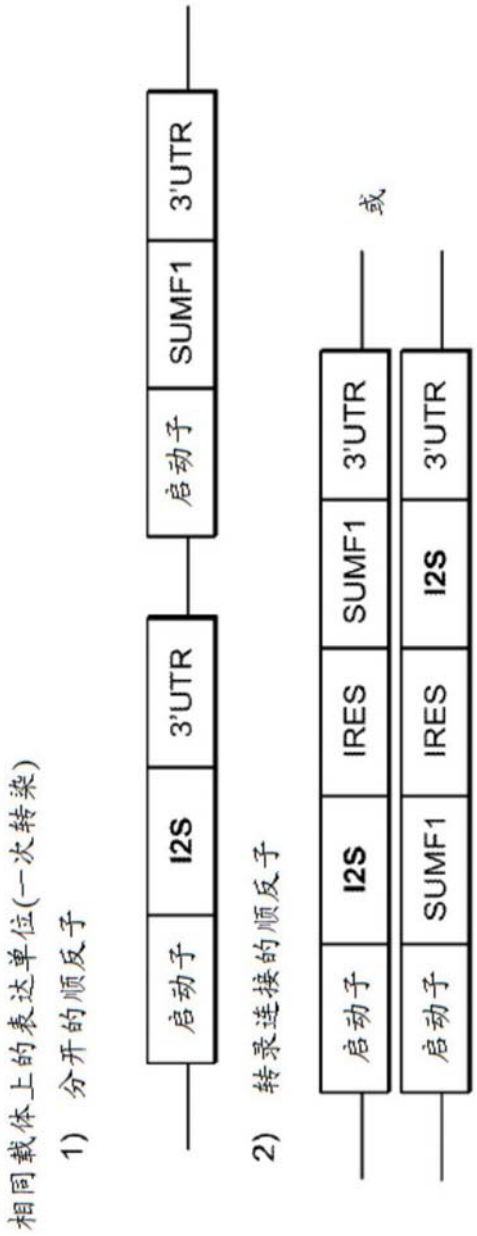

相同载体上的表达单位(一次转染)
1) 分开的顺反子
启动子
I2S
3'UTR
启动子
SUMF1
3'UTR
2) 转录连接的顺反子
启动子
I2S
IRES
SUMF1
3'UTR
或
启动子
SUMF1
IRES
I2S
3'UTR

图 2B

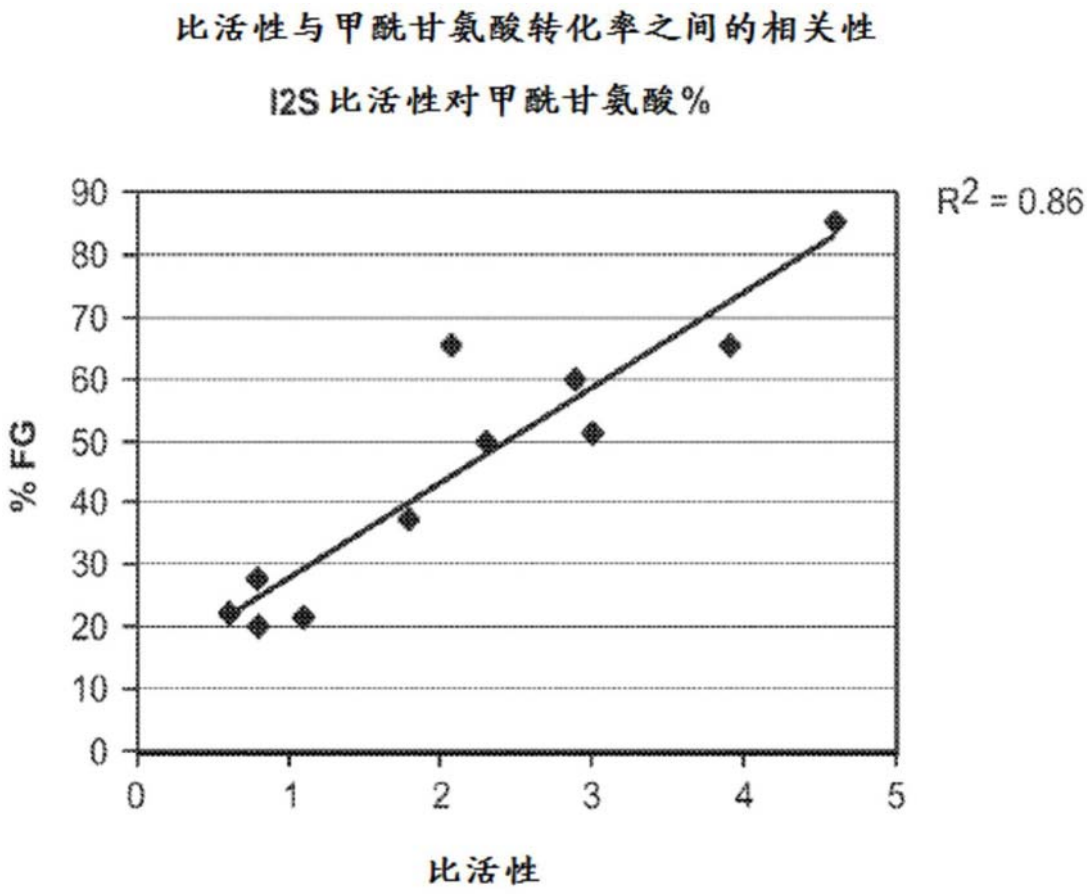

比活性与甲酰甘氨酸转化率之间的相关性
I2S 比活性对甲酰甘氨酸%
R^2 = 0.86
% FG
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
比活性

图 3

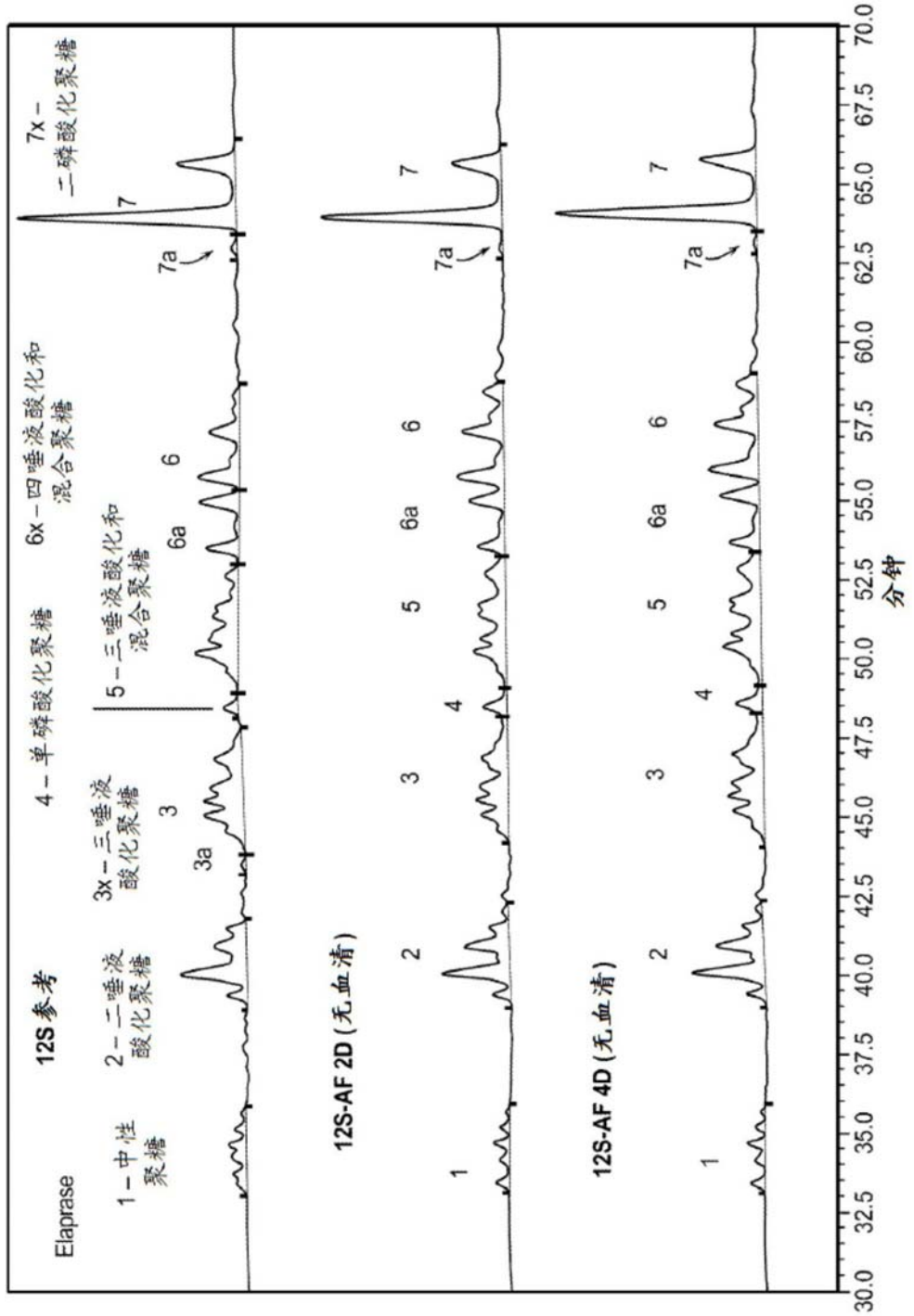

Elaprase
12S 参考
1 – 中性聚糖
2 – 二唾液酸化聚糖
3x – 三唾液酸化聚糖
3
3a
4 – 单磷酸化聚糖
5 – 三唾液酸化和混合聚糖
6
6a
6x – 四唾液酸化和混合聚糖
7
7a
7x – 二磷酸化聚糖
12S-AF 2D (无血清)
12S-AF 4D (无血清)
1
2
3
4
5
6a
6
7
7a
30.0
32.5
35.0
37.5
40.0
42.5
45.0
47.5
50.0
52.5
55.0
57.5
60.0
62.5
65.0
67.5
70.0
分钟

图 4

Abstract

The present invention provides, among other things, methods and compositions for production of recombinant I2S protein with improved potency and activity using cells co- express I2S and FGE protein. In some embodiments, cells according to the present invention are engineered to simultaneously over-express recombinant I2S and FGE proteins. Cells according to the invention are adaptable to various cell culture conditions. In some embodiments, cells of the present invention adaptable to a large-scale suspension serum- free culture.

FIG. 2A

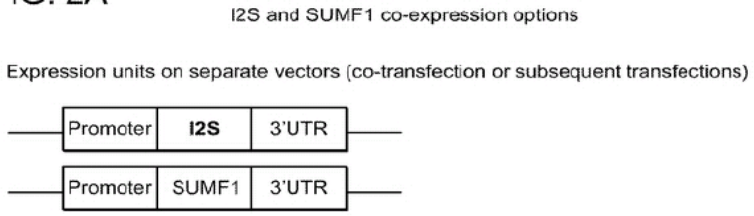


FIG. 2B