

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2005224672 B2**

(54) Title
Method of engineering a cytidine monophosphate-sialic acid synthetic pathway in fungi and yeast

(51) International Patent Classification(s)
C12N 1/14 (2006.01) **C12N 9/90** (2006.01)
C12N 9/12 (2006.01) **C12N 15/52** (2006.01)
C12N 9/16 (2006.01) **C12P 19/26** (2006.01)
C12N 9/88 (2006.01) **C12P 21/00** (2006.01)

(21) Application No: **2005224672** (22) Date of Filing: **2005.03.17**

(87) WIPO No: **WO05/090552**

(30) Priority Data

(31) Number	(32) Date	(33) Country
60/554,139	2004.03.17	US

(43) Publication Date: **2005.09.29**

(44) Accepted Journal Date: **2011.06.02**

(71) Applicant(s)
GlycoFi, Inc.

(72) Inventor(s)
Hamilton, Stephen R.

(74) Agent / Attorney
Cullens Patent and Trade Mark Attorneys, Level 32 239 George Street, Brisbane, QLD, 4000

(56) Related Art
WO 2000/052135 A2

(19) World Intellectual Property Organization
International Bureau



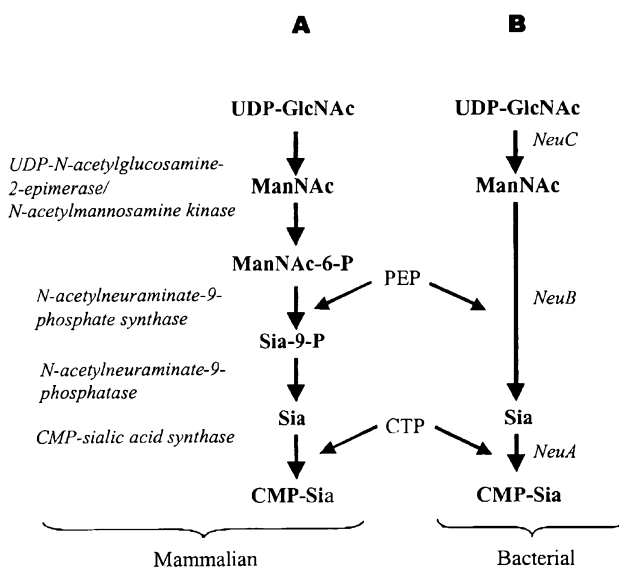
(43) International Publication Date
29 September 2005 (29.09.2005)

PCT

(10) International Publication Number
WO 2005/090552 A3

- (51) International Patent Classification⁷: C12P 21/00, C12R 1/645, C12N 1/14, 15/52, 5/10
- (21) International Application Number: PCT/US2005/009095
- (22) International Filing Date: 17 March 2005 (17.03.2005)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/554,139 17 March 2004 (17.03.2004) US
- (71) Applicant (for all designated States except US): GLYCOFI, INC. [US/US]; 21 Lafayette Street, Suite 200, Lebanon, NH 03766 (US).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): HAMILTON, Stephen, R. [GB/US]; 98 Algonquin Road, Enfield, NH 03748 (US).
- (74) Agents: HALEY, James, F., Jr. et al.; Fish & Neave IP Group, Ropes & Gray LLP, 1251 Avenue of the Americas, New York, NY 10020 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— with international search report
- (88) Date of publication of the international search report:
26 January 2006
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD OF ENGINEERING A CYTIDINE MONOPHOSPHATE-SIALIC ACID SYNTHETIC PATHWAY IN FUNGI AND YEAST



(57) Abstract: The present invention provides methods for generating CMP-sialic acid in a non-human host which lacks endogenous CMP-Sialic by providing the host with enzymes involved in CMP-sialic acid synthesis from a bacterial, mammalian or hybrid CMP-sialic acid biosynthetic pathway. Novel fungal hosts expressing a CMP-sialic acid biosynthetic pathway for the production of sialylated glycoproteins are also provided.

WO 2005/090552 A3

METHOD OF ENGINEERING A
CYTIDINE MONOPHOSPHATE-SIALIC ACID
SYNTHETIC PATHWAY IN FUNGI AND YEAST

RELATED APPLICATIONS

- 5 [0001] This application claims the benefit of U.S. Provisional Application No. 60/554,139, filed March 17, 2004, the disclosure of which is hereby incorporated by reference herein in its entirety.

FIELD OF THE INVENTION

- 10 [0002] The present invention relates to the field of protein glycosylation. The present invention further relates to novel host cells comprising genes encoding activities in the cytidine monophosphate-sialic acid (CMP-Sia) pathway, which are particularly useful in the sialylation of glycoproteins in non-human host cells which lack endogenous CMP-Sia.

BACKGROUND OF THE INVENTION

- 15 [0003] Sialic acids (Sia) are a unique group of N- or O-substituted derivatives of N-acetylneuraminic acid (Neu5Ac) that are ubiquitous in animals of the deuterostome lineage, from starfish to humans. In other organisms, including most plants, protists, Archaea, and eubacteria, these compounds are thought to be absent (Warren, L. 1994). Exceptions have been identified, all of which are in pathogenic
20 organisms, including certain bacteria, protozoa and fungi (Kelm, S. and Schauer, R. 1997) (Parodi, A.J. 1993) (Alviano, C.S., Travassos, L.R., et al. 1999). The

mechanism by which pathogenic fungi, including *Cryptococcus neoformans* and *Candida albicans*, acquire sialic acid on cell surface glycoproteins and glycolipids remains undetermined (Alviano, C.S., Travassos, L.R., et al. 1999). It has been demonstrated, however, that when these organisms are grown in sialic acid-free
5 media, sialic acid residues are found on cellular glycans, suggesting *de novo* synthesis of sialic acid. To date, no enzymes have been identified in fungi that are involved in the biosynthesis of sialic acid. The mechanism by which protozoa sialylate cell surface glycans has been well characterized. Protozoa, such as *Trypanosoma cruzi*, possess an external trans-sialidase that adds sialic acid to cell
10 surface glycoproteins and glycolipids in a CMP-Sia independent mechanism (Parodi, A.J. 1993) The identification of a similar trans-sialidase in fungi would help to elucidate the mechanism of sialic acid transfer on cellular glycans, but such a protein has not yet been identified or isolated.

[0004] Despite the absence and/or ambiguity of sialic acid biosynthesis in fungi,
15 sialic acid biosynthesis in pathogenic bacteria and mammalian cells is well understood. A group of pathogenic bacteria have been identified which possess the ability to synthesize sialic acids *de novo* to generate sialylated glycolipids that occur on the cell surface (Vimr, E., Steenbergen, S., et al. 1995). Although sialic acids on the surface of these pathogenic organisms are predominantly thought to be
20 a means of evading the host immune system, it has been shown that these same sialic acid molecules are also involved in many processes in higher organisms, including protein targeting, cell-cell interaction, cell-substrate recognition and adhesion (Schauer, et al., 2000).

[0005] The presence of sialic acids can affect biological activity and *in vivo* half-
25 life (MacDougall et al., 1999). For example, the importance of sialic acids has been demonstrated in studies of the human erythropoietin (hEPO). The terminal sialic acid residues on the carbohydrate chains of the N-linked glycan of this glycoprotein prevent rapid clearance of hEPO from the blood and improve *in vivo* activity. Asialylated-hEPO (asialo-hEPO), which terminates in a galactose
30 residue, has dramatically decreased erythropoietic activity *in vivo*. This decrease is caused by the increased clearance of the asialo-hEPO by the hepatic asialoglycoprotein receptor (Fukuda, M.N., Sasaki, H., et al. 1989) (Spivak, J.L.

and Hogans, B.B. 1989). Similarly, the absence of the terminal sialic acid on many therapeutic glycoproteins can reduce efficacy, and thus require more frequent dosing.

[0006] Although many of the currently available therapeutic glycoproteins are made in mammalian cell lines, these systems are expensive and typically yield low product titers. To overcome these shortcomings the pharmaceutical industry is currently investigating new approaches. One approach is the production of glycoproteins in fungal systems. Fungal expression systems are less expensive to maintain, and are capable of producing higher titers per unit culture (Cregg, J.M. et al., 2000). The disadvantage, however, is that fungal and mammalian glycosylation differ greatly, and therapeutic proteins with non-human glycosylation have a high risk of eliciting an immune response in humans (Ballou, C.E., 1990). Although the initial stages of N-linked glycosylation in the endoplasmic reticulum are similar in fungi and mammals, subsequent processing in the Golgi results in dramatically different glycans. Nonetheless, these divergent glycosylation pathways can be overcome by genetically engineering the fungal host to produce human-like glycoproteins as described in WO 02/00879, WO 03/056914, US 2004/0018590, Choi et al., 2003 and Hamilton et al., 2003. It is, therefore, desirable to have a novel protein expression system (e.g., fungal system) that is capable of producing fully sialylated human-like glycoproteins.

[0007] A method to engineer a CMP-Sia biosynthetic pathway into non-human host cells which lack endogenous CMP-Sia is needed. Non-human hosts which lack endogenous CMP-Sia include most lower eukaryotes such as fungi, most plants and non-pathogenic bacteria.

[0008] To date, no fungal system has been identified that generates sialylated glycoproteins from an endogenous pool of the sugar substrate CMP-Sia. What is needed, therefore, is a method to engineer a CMP-Sia biosynthetic pathway into a non-human host which lacks endogenous CMP-Sia, such as a fungal host, to ensure that substrates required for sialylation are present in useful quantities for the production of therapeutic glycoproteins.

SUMMARY OF THE INVENTION

[0009] A method for engineering a functional CMP-sialic acid (CMP-Sia) biosynthetic pathway into a non-human host cell lacking endogenous CMP-Sia, such as a fungal host cell, is provided. The method involves the cloning and
5 expression of several enzymes of mammalian origin, bacterial origin or both, in a host cell, particularly a fungal host cell. The engineered CMP-Sia biosynthetic pathway is useful for producing sialylated glycolipids, O-glycans and N-glycans *in vivo*. The present invention is thus useful for facilitating the generation of sialylated therapeutic glycoproteins in non-human host cells lacking endogenous
10 sialylation, such as fungal host cells.

Modified Hosts Comprising A Cellular Pool Of
CMP-Sia Or A CMP-Sia Biosynthetic Pathway

[0010] The invention comprises a recombinant non-human host cell comprising a cellular pool of CMP-Sia, wherein the host cell lacks endogenous CMP-Sia. In
15 one embodiment, the CMP-Sia comprises a sialic acid selected from Neu5Ac, N-glycolylneuraminic acid (Neu5Gc), and keto-3-deoxy-D-glycero-D-galactononic acid (KDN).

[0011] The invention further comprises a recombinant non-human host cell comprising a CMP-Sia biosynthetic pathway, wherein the host cell lacks
20 endogenous CMP-Sia.

[0012] In another embodiment, the invention comprises a non-human host cell comprising one or more recombinant enzymes that participate in the biosynthesis of CMP-Sia, wherein the host cell lacks endogenous CMP-Sia.

[0013] In one embodiment, the host cell of the invention is a fungal host cell.

[0014] In one embodiment, the host cell of the invention produces at least one intermediate selected from the group consisting of UDP-GlcNAc, ManNAc, ManNAc-6-P, Sia-9-P and Sia. In one embodiment, the intermediate is UDP-GlcNAc. In one embodiment, the intermediate is ManNAc. In one embodiment, the intermediate is ManNAc-6-P. In one embodiment, the intermediate is Sia-9-P.
25
30 In one embodiment, the intermediate is Sia.

[0015] In one embodiment, the host cell of the invention comprises a cellular pool of CMP-Sia. In one embodiment, the CMP-Sia comprises a sialic acid

selected from Neu5Ac, N-glycolylneuraminic acid (Neu5Gc), and keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN).

[0016] In one embodiment, the host cell of the invention expresses one or more enzyme activities selected from *E. coli NeuC*, *E. coli NeuB* and *E. coli NeuA*.

5 [0017] In one embodiment, the host cell of the invention expresses one or more enzyme activities selected from *E. coli NeuC*, *E. coli NeuB* and a mammalian CMP-sialate synthase activity.

[0018] In one embodiment, the host cell of the invention expresses one or more enzyme activities selected from *E. coli NeuC*, *E. coli NeuB* and a mammalian
10 CMP-sialate synthase activity, and further expresses at least one enzyme activity selected from UDP-GlcNAc epimerase, sialate synthase, CMP-sialate synthase, UDP-N-acetylglucosamine-2-epimerase, N-acetylmannosamine kinase, N-acetylneuraminate-9-phosphate synthase, N-acetylneuraminate-9-phosphatase and CMP-sialic acid synthase.

15 [0019] In one embodiment, the host cell of the invention expresses at least one enzyme activity selected from UDP-GlcNAc epimerase, sialate synthase, CMP-sialate synthase, UDP-N-acetylglucosamine-2-epimerase, N-acetylmannosamine kinase, N-acetylneuraminate-9-phosphate synthase, N-acetylneuraminate-9-phosphatase and CMP-sialic acid synthase.

20 [0020] In one embodiment, the host cell of the invention expresses *E. coli NeuC*. In one embodiment, the host cell expresses *E. coli NeuB*. In one embodiment, the host cell expresses *E. coli NeuA*.

[0021] In one embodiment, the host cell of the invention expresses the enzyme activity of UDP-GlcNAc epimerase. In one embodiment, the host cell of the
25 invention expresses the enzyme activity of sialate synthase. In one embodiment, the host cell of the invention expresses the enzyme activity of CMP-sialate synthase. In one embodiment, the host cell of the invention expresses the enzyme activity of UDP-N-acetylglucosamine-2-epimerase. In one embodiment, the host cell of the invention expresses the enzyme activity of N-acetylmannosamine
30 kinase. In one embodiment, the host cell of the invention expresses the enzyme activity of N-acetylneuraminate-9-phosphate synthase. In one embodiment, the host cell of the invention expresses the enzyme activity of N-acetylneuraminate-9-

phosphatase. In one embodiment, the host cell of the invention expresses the enzyme activity of CMP-sialic acid synthase.

5 [0022] In one embodiment, the enzyme activity of *NeuC* is expressed from a nucleic acid comprising the nucleic acid sequence of SEQ ID NO:13, or a portion thereof. In one embodiment, the enzyme activity of *NeuC* is from a polypeptide comprising the amino acid sequence of SEQ ID NO:14 or a fragment thereof.

10 [0023] In one embodiment, the enzyme activity of *NeuB* is expressed from a nucleic acid comprising the nucleic acid sequence of SEQ ID NO:15, or a portion thereof. In one embodiment, the enzyme activity of *NeuB* is from a polypeptide comprising the amino acid sequence of SEQ ID NO:16 or a fragment thereof.

[0024] In one embodiment, the enzyme activity of *NeuA* is expressed from a nucleic acid comprising the nucleic acid sequence of SEQ ID NO:17, or a portion thereof. In one embodiment, the enzyme activity of *NeuA* is from a polypeptide comprising the amino acid sequence of SEQ ID NO:18 or a fragment thereof.

15 [0025] In one embodiment, the enzyme activity of CMP-synthase is expressed from a nucleic acid comprising the nucleic acid sequence of SEQ ID NO:19, or a portion thereof. In one embodiment, the enzyme activity of CMP-synthase is from a polypeptide comprising the amino acid sequence of SEQ UD NO:20 or a fragment thereof.

20 [0026] In one embodiment, the enzyme activity of CMP-synthase is expressed from a nucleic acid comprising the nucleic acid sequence of GenBank Accession No. AF397212, or a portion thereof. In one embodiment, the enzyme activity of CMP-synthase is from a polypeptide comprising the amino acid sequence of AAM90588 or a fragment thereof.

25 [0027] In one embodiment, the enzyme activity of GlcNAc epimerase is expressed from a nucleic acid comprising the nucleic acid sequence of SEQ ID NO:21, or a portion thereof. In one embodiment, the enzyme activity of GlcNAc is from a polypeptide comprising the amino acid sequence of SEQ ID NO:22 or a fragment thereof.

30 [0028] In one embodiment, the enzyme activity of sialate aldolase is expressed from a nucleic acid comprising the nucleic acid sequence of SEQ ID NO:23, or a portion thereof. In one embodiment, the enzyme activity of sialate aldolase is from

a polypeptide comprising the amino acid sequence of SEQ ID NO:24 or a fragment thereof.

[0029] In one embodiment, the host cell of the invention produces at least one intermediate selected from the group consisting of UDP-GlcNAc, ManNAc, 5 ManNAc-6-P, Sia-9-P and Sia. In one embodiment, the intermediate is UDP-GlcNAc. In one embodiment, the intermediate is ManNAc. In one embodiment, the intermediate is ManNAc-6-P. In one embodiment, the intermediate is Sia-9-P. In one embodiment, the intermediate is Sia.

[0030] In one embodiment, the host cell of the invention expresses a 10 heterologous therapeutic protein. In one embodiment, said therapeutic protein is selected from the group consisting of: erythropoietin, cytokines, interferon- α , interferon- β , interferon- γ , interferon- ω , TNF- α , granulocyte-CSF, GM-CSF, interleukins, IL-1ra, coagulation factors, factor VIII, factor IX, human protein C, antithrombin III and thrombopoietin, IgA antibodies or fragments thereof, IgG 15 antibodies or fragments thereof, IgA antibodies or fragments thereof, IgD antibodies or fragments thereof, IgE antibodies or fragments thereof, IgM antibodies and fragments thereof, soluble IgE receptor α -chain, urokinase, chymase, urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, FSH, annexin V fusion protein, angiostatin, 20 vascular endothelial growth factor-2, myeloid progenitor inhibitory factor-1, osteoprotegerin, α -1 antitrypsin, DNase II, α -feto proteins and glucocerebrosidase.

[0031] In one embodiment, the host cell is from a fungal host. In one embodiment, the fungal host is selected from the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia minuta*, *Ogataea minuta*, *Pichia lindneri*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia sp.*, *Saccharomyces cerevisiae*, *Saccharomyces sp.*, *Hansenula polymorpha*, *Kluyveromyces sp.*, *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus 30 oryzae*, *Aspergillus sp.*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum* and *Neurospora crassa*. In one embodiment, the fungal host is *P. pastoris*.

[0032] In one embodiment, the host cell of the invention is from a non-pathogenic bacteria. In another embodiment, the host cell of the invention is from a plant.

5 [0033] In one embodiment, the enzyme activity is expressed under the control of a constitutive promoter.

[0034] In another embodiment, the enzyme activity is expressed under the control of an inducible promoter.

[0035] In one embodiment, the expressed enzyme activity is from a partial ORF encoding that enzymatic activity.

10 [0036] In another embodiment, the expressed enzyme is a fusion to another protein or peptide.

[0037] In another embodiment, the expressed enzyme has been mutated to enhance or attenuate the enzymatic activity.

15 [0038] In one embodiment, the recombinant host cells of the invention have modified oligosaccharides which may be modified further by heterologous expression of a set of glycosyltransferases, sugar transporters and mannosides as described in WO02/00879, WO03/056914 and US 2004/0018590.

Method Of Producing CMP-Sia In A Host

20 [0039] The invention further comprises a method for producing CMP-Sia in a recombinant non-human host comprising expressing a CMP-Sia biosynthetic pathway.

[0040] In one embodiment, the invention comprises a method for producing CMP-Sia, comprising expressing in a non-human host cell one or more recombinant enzymes that participate in the biosynthesis of CMP-Sia.

25 [0041] In one embodiment, the host cell of the invention is a fungal host cell.

[0042] In one embodiment, the method of the invention comprises expressing at least one enzyme activity from a prokaryotic CMP-Sia biosynthetic pathway. In one embodiment, the method of the invention comprises expressing at least one enzyme activity selected from the group consisting of *E. coli NeuC*, *E. coli NeuB* and *E. coli NeuA* activity.

30 [0043] In another embodiment, the method of the invention comprises expressing at least one enzyme activity from a mammalian CMP-Sia biosynthetic pathway.

[0044] In one embodiment, the method of the invention comprises expressing a mammalian CMP-sialate synthase activity. In one embodiment, the CMP-sialate synthase activity localizes in the nucleus.

5 [0045] In one embodiment, the method of the invention comprises expressing a hybrid CMP-Sia biosynthetic pathway. In one embodiment, the method of the invention comprises expressing at least one enzyme activity selected from *E. coli NeuC*, *E. coli NeuB* and a mammalian CMP-sialate synthase activity. In one embodiment, the CMP-sialate synthase activity localizes in the nucleus.

10 [0046] In one embodiment, the enzyme activity of *NeuB* is expressed from a nucleic acid comprising the nucleic acid sequence of SEQ ID NO:15, or a portion thereof. In one embodiment, the enzyme activity of *NeuB* is from a polypeptide comprising the amino acid sequence of SEQ ID NO:16 or a fragment thereof.

15 [0047] In one embodiment, the enzyme activity of *NeuA* is expressed from a nucleic acid comprising the nucleic acid sequence of SEQ ID NO:17, or a portion thereof. In one embodiment, the enzyme activity of *NeuA* is from a polypeptide comprising the amino acid sequence of SEQ ID NO:18 or a fragment thereof.

[0048] In one embodiment, the enzyme activity of CMP-synthase is expressed from a nucleic acid comprising the nucleic acid sequence of SEQ ID NO:19, or a portion thereof. In one embodiment, the enzyme activity of CMP-synthase is from a polypeptide comprising the amino acid sequence of SEQ ID NO:20 or a fragment thereof.

20 [0049] In one embodiment, the enzyme activity of CMP-synthase is expressed from a nucleic acid comprising the nucleic acid sequence of GenBank Accession No. AF397212, or a portion thereof. In one embodiment, the enzyme activity of
25 CMP-synthase is from a polypeptide comprising the amino acid sequence of AAM90588 or a fragment thereof.

[0050] In one embodiment, the method of the invention comprises using a host cell which expresses a heterologous therapeutic protein. In one embodiment, said therapeutic protein is selected from the group consisting of: erythropoietin,
30 cytokines, interferon- α , interferon- β , interferon- γ , interferon- ω , TNF- α , granulocyte-CSF, GM-CSF, interleukins, IL-1ra, coagulation factors, factor VIII, factor IX, human protein C, antithrombin III and thrombopoietin, IgA antibodies or

fragments thereof, IgG antibodies or fragments thereof, IgA antibodies or fragments thereof, IgD antibodies or fragments thereof, IgE antibodies or fragments thereof, IgM antibodies and fragments thereof, soluble IgE receptor α -chain, urokinase, chymase, urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, FSH, annexin V fusion protein, angiostatin, vascular endothelial growth factor-2, myeloid progenitor inhibitory factor-1, osteoprotegerin, α -1 antitrypsin, DNase II, α -feto proteins and glucocerebrosidase.

[0051] In one embodiment, the non-human host cell to be used is from a fungal host. In one embodiment, the fungal host is selected from the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia sp.*, *Saccharomyces cerevisiae*, *Saccharomyces sp.*, *Hansenula polymorpha*, *Kluyveromyces sp.*, *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Aspergillus sp.*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum* and *Neurospora crassa*. In one embodiment, the fungal host is *Pichia pastoris*.

[0052] In one embodiment, the host cell of the invention is from a non-pathogenic bacteria. In another embodiment, the host cell of the invention is from a plant.

[0053] In one embodiment, the CMP-Sia synthesis is enhanced by supplementing a medium for growing the non-human host cell with one or more intermediate substrates used in the CMP-Sia synthesis. In one embodiment, the intermediates are selected from the group consisting of UDP-GlcNAc, ManNAc, ManNAc-6-P, Sia-9-P and Sia.

[0054] In one embodiment, the enzyme activity is expressed under the control of a constitutive promoter.

[0055] In another embodiment, the enzyme activity is expressed under the control of an inducible promoter.

[0056] In one embodiment, the expressed enzyme activity is from a partial ORF encoding that enzymatic activity.

[0057] In another embodiment, the expressed enzyme is a fusion to another protein or peptide.

5 [0058] In another embodiment, the expressed enzyme has been mutated to enhance or attenuate the enzymatic activity.

[0059] In one embodiment the methods described above comprise the use of a host having modified oligosaccharides which may be modified further by heterologous expression of a set of glycosyltransferases, sugar transporters and
10 mannosides as described in WO02/00879, WO03/056914 and US 2004/0018590.

Methods Of Producing Recombinant Glycoproteins

[0060] In one embodiment, the invention provides a method for producing recombinant glycoprotein comprising the step of producing a cellular pool of CMP-Sia in a recombinant non-human host cell which lacks endogenous CMP-Sia
15 and expressing the glycoprotein in said host. In one embodiment, the host is a fungal host.

[0061] In another embodiment, the invention provides a method for producing recombinant glycoprotein comprising the step of engineering a CMP-Sia biosynthetic pathway in a non-human host cell which lacks endogenous CMP-Sia
20 and expressing the glycoprotein said host. In one embodiment, the host is a fungal host. In one embodiment, the CMP-Sia pathway results in the formation of a cellular pool of CMP-Sia.

[0062] In another embodiment, the invention provides a method for producing recombinant glycoprotein comprising the step of expressing one or more
25 recombinant enzymes that participate in the biosynthesis of CMP-Sia in a non-human host cell which lacks endogenous CMP-Sia and expressing the glycoprotein in said host. In one embodiment, the host is a fungal host.

[0063] In any of the embodiments of the invention, the recombinant non-human host cell may have modified oligosaccharides which may be modified further by
30 heterologous expression of recombinant glycosylation enzymes (such as sialyltransferases, mannosidases, fucosyltransferases, galactosyltransferases, GcINAc transferases, ER and Golgi specific transporters, enzymes involved in the

- 12 -

processing of oligosaccharides, and enzymes involved in the synthesis of activated oligosaccharide precursors such as UDP-galactose and CMP-N-acetylneuraminic acid) which may be necessary for the production of a human-like glycoprotein in a non-human host as described in WO02/00879, WO03/056914 and US2004/0018590.

5 [0064] In any of the embodiments of the invention, the host cell may express a heterologous therapeutic protein. In one embodiment, said therapeutic protein is selected from the group consisting of: erythropoietin, cytokines, interferon- α , interferon- β , interferon- γ , interferon- ω , TNF- α , granulocyte-CSF, GM-CSF, interleukins, IL-1ra, coagulation factors, factor VIII, factor IX, human protein C, antithrombin III and thrombopoietin, IgA antibodies or fragments thereof, IgG antibodies or fragments thereof, IgD antibodies or fragments thereof, IgE antibodies or fragments thereof, IgM antibodies or fragments thereof, soluble IgE receptor α -chain, urokinase, chymase, urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, FSH, annexin V fusion protein, angiostatin, vascular endothelial growth factor-2, myeloid progenitor inhibitory factor-1, osteoprotegerin, α -1 antitrypsin, DNase II, α -feto proteins and glucocerebrosidase.

10 [0065] It is to be understood that single or multiple enzymatic activities may be introduced into a non-human host cell in any fashion, by use of one or more nucleic acid molecules, without necessarily using a nucleic acid, plasmid or vector that is specifically disclosed in the foregoing description of the invention.

[0065a] Definitions of the specific embodiments of the invention as claimed herein follow.

20 According to a first embodiment of the invention, there is provided a recombinant fungal host cell comprising a CMP-Sia biosynthetic pathway consisting of *E. coli NeuC* enzyme activity, *E. coli NeuB* enzyme activity, and a mammalian CMP-sialate synthase activity or a fusion protein comprising an *E. coli* CMP-sialate synthase (*NeuA*) catalytic domain and a cellular targeting signal peptide not normally associated with the catalytic domain selected to target the catalytic domain to the nucleus of the host cell and a sialyltransferase wherein the host cell is capable of producing sialylated glycolipids, O-glycans or N-glycans.

25 [0065b] According to a second embodiment of the invention, there is provided a method for producing sialylated glycolipids, O-glycans or N-glycans in a fungal host cell comprising expressing a CMP-Sia biosynthetic pathway consisting of *E. coli NeuC* enzyme activity, *E. coli NeuB* enzyme activity, and a mammalian CMP-sialate synthase activity or a fusion protein comprising an *E. coli* CMP-sialate synthase (*NeuA*) catalytic domain and a cellular targeting signal peptide not normally associated with the catalytic domain selected to target the catalytic domain to the nucleus of the host cell and a sialyltransferase.

30 [0065c] According to a third embodiment of the invention, there is provided a method for producing a recombinant glycoprotein comprising expressing said glycoprotein in the host cell of the first embodiment.

2005224672 19 May 2011

- 12a -

19 May 2011

2005224672

[0065d] According to a fourth embodiment of the invention, there is provided a recombinant fungal host cell comprising a CMP-Sia biosynthetic pathway consisting of *E. coli NeuC* enzyme activity, *E. coli NeuB* enzyme activity, and a mammalian CMP-sialate synthase activity or a fusion protein comprising an *E. coli* CMP-sialate synthase (*NeuA*) catalytic domain and a cellular targeting signal peptide not normally associated with the catalytic domain selected to target the catalytic domain to the nucleus of the host cell.

[0065e] According to a fifth embodiment of the invention, there is provided a method for producing a cellular pool of CMP-sialic acid in a fungal host cell wherein the CMP-sialic acid in the pool consists of *N*-acetyl-5-neuraminic acid in a fungal host cell comprising:

(a) providing a yeast host cell comprising a CMP-Sia biosynthetic pathway consisting of *E. coli NeuC* enzyme activity, *E. coli NeuB* enzyme activity, and a mammalian CMP-sialate synthase activity or a fusion protein comprising an *E. coli* CMP-sialate synthase (*NeuA*) catalytic domain and a cellular targeting signal peptide not normally associated with the catalytic domain selected to target the catalytic domain to the nucleus of the host cell; and

(b) growing the yeast host cell under conditions effective for producing the cellular pool of CMP-sialic acid wherein the CMP-sialic acid in the pool consists of *N*-acetyl-5-neuraminic acid.

BRIEF DESCRIPTION OF THE DRAWINGS

[0066] **Figure 1** illustrates the CMP-sialic acid biosynthetic pathway in mammals and bacteria.

Enzymes involved in each pathway are italicized. The primary substrates, intermediates and products are in bold. (PEP: phosphoenol pyruvate; CTP: cytidine triphosphate)

[0067] **Figure 2** shows the open reading frame (ORF) of *E. coli* protein *NeuC* (Genbank: M84026.1; SEQ ID NO: 13) and the predicted amino acid sequence (SEQ ID NO:14). The underlined DNA sequences are regions to which primers have been designed to amplify the ORF.

[Text continues on page 13.]

[0068] **Figure 3** shows the ORF of *E. coli* protein *NeuB* (Genbank: U05248.1; SEQ ID NO:15) and the predicted amino acid sequence (SEQ ID NO:16). The underlined DNA sequences are regions to which primers have been designed to amplify the ORF.

5 [0069] **Figure 4** shows the ORF of *E. coli* protein *NeuA* (Genbank: J05023.1; SEQ ID NO:17) and the predicted amino acid sequence (SEQ ID NO:18). The underlined DNA sequences are regions to which primers have been designed to amplify the ORF.

[0070] **Figure 5** shows the ORF of *Mus musculus* CMP-Sia synthase (Genbank: 10 AJ006215; SEQ ID NO:19) and the amino acid sequence (SEQ ID NO:20). The underlined DNA sequences are regions to which primers have been designed to amplify the ORF.

[0071] **Figure 6** illustrates an alternative biosynthetic route for generating *N*-acetylmannosamine (ManNAc) *in vivo*. Enzymes involved in each pathway are 15 italicized. The primary substrates, intermediates and products are in bold.

[0072] **Figure 7** shows the ORF of *Sus scrofa* GlcNAc epimerase (Genbank: D83766; SEQ ID NO: 21) and the amino acid sequence (SEQ ID NO:22). The underlined DNA sequences are regions to which primers have been designed to amplify the ORF.

20 [0073] **Figure 8** illustrates the reversible reaction catalyzed by sialate aldolase and its dependence on sialic acid (Sia) concentration. Enzymes involved in each pathway are italicized. The primary substrates, intermediates and products are in bold.

[0074] **Figure 9** shows the ORF of *E. coli* sialate aldolase (Genbank: X03345; 25 SEQ ID NO:23) and the amino acid sequence (SEQ ID NO:24). The underlined DNA sequences are regions to which primers have been designed to amplify the ORF.

[0075] **Figure 10** shows a HPLC of negative control of cell extracts from strain YSH99a incubated under assay conditions (**Example 10**) in the absence of 30 acceptor glycan. The doublet peak eluting at 26.5 min results from contaminating cellular component(s).

[0076] **Figure 11** shows a HPLC of positive control cell extract from strain YSH99a incubated under assay conditions (**Example 10**) in the presence of 2-AB (aminobenzamide) labeled acceptor glycan and supplemented with CMP-sialic acid. The peak eluting at 23 min corresponds to sialylation on each branch of a
5 biantennary galactosylated N-glycan. The doublet peak eluting at 26.5 min results from contaminating cellular component(s).

[0077] **Figure 12** shows a HPLC of a cell extract from strain YSH99a incubated under assay conditions (**Example 10**) in the presence of acceptor glycan with no exogenous CMP-sialic acid. The peaks eluting at 20 and 23 min correspond to
10 mono- and di-sialylation of a biantennary galactosylated N-glycan. The doublet peak eluting at 26.5 min results from contaminating cellular component(s).

[0078] **Figure 13** shows sialidase treatment of N-glycans from YSH99a extract incubation. The sample illustrated in **Figure 12** was incubated overnight at 37°C in the presence of 100 U sialidase (New England Biolabs, Beverly, MA). The
15 peaks eluting at 20 and 23 min, corresponding to mono- and di-sialylated N-glycan, have been removed. The contaminating peak at 26 min remains.

[0079] **Figure 14** shows commercial mono- and di-sialylated N-glycan standards. The peaks eluting at 20 and 23 min correspond to mono- and di-sialylation of the commercial standards A1 and A2 (Glyko Inc., San Rafael, CA).

20

DETAILED DESCRIPTION OF THE INVENTION

[0080] Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include
25 the singular. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art. Generally, nomenclatures used in connection with, and techniques of biochemistry, enzymology, molecular and cellular biology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well
30 known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are

cited and discussed throughout the present specification unless otherwise indicated. See, e.g., Sambrook, J. and Russell, D.W. (2001); Ausubel et al., *Current Protocols in Molecular Biology*, Greene Publishing Associates (1992, and Supplements to 2002); *Harlow and Lane Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1990); *Introduction to Glycobiology*, Maureen E. Taylor, Kurt Drickamer, Oxford Univ. Press (2003); *Worthington Enzyme Manual*, Worthington Biochemical Corp. Freehold, NJ; *Handbook of Biochemistry: Section A Proteins Vol I* 1976 CRC Press; *Handbook of Biochemistry: Section A Proteins Vol II* 1976 CRC Press; *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press (1999). The nomenclatures used in connection with, and the laboratory procedures and techniques of, biochemistry and molecular biology described herein are those well known and commonly used in the art.

[0081] All publications, patents and other references mentioned herein are incorporated by reference.

[0082] The following terms, unless otherwise indicated, shall be understood to have the following meanings:

[0083] The term "polynucleotide" or "nucleic acid molecule" refers to a polymeric form of nucleotides of at least 10 bases in length. The term includes DNA molecules (e.g., cDNA or genomic or synthetic DNA) and RNA molecules (e.g., mRNA or synthetic RNA), as well as analogs of DNA or RNA containing non-natural nucleotide analogs, non-native internucleoside bonds, or both. The nucleic acid can be in any topological conformation. For instance, the nucleic acid can be single-stranded, double-stranded, triple-stranded, quadruplexed, partially double-stranded, branched, hairpinned, circular, or in a padlocked conformation. The term includes single and double stranded forms of DNA.

[0084] Unless otherwise indicated, a "nucleic acid comprising SEQ ID NO:X" refers to a nucleic acid, at least a portion of which has either (i) the sequence of SEQ ID NO:X, or (ii) a sequence complementary to SEQ ID NO:X. The choice between the two is dictated by the context. For instance, if the nucleic acid is used as a probe, the choice between the two is dictated by the requirement that the probe be complementary to the desired target.

[0085] An "isolated" or "substantially pure" nucleic acid or polynucleotide (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, e.g., ribosomes, polymerases, and genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, or (4) does not occur in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems.

[0086] However, "isolated" does not necessarily require that the nucleic acid or polynucleotide so described has itself been physically removed from its native environment. For instance, an endogenous nucleic acid sequence in the genome of an organism is deemed "isolated" herein if a heterologous sequence (i.e., a sequence that is not naturally adjacent to this endogenous nucleic acid sequence) is placed adjacent to the endogenous nucleic acid sequence, such that the expression of this endogenous nucleic acid sequence is altered. By way of example, a non-native promoter sequence can be substituted (e.g., by homologous recombination) for the native promoter of a gene in the genome of a human cell, such that this gene has an altered expression pattern. This gene would now become "isolated" because it is separated from at least some of the sequences that naturally flank it.

[0087] A nucleic acid is also considered "isolated" if it contains any modifications that do not naturally occur to the corresponding nucleic acid in a genome. For instance, an endogenous coding sequence is considered "isolated" if it contains an insertion, deletion or a point mutation introduced artificially, e.g., by human intervention. An "isolated nucleic acid" also includes a nucleic acid integrated into a host cell chromosome at a heterologous site, a nucleic acid construct present as an episome. Moreover, an "isolated nucleic acid" can be substantially free of other cellular material, or substantially free of culture medium

when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[0088] As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence encompasses nucleic acid sequences that can be translated, according to
5 the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

[0089] The term "percent sequence identity" or "identical" in the context of nucleic acid sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence. The length of sequence identity
10 comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For
15 instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, 1990, (herein incorporated by reference). For instance,
20 percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

[0090] The term "substantial homology" or "substantial similarity," when
25 referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more
30 preferably at least about 95%, 96%, 97%, 98% or 99% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

[0091] Alternatively, substantial homology or similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under stringent hybridization conditions. "Stringent hybridization conditions" and "stringent wash conditions" in
5 the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be
10 readily appreciated by those skilled in the art. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization.

[0092] In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of
15 conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook, J. and Russell, D.W. (2001), *supra*, page 9.51, hereby incorporated by reference. For purposes herein, "high stringency
20 conditions" are defined for solution phase hybridization as aqueous hybridization (i.e., free of formamide) in 6X SSC (where 20X SSC contains 3.0 M NaCl and 0.3 M sodium citrate), 1% SDS at 65°C for 8-12 hours, followed by two washes in 0.2X SSC, 0.1% SDS at 65°C for 20 minutes. It will be appreciated by the skilled
25 worker that hybridization at 65°C will occur at different rates depending on a number of factors including the length and percent identity of the sequences which are hybridizing.

[0093] The nucleic acids (also referred to as polynucleotides) of this invention may include both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. They may be modified
30 chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the

naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids, etc.). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

[0094] The term "mutated" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. A nucleic acid sequence may be mutated by any method known in the art including but not limited to mutagenesis techniques such as "error-prone PCR" (a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. See, e.g., Leung, D.W., et al., *Technique*, 1, pp. 11-15 (1989) and Caldwell, R. C. & Joyce G.F., *PCR Methods Applic.*, 2, pp. 28-33 (1992)); and "oligonucleotide-directed mutagenesis" (a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. See, e.g., Reidhaar-Olson, J.F. & Sauer, R.T., et al., *Science*, 241, pp. 53-57 (1988)).

[0095] The term "vector" as used herein is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome (discussed in more detail below). Certain

vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., vectors having an origin of replication which functions in the host cell). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and are thereby replicated along with the host genome. Moreover, certain preferred vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors").

5
10
[0096] "Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

[0097] The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (e.g., ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

20
25
30
[0098] The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell that has been genetically engineered. A recombinant host cell includes a cell into which a recombinant vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may

occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein. A recombinant host cell may be an isolated cell or cell line grown in culture or may
5 be a cell which resides in a living tissue or organism. The term "host" refers to any organism or plant comprising one or more "host cells", or to the source of the "host cells".

[0099] Moreover, as used herein a "host cell which lacks endogenous CMP-Sia" refers to a cell that does not endogeneously produce CMP-Sia, including cells
10 which lack a CMP-Sia pathway. As used herein a "fungal host cell" refers to a fungal host cell that lacks CMP-Sia.

[0100] The term "peptide" as used herein refers to a short polypeptide, e.g., one that is typically less than about 50 amino acids long and more typically less than about 30 amino acids long. The term as used herein encompasses analogs and
15 mimetics that mimic structural and thus biological function.

[0101] The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins, and fragments, mutants, homologs, variants, derivatives and analogs thereof. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different domains each of which
20 has one or more distinct activities.

[0102] The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) when it exists in a purity not found in nature, where purity can be adjudged with
25 respect to the presence of other cellular material (e.g., is free of other proteins from the same species) (3) is expressed by a cell from a different species, or (4) does not occur in nature (e.g., it is a fragment of a polypeptide found in nature or it includes amino acid analogs or derivatives not found in nature or linkages other than standard peptide bonds). Thus, a polypeptide that is chemically synthesized or
30 synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally

associated components by isolation, using protein purification techniques well known in the art. As thus defined, "isolated" does not necessarily require that the protein, polypeptide, peptide or oligopeptide so described has been physically removed from its native environment.

5 [0103] The term "polypeptide fragment" as used herein refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments
10 typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

[0104] A "recombinant protein", "recombinant glycoprotein" or "recombinant
15 enzyme" refers to a protein, glycoprotein or enzyme (respectively) produced by genetic engineering. A recombinant protein, glycoprotein or enzyme includes a heterologous protein, glycoprotein or enzyme (respectively) expressed from a nucleic acid which has been introduced into a host cell.

[0105] A "modified derivative" or a "derivative" refers to polypeptides or
20 fragments thereof that are substantially homologous in primary structural sequence but which include, e.g., *in vivo* or *in vitro* chemical and biochemical modifications or which incorporate amino acids that are not found in the native polypeptide. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, e.g., with radionuclides,
25 and various enzymatic modifications, as will be readily appreciated by those well skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ¹²⁵I, ³²P, ³⁵S, and ³H, ligands which bind to labeled antiligands (e.g., antibodies), fluorophores, chemiluminescent agents,
30 enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation.

Methods for labeling polypeptides are well known in the art. See Ausubel et al., 1992, hereby incorporated by reference.

[0106] The term "fusion protein" refers to a polypeptide comprising a polypeptide or fragment coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

[0107] The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic". See, e.g., Jones, (1992) *Amino Acid and Peptide Synthesis*, Oxford University Press; Jung, (1997) *Combinatorial Peptide and Nonpeptide Libraries: A Handbook*, John Wiley; Bodanszky et al. (1993), *Peptide Chemistry -- A Practical Textbook*, Springer Verlag; "Synthetic Peptides: A Users Guide", G.A. Grant, Ed, W.H., Freeman and Co. (1992); Evans et al. *J. Med. Chem.* 30:1229 (1987); Fauchere, *J. Adv. Drug Res.* 15:29 (1986); Veber and Freidinger, *TINS* p.392 (1985); and references cited in each of the above, which are incorporated herein by reference. Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides of the invention may be used to produce an equivalent effect and are therefore envisioned to be part of the invention.

[0108] A "polypeptide mutant" or "mutein" or "variant" refers to a polypeptide whose sequence contains an insertion, duplication, deletion, rearrangement or substitution of one or more amino acids compared to the amino acid sequence of a native or wild type protein. A mutein may have one or more amino acid point

substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both
5 the amino or carboxy termini. A mutein may have the same but preferably has a different biological activity compared to the naturally-occurring protein.

[0109] A mutein has at least 70% overall sequence homology to its wild-type counterpart. Even more preferred are muteins having 80%, 85% or 90% overall sequence homology to the wild-type protein. In an even more preferred
10 embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99% overall sequence identity. Sequence homology may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

[0110] Preferred amino acid substitutions are those which: (1) reduce
15 susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs.

[0111] As used herein, the twenty conventional amino acids and their
20 abbreviations follow conventional usage. See *Immunology – A Synthesis* (2nd Edition, E.S. Golub and D.R. Gren, Eds., Sinauer Associates, Sunderland, Mass. (1991)), which is incorporated herein by reference. Stereoisomers (e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α -, α -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino
25 acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ -carboxyglutamate, ϵ -N,N,N-trimethyllysine, ϵ -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (e.g.,
30 4-hydroxyproline). In the polypeptide notation used herein, the left-hand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

[0112] A protein has "homology" or is "homologous" to a second protein if the nucleic acid sequence that encodes the protein has a similar sequence to the nucleic acid sequence that encodes the second protein. Alternatively, a protein has homology to a second protein if the two proteins have "similar" amino acid sequences. (Thus, the term "homologous proteins" or "homologs" is defined to mean that the two proteins have similar amino acid sequences). In a preferred embodiment, a homologous protein is one that exhibits 50% sequence homology to the wild type protein, more preferred is 60% sequence homology. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence homology to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence identity. As used herein, homology between two regions of amino acid sequence (especially with respect to predicted structural similarities) is interpreted as implying similarity in function.

[0113] When "homologous" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of homology may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art (see, e.g., Pearson et al., 1994, herein incorporated by reference).

[0114] The following six groups each contain amino acids that are conservative substitutions for one another: 1) Serine (S), Threonine (T); 2) Aspartic Acid (D), Glutamic Acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

[0115] Sequence homology for polypeptides, which is also referred to as percent sequence identity, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group (GCG), University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisconsin 53705. Protein analysis software matches similar sequences using measure of homology assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. See, e.g., GCG Version 6.1.

[0116] A preferred algorithm when comparing an inhibitory molecule sequence to a database containing a large number of sequences from different organisms is the computer program BLAST (Altschul, S.F. et al. (1990) *J. Mol. Biol.* 215:403-410; Gish and States (1993) *Nature Genet.* 3:266-272; Madden, T.L. et al. (1996) *Meth. Enzymol.* 266:131-141; Altschul, S.F. et al. (1997) *Nucleic Acids Res.* 25:3389-3402; Zhang, J. and Madden, T.L. (1997) *Genome Res.* 7:649-656, especially blastp or tblastn (Altschul et al., 1997)). Preferred parameters for BLASTp are: Expectation value: 10 (default); Filter: seg (default); Cost to open a gap: 11 (default); Cost to extend a gap: 1 (default); Max. alignments: 100 (default); Word size: 11 (default); No. of descriptions: 100 (default); Penalty Matrix: BLOWSUM62.

[0117] The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences. Database searching using amino acid sequences can be measured by algorithms other than blastp known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA provides alignments and percent sequence identity of the regions of

the best overlap between the query and search sequences (Pearson, 1990, herein incorporated by reference). For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, 5 herein incorporated by reference.

[0118] "Specific binding" refers to the ability of two molecules to bind to each other in preference to binding to other molecules in the environment. Typically, "specific binding" discriminates over adventitious binding in a reaction by at least two-fold, more typically by at least 10-fold, often at least 100-fold. Typically, the 10 affinity or avidity of a specific binding reaction is at least about 10^{-7} M (e.g., at least about 10^{-8} M or 10^{-9} M).

[0119] The term "region" as used herein refers to a physically contiguous portion of the primary structure of a biomolecule. In the case of proteins, a region is defined by a contiguous portion of the amino acid sequence of that protein.

15 [0120] The term "domain" as used herein refers to a structure of a biomolecule that contributes to a known or suspected function of the biomolecule. Domains may be co-extensive with regions or portions thereof; domains may also include distinct, non-contiguous regions of a biomolecule. Examples of protein domains include, but are not limited to, an Ig domain, an extracellular domain, a 20 transmembrane domain, and a cytoplasmic domain.

[0121] As used herein, the term "molecule" means any compound, including, but not limited to, a small molecule, peptide, protein, sugar, nucleotide, nucleic acid, lipid, etc., and such a compound can be natural or synthetic.

[0122] As used herein, a "CMP-Sialic acid biosynthetic pathway" or a "CMP-Sia 25 biosynthetic pathway" refers to one or more glycosylation enzymes which results in the formation of CMP-Sia in a host.

[0123] As used herein, a "CMP-Sia pool" refers to a detectable level of cellular CMP-Sia.

[0124] As used herein, the term "N-glycan" refers to an N-linked 30 oligosaccharide, e.g., one that is attached by an asparagine-N-acetylglucosamine linkage to an asparagine residue of a polypeptide. N-glycans have a common pentasaccharide core of $\text{Man}_3\text{GlcNAc}_2$ ("Man" refers to mannose; "Glc" refers to

glucose; and "NAc" refers to N-acetyl; GlcNAc refers to N-acetylglucosamine). The term "trimannose core" used with respect to the N-glycan also refers to the structure $\text{Man}_3\text{GlcNAc}_2$ ("Man3"). N-glycans differ with respect to the number of branches (antennae) comprising peripheral sugars (e.g., GlcNAc, galactose and sialic acid) that are added to the Man_3 core structure. N-glycans are classified according to their branched constituents (e.g., high mannose, complex or hybrid).

[0125] A "high mannose" type N-glycan has five or more mannose residues. A "complex" type N-glycan typically has at least one GlcNAc attached to the 1,3 mannose arm and at least one GlcNAc attached to the 1,6 mannose arm of the trimannose core. Complex N-glycans may also have galactose ("Gal") residues that are optionally modified with sialic acid or derivatives ("NeuAc", where "Neu" refers to neuraminic acid and "Ac" refers to acetyl). A complex N-glycan typically has at least one branch that terminates in an oligosaccharide such as, for example: NeuAc-; NeuAc α 2-6GalNAc α 1-; NeuAc α 2-3Gal β 1-3GalNAc α 1-; NeuAc α 2-3/6Gal β 1-4GlcNAc β 1-; GlcNAc α 1-4Gal β 1-(mucins only); Fuc α 1-2Gal β 1-(blood group H). Sulfate esters can occur on galactose, GalNAc, and GlcNAc residues, and phosphate esters can occur on mannose residues. NeuAc (Neu: neuraminic acid; Ac:acetyl) can be O-acetylated or replaced by NeuGI (N-glycolylneuraminic acid). Complex N-glycans may also have intrachain substitutions comprising "bisecting" GlcNAc and core fucose ("Fuc"). A "hybrid" N-glycan has at least one GlcNAc on the terminal of the 1,3 mannose arm of the trimannose core and zero or more mannoses on the 1,6 mannose arm of the trimannose core.

[0126] The substrate UDP-GlcNAc is the abbreviation for UDP-*N*-acetylglucosamine. The intermediate ManNAc is the abbreviation for *N*-acetylmannosamine. The intermediate ManNAc-6-P is the abbreviation for *N*-acetylmannosamine-6-phosphate. The intermediate Sia-9-P is the abbreviation for sialate-9-phosphate. The intermediate Cytidine monophosphate-sialic acid is abbreviated as "CMP-Sia." Sialic acid is abbreviated as "Sia," "Neu5Ac," "NeuAc" or "NANA" herein.

[0127] As used herein, the term "sialic acid" refers to a group of molecules where the common molecule includes N-acetyl-5-neuraminic acid (Neu5Ac) having the basic 9-carbon neuraminic acid core modified at the 5-carbon position with an

attached acetyl group. Common derivatives of Neu5Ac at the 5-carbon position include: 2-keto-3-deoxy-d-glycero-d-galactononic acid (KDN) which possesses a hydroxyl group in place of the acetyl group; de-N-acetylation of the 5-N-acetyl group produces neuraminic (Neu); hydroxylation of the 5-N-acetyl group produces N-glycolylneuraminic acid (Neu5Gc). The hydroxyl groups at positions 4-, 7-, 8- and 9- of these four molecules (Neu5Ac, KDN, Neu and Neu5Gc) can be further substituted with O-acetyl, O-methyl, O-sulfate and phosphate groups to enlarge this group of compounds. Furthermore, unsaturated and dehydro forms of sialic acids are known to exist.

10 [0128] The gene encoding for the UDP-GlcNAc epimerase is abbreviated as "*NeuC*." The gene encoding for the sialate synthase is abbreviated as "*NeuB*." The gene encoding for the CMP-Sialate synthase is abbreviated as "*NeuA*."

[0129] Sialate aldolase is also commonly referred to as sialate lyase and sialate pyruvate-lyase. More specifically in *E. coli*, sialate aldolase is referred to as *NanA*.

15 [0130] The term "enzyme," when used herein in connection with altering host cell glycosylation, refers to a molecule having at least one enzymatic activity, and includes full-length enzymes, catalytically active fragments, chimerics, complexes, and the like.

[0131] A "catalytically active fragment" of an enzyme refers to a polypeptide
20 having a detectable level of functional (enzymatic) activity.

[0132] As used herein, the term "secretion pathway" refers to the assembly line of various glycosylation enzymes to which a lipid-linked oligosaccharide precursor and an N-glycan substrate are sequentially exposed, following the molecular flow of a nascent polypeptide chain from the cytoplasm to the endoplasmic reticulum
25 (ER) and the compartments of the Golgi apparatus. Enzymes are said to be localized along this pathway. An enzyme X that acts on a lipid-linked glycan or an N-glycan before enzyme Y is said to be or to act "upstream" to enzyme Y; similarly, enzyme Y is or acts "downstream" from enzyme X.

[0133] The term "polynucleotide" or "nucleic acid molecule" refers to a
30 polymeric form of nucleotides of at least 10 bases in length. The term includes DNA molecules (e.g., cDNA or genomic or synthetic DNA) and RNA molecules (e.g., mRNA or synthetic RNA), as well as analogs of DNA or RNA containing

non-natural nucleotide analogs, non-native internucleoside bonds, or both. The nucleic acid can be in any topological conformation. For instance, the nucleic acid can be single-stranded, double-stranded, triple-stranded, quadruplexed, partially double-stranded, branched, hairpinned, circular, or in a padlocked conformation.

5 The term includes single and double stranded forms of DNA. A nucleic acid molecule of this invention may include both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. They may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the
10 art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), pendent moieties (e.g.,
15 polypeptides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids, etc.) Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide
20 linkages substitute for phosphate linkages in the backbone of the molecule.

[0134] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Exemplary methods and materials are described below, although methods and materials similar or equivalent to those described
25 herein can also be used in the practice of the present invention and will be apparent to those of skill in the art. All publications and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. The materials, methods, and examples are illustrative only and not intended to be limiting.

30 [0135] Throughout this specification and claims, the word "comprise" or variations such as "comprises" or "comprising", will be understood to imply the

inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

Methods For Producing CMP-Sia For The
Generation Of Recombinant N-Glycans In Fungal Cells

5 [0136] The present invention provides methods for production of a functional
CMP-Sia biosynthetic pathway in a host cell which lacks endogenous CMP-Sia,
such as a fungal cell. The present invention also provides a method for creating a
host which has been modified to express a CMP-Sia pathway. The invention
further provides a method for creating a host cell which comprises a cellular pool
10 of CMP-Sia.

[0137] The methods involve the cloning and expression of several genes
encoding enzymes of the CMP-Sia biosynthetic pathway resulting in a cellular
pool of CMP-Sia which can be utilized in the production of sialylated glycans on
proteins of interest. In general, the addition of sialic acids to glycans requires the
15 presence of the sialyltransferase, a glycan acceptor (e.g.,
 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$) and the sialyl donor molecule, CMP-Sia. The
synthesis of the CMP-Sia donor molecule in higher organisms (e.g., mammals) is a
four enzyme, multiple reaction process starting with the substrate UDP-GlcNAc
and resulting in CMP-Sia (**Figure 1A**). The process initiates in the cytoplasm
20 producing sialic acid which is then translocated into the nucleus where Sia is
converted to CMP-Sia. Subsequently, CMP-Sia exits the nucleus into the
cytoplasm and is then transported into the Golgi where sialyltransferases catalyze
the transfer of sialic acid onto the acceptor glycan. In contrast, the bacterial
pathway for synthesizing CMP-Sia from UDP-GlcNAc involves only three
25 enzymes and two intermediates (**Figure 1B**), with all reactions occurring in the
cytoplasm.

[0138] Accordingly, the methods of the invention involve generating a pool of
CMP-Sia in a non-human host cell which lacks endogenous CMP-Sia by
introducing a functional CMP-Sia biosynthetic pathway. With readily available
30 DNA sequence information from genetic databases (e.g., GenBank, Swissprot),
enzymes and/or activities involved in the CMP-Sia pathways (**Example 1**) are
cloned. Using standard techniques known to those skilled in the art, nucleic acid

molecules encoding enzymes (or catalytically active fragments thereof) involved in the biosynthesis of CMP-Sia are inserted into appropriate expression vectors under the transcriptional control of promoters and/or other expression control sequences capable of driving transcription in a selected host cell of the invention (e.g., a
5 fungal host cell). The functional expression of such enzymes in the selected host cells of the invention can be detected. In one embodiment, the functional expression of such enzymes in the selected host cells of the invention can be detected by measuring the intermediate formed by the enzyme. The methods of the invention are not limited to the use of the specific enzyme sources disclosed
10 herein.

Engineering A Mammalian CMP-Sialic Acid Biosynthetic Pathway In Fungi

[0139] In one aspect of the invention, a method for synthesizing a mammalian CMP-sialic acid pathway in a host cell which lacks endogenous CMP-Sia is provided. In mammals and higher eukaryotes, synthesis of CMP-sialic acid is
15 initiated in the cytoplasm where the enzyme activities (UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase, N-acetylneuraminate-9-phosphate synthase, N-acetylneuraminate-9-phosphatase) convert UDP-GlcNAc to sialic acid (**Figure 1A**). The sialic acid then enters the nucleus where it is converted to CMP-sialic acid by CMP-sialic acid synthase.

20 [0140] In one embodiment of the invention, the method involves cloning several genes encoding enzymes in the CMP-Sia biosynthetic pathway, including UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase, N-acetylneuraminate-9-phosphate synthase, N-acetylneuraminate-9-phosphatase, and CMP-sialic acid synthase, in a host cell which lacks endogenous CMP-Sia, such as
25 a fungal host cell. The genes are expressed to generate each enzyme, producing intermediates that are used for subsequent enzymatic reactions. **Examples 5-8** describe methods for the introduction of these enzymes into a fungal host (e.g., *P. pastoris*) using a selection marker. Alternatively, the enzymes are expressed together to produce or increase downstream intermediates whereby subsequent
30 enzymes are able to act upon them.

[0141] The first enzyme in the pathway is a bi-functional enzyme that is both an UDP-GlcNAc epimerase and an N-acetylmannosamine kinase, converting UDP-

GlcNAc through N-acetylmannosamine (ManNAc) to N-acetylmannosamine-6-phosphate (ManNAc-6-P) (Hinderlich, S., Stasche, R., et al. 1997). This enzyme was originally cloned from a rat liver cDNA library (Stasche, R., Hinderlich, S., et al. 1997). In a preferred embodiment, a gene encoding the functional UDP-N-acetylglucosamine-2-epimerase enzyme, including homologs, variants and derivatives thereof, is cloned and expressed in a non-human host cell which lacks endogenous CMP-Sia, such as a fungal host cell. In another preferred embodiment, a gene encoding the functional N-acetylmannosamine kinase enzyme, including homologs, variants and derivatives thereof, is cloned and expressed in a host cell, such as a fungal host cell. In a more preferred embodiment, a gene encoding the bifunctional UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase enzyme, including homologs, variants and derivatives thereof, is cloned and expressed in a non-human host cell which lacks endogenous CMP-Sia, such as a fungal host cell (e.g., *P. pastoris*). The functional expression of these genes can be detected using a functional assay. In one embodiment, the functional expression of such genes can be detected by detecting the formation of ManNAc and ManNAc-6-P intermediates.

[0142] The second enzyme in the pathway, N-acetylneuraminic acid phosphate synthase, was cloned from human liver based on its homology to the *E. coli* sialic acid synthase gene, *NeuB* (Lawrence, S.M., Huddleston, K.A., et al. 2000). This enzyme catalyzes the conversion of ManNAc-6-P to sialate 9-phosphate (also referred to as Sia-9P, N-acetylneuraminate 9-phosphate, or Neu5Ac-9P). Accordingly, in a preferred embodiment, a gene encoding the functional N-acetylneuraminate 9-phosphate synthase enzyme, including homologs, variants and derivatives thereof, is cloned and expressed in a non-human host cell which lacks endogenous CMP-Sia, such as a fungal host cell. The functional expression N-acetylneuraminic acid phosphate synthase in the host can be detected using a functional assay. In one embodiment, the functional expression of N-acetylneuraminic acid phosphate synthase can be detected by detecting the formation of Sia-9P.

[0143] The third enzyme in the pathway, N-acetylneuraminate 9-phosphatase (Sia-9-phosphatase), has yet to be cloned but is involved in the conversion of Sia-

9-P to sialic acid. Although the activity of this enzyme has been detected in mammalian cells, no such activity has been identified in fungal cells. Therefore, the lack of Sia-9-phosphatase would cause a break in the pathway. Accordingly, in a preferred embodiment, the method of the present invention involves isolating and cloning a Sia-9-phosphatase gene into a non-human host cell, such as a fungal host cell. Such hosts include yeast, fungal, insect and bacterial cells. In a more preferred embodiment, the Sia-9-phosphatase gene, including homologs, variants and derivatives thereof, is expressed in a non-human host cell which lacks endogenous CMP-Sia, such as a fungal host. The functional expression of Sia-9-phosphatase in the host can be detected using a functional assay. In one embodiment, the functional expression of Sia-9-phosphatase can be detected by detecting the formation of sialic acid.

[0144] The next enzyme in the mammalian pathway, CMP-Sia synthase, was originally cloned from the murine pituitary gland by functional complementation of a cell line deficient in this enzyme (Munster, A.K., Eckhardt, M., et al. 1998). This enzyme converts sialic acid to CMP-Sia, which is the donor substrate in a sialyltransferase reaction in the Golgi. Accordingly, in an even more preferred embodiment, a gene encoding the functional CMP-Sia synthase enzyme, including homologs, variants and derivatives thereof, is cloned and expressed in a non-human host cell which lacks endogenous CMP-Sia, such as a fungal host cell. The functional expression of CMP-Sia synthase in the host can be detected using a functional assay. In one embodiment, the functional expression of CMP-Sia synthase can be detected by detecting the formation of CMP-Sia.

[0145] The method of the present invention further involves the production of the intermediates produced in a non-human host as a result of expressing the above enzymes in the CMP-Sia pathway. Preferably, the intermediates produced include one or more of the following: UDP-GlcNAc, ManNAc, ManNAc-6-P, Sia-9-P, Sia and CMP-Sia. Additionally, each intermediate produced by the enzymes is preferably detected. For example, to detect the presence or absence of an intermediate, an assay as described in **Example 10** is used. Accordingly, the method also involves assays to detect the N-glycan intermediates produced in a non-human host cell which lacks endogenous CMP-Sia, such as a fungal host cell.

[0146] A skilled artisan recognizes that the mere availability of one or more enzymes in the CMP-sialic acid biosynthetic pathway does not suggest that such enzymes can be functionally expressed in a host cell which lacks endogenous CMP-Sia, such as a fungal host cell. To date, the ability of such host cell to
5 express these mammalian enzymes to create a functional *de novo* CMP-Sia biosynthetic pathway has not been described. The present invention provides for the first time the functional expression of at least one mammalian enzyme involved in CMP-Sia biosynthesis in a fungal host: the mouse CMP-Sia synthase (**Example 8**), suggesting that production of CMP-Sia via the mammalian pathway (in whole
10 or in part) is possible in a fungal host and in other non-human hosts which lack endogenous CMP-Sia.

[0147] The invention described herein is not limited to the use of the specific enzymes, genes, plasmids and constructs disclosed herein. A person of skill could use any homologs, variants and derivatives of the genes involved in the synthesis
15 of CMP-Sia.

[0148] To produce sialylated, recombinant glycoproteins in a non-human host cell which lacks endogenous CMP-Sia (e.g., a fungal host such as *P. pastoris*), the above mentioned mammalian enzymes can be expressed using a combinatorial DNA library as disclosed in WO 02/00879, generating a pool of CMP-Sia, which
20 is transferred onto galactosylated N-glycans in the presence of a sialyltransferase. Accordingly, the present invention provides a method for engineering a CMP-Sia biosynthetic pathway into a fungal host by expressing each of the enzymes such that they function, preferably so that they function optimally, in the fungal host. Mammalian, bacterial or hybrid engineered CMP-Sia biosynthetic pathways are
25 provided.

Engineering A Bacterial CMP-Sialic Acid Biosynthetic Pathway In Fungi

[0149] The metabolic intermediate UDP-GlcNAc is common to eukaryotes and prokaryotes, providing an endogenous substrate from which to initiate the synthesis of CMP-Sia (**Figure 1**). Based on the presence of this common
30 intermediate, the CMP-Sia biosynthetic pathway can be engineered into non-human host cells which lack endogenous CMP-Sia by integrating the genes encoding the bacterial UDP-GlcNAc epimerase, sialate synthase and CMP-Sia

synthase. Accordingly, another aspect of the present invention involves engineering the bacterial CMP-Sia biosynthetic pathway into host cells which lack an endogenous CMP-Sia pathway. The expression of bacterial *Neu* genes in cells which lack an endogenous CMP-Sia biosynthetic pathway enables the generation
5 of a cellular CMP-Sia pool, which can subsequently facilitate the production of recombinant N-glycans having detectable level of sialylation on a protein of interest, such as recombinantly expressed glycoproteins. The bacterial enzymes involved in the synthesis of CMP-Sia include UDP-GlcNAc epimerase (*NeuC*), sialate synthase (*NeuB*) and CMP-Sia synthase (*NeuA*). In one embodiment, the
10 *NeuC*, *NeuB*, and *NeuA* genes which encode these functional enzymes, respectively, including homologs, variants and derivatives thereof, are cloned and expressed in non-human host cells which lack an endogenous CMP-Sia pathway, such as a fungal host. The sequences of *NeuC*, *NeuB* and *NeuA* genes are shown in **Figures 2-4**, respectively. The expression of these genes generates the
15 intermediate molecules in the biosynthetic pathway of CMP-sialic acid (**Figure 1B**).

[0150] In addition to these three enzymes, the method for synthesizing the bacterial CMP-Sia biosynthetic pathway from UDP-GlcNAc involves generating two intermediates: ManNAc and Sia (**Figure 1B**). The conversion of UDP-
20 GlcNAc to ManNAc is facilitated by the *NeuC* gene. The conversion of ManNAc to Sia is facilitated by the *NeuB* gene and the conversion of substrates Sia to CMP-Sia is facilitated by the *NeuA* gene. These three enzymes (or homologs thereof) have thus far been found together in pathogenic bacteria -- i.e., not one of the genes has not been found without the other two. In comparison to the mammalian
25 pathway, the introduction of the bacterial pathway into a host, such as a fungal host, requires the manipulation of fewer genes.

[0151] The *E. coli* UDP-GlcNAc epimerase, encoded by the *E. coli NeuC* gene, is the first enzyme involved in the bacterial synthesis of polysialic acid (Ringenberg, M., Lichtensteiger, C., et al. 2001). The *NeuC* gene (Genbank:
30 M84026.1; SEQ ID NO:13) encoding this enzyme was isolated from the pathogenic *E. coli* K1 strain and encodes a protein of 391 amino acids (SEQ ID NO:14) (**Figure 2**) (Zapata, G., Crowley, J.M., et al. 1992). The encoded UDP-

GlcNAc epimerase catalyzes the conversion of UDP-GlcNAc to ManNAc. Homologs of this enzyme have been identified in several pathogenic bacteria, including *Streptococcus agalactiae*, *Synechococcus* sp. WH 8102, *Clostridium thermocellum*, *Vibrio vulnificus*, *Legionella pneumophila*, and *Campylobacter jejuni*. In one embodiment, a gene encoding the functional *E. coli* UDP-GlcNAc epimerase enzyme (*NeuC*), including homologs, variants and derivatives thereof, is cloned and expressed in a non-human host cell, such as a fungal host. The functional expression of *NeuC* in the host can be detected using a functional assay. In one embodiment, the functional expression *NeuC* can be detected by detecting the formation of ManNAc.

[0152] The second enzyme in the bacterial pathway is sialate synthase which directly converts ManNAc to Sia, bypassing several enzymes and intermediates present in the mammalian pathway. This enzyme of 346 amino acids (SEQ ID NO:16), is encoded by the *E. coli NeuB* gene (Genbank: U05248.1; SEQ ID NO:15) (**Figure 3**) (Annunziato, P.W., Wright, L.F., et al. 1995). In another embodiment, a gene encoding a functional *E. coli* sialate synthase enzyme (*NeuB*), including homologs, variants and derivatives thereof, is cloned and expressed in a non-human host cell, such as a fungal host cell. The functional expression of *NeuB* in the host can be detected using a functional assay. In one embodiment, the functional expression *NeuB* can be detected by detecting the formation of Sia.

[0153] The third enzyme in this bacterial pathway is CMP-Sia synthase, consisting of 419 amino acids (SEQ ID NO:18) and encoded by the *E. coli NeuA* gene (Genbank: J05023; SEQ ID NO:17) (**Figure 4**). CMP-Sia synthase converts Sia to CMP-Sia (Zapata, G., Vann, W.F., et al. 1989). The *NeuA* gene is found in the same organisms as the *NeuC* and *NeuB* genes. Accordingly, in yet another embodiment, a gene (*NeuA*) encoding a functional *E. coli* CMP-Sia synthase enzyme, including homologs, variants and derivatives thereof, is cloned and expressed in a non-human host cell, such as a fungal host cell. In one embodiment, the functional expression *NeuA* can be detected by detecting the formation of CMP-Sia.

[0154] In yet another embodiment, the gene encoding a functional bacterial CMP-Sia synthase (e.g. *NeuA*) encodes a fusion protein comprising a: catalytic

domain having the activity of a bacterial CMP-Sia synthase and a cellular targeting signal peptide (not normally associated with the catalytic domain) selected to target the enzyme to the nucleus of the host cell. In one embodiment, said cellular targeting signal peptide comprises a domain of the SV40 capsid polypeptide VP1.

5 In another embodiment, the signal peptide comprises one or more endogenous signaling motifs from a mammalian CMP-Sia synthase that ensure correct localization of the enzyme to the nucleus. The methods of making said fusion protein are well known in the art.

[0155] After PCR amplification of the *E. coli NeuA*, *NeuB* and *NeuC* genes, the
10 amplified fragments were ligated into a selectable yeast integration vector under the control of a promoter (**Example 2**). After transforming a host strain (e.g., *P. pastoris*), with each vector carrying the *Neu* gene fragments, colonies were screened by applying positive selection. These transformants were grown in YPD media. An assay for *Neu* gene enzymatic activity is carried out after each
15 transformation. The ability of a non-human host which lacks endogenous sialylation to express the bacterial enzymes involved in creating a *de novo* CMP-Sia biosynthetic pathway is provided for the first time herein.

Engineering A Hybrid Mammalian/Bacterial CMP-Sialic Acid Biosynthetic Pathway In Fungi

20 [0156] Both mammalian and bacterial CMP-Sia biosynthetic pathways require that both CTP and sialic acid be available to the CMP-Sia synthase. Although similar in enzymatic function to the corresponding bacterial enzyme, the mammalian CMP-Sia synthase may include one or more endogenous signaling motifs that ensure correct localization to the nucleus. Because eukaryotes have a
25 nucleus-localized pool of CTP and the prokaryotic CMP-Sia synthase may not localize to this compartment, a hybrid CMP-Sia biosynthetic pathway combining both mammalian and bacterial enzymes is a preferred method for the production of sialic acid and its intermediates in a non-human host cell, such as a fungal host cell. To this end, a pathway can be engineered into the host cell which involves
30 the integration of both *NeuC* and *NeuB* as well as a mammalian CMP-Sia synthase. The CMP-Sia synthase enzyme may be selected from several mammalian homologs that have been cloned and characterized (Genbank: AJ006215; SEQ ID

NO:19) (Munster, A.K., Eckhardt, M., et al. 1998) (see e.g., the murine CMP-Sia synthase) (Figure 5). Preferably, the host cell is transformed with UDP-GlcNAc epimerase (*E. coli NeuC*) and sialate synthase (*E. coli NeuB*) in combination with the mouse CMP-Sia synthase. The host engineered with this hybrid CMP-Sia biosynthetic pathway produces a cellular pool of the donor molecule CMP-Sia (Figure 12). In a more preferred embodiment, the combination of the enzymes expressed in the host enhances production of the donor molecule CMP-Sia.

Engineering Enzymes Involved In Alternative Routes For Enhancing The Production Of CMP-Sialic Acid Pathway Intermediates In Fungi

10 [0157] In yet another aspect of the invention, enzymes involved in alternate pathways of CMP-sialic acid biosynthesis are engineered into non-human host cells, such as fungal host cells. For example, it is contemplated that when an intermediate becomes limiting during one of the methods outlined above, the introduction of an enzyme that uses an alternate mechanism to produce that intermediate will serve as a sufficient substitute in the production of CMP-sialic acid, or any intermediate along this pathway. Embodiments are described herein for the production of the intermediates ManNAc and Sia, though this approach may be extended to produce other intermediates. Furthermore, any of these enzymes can be incorporated into either the mammalian, bacterial or hybrid pathways, either in the absence of the enzymes mentioned previously (i.e., enzymes producing the same intermediate) or in the presence of enzymes mentioned previously, i.e., to enhance overall production.

25 [0158] In the above mentioned embodiments, ManNAc is produced from UDP-GlcNAc by either the mammalian enzyme UDP-GlcNAc-2-epimerase/ManNAc kinase or by the bacterial enzyme *NeuC*. The substrate for this reaction, UDP-GlcNAc, is predicted to be present in sufficient quantities in cells for the synthesis of CMP-Sia due to its requirement in producing several classes of molecules, including endogenous N-glycans. However, if ManNAc does become limiting -- potentially due to the increased demand for ManNAc from the sialic acid biosynthetic pathway -- then the cellular supply of ManNAc may be increased by introducing a GlcNAc epimerase which reacts with the substrate GlcNAc to produce ManNAc.

[0159] Accordingly, in one embodiment, a gene encoding a functional GlcNAc epimerase enzyme, including homologs, variants and derivatives thereof, is cloned and expressed in a host cell, such as a fungal host cell. Using GlcNAc epimerase to directly convert GlcNAc to ManNAc is a shorter, more efficient approach compared with the two-step process involving the synthesis of UDP-GlcNAc (Figure 6). The GlcNAc epimerase is readily available and, to date, the only confirmed GlcNAc epimerase to have been cloned is from the pig kidney (Maru, I., Ohta, Y., et al. 1996) (Example 3). The gene (Genbank: D83766; SEQ ID NO: 21) isolated from pig kidney encodes a protein of 402 amino acids (SEQ ID NO:22) (Figure 7). When this enzyme was cloned, it was found to be identical to the pig renin-binding protein cloned previously (Inoue, H., Fukui, K., et al. 1990). Although this is the only protein with confirmed GlcNAc epimerase activity, several other renin-binding proteins have been isolated from other organisms, including humans, mouse, rat and bacteria, among others. All are shown to have significant homology. For example, the human GlcNAc epimerase homolog (Genbank: D10232.1) has 87% identity and 92% similarity to the pig GlcNAc epimerase protein. Although these homologs are very similar in sequence, the pig protein is the only one having demonstrable epimerase activity to date. The methods of the invention could be performed using any gene encoding a functional GlcNAc epimerase activity. Based on the presence of the activity of GlcNAc epimerase, the cloning and expression of this gene in a non-human host cell, such as a fungal host cell, is predicted to enhance the cellular levels of ManNAc, thereby, providing sufficient substrate for the enzymes that utilize ManNAc in the CMP-sialic acid biosynthetic pathway.

[0160] In another embodiment, sialate aldolase is used to increase cellular levels of sialic acid, as illustrated in Figure 8. This enzyme (also known as sialate lyase and sialate pyruvate-lyase) directly catalyzes the reversible reaction of ManNAc to sialic acid. In the presence of low concentrations of Sia, this enzyme catalyzes the condensation of ManNAc and pyruvate to produce Sia. Conversely, when Sia concentrations are high, the enzyme causes the reverse reaction to proceed, producing ManNAc and pyruvate (Vimr, E.R. and Troy, F.A. 1985). In the above embodiments, the presence of CMP-Sia synthase converts substantially all Sia to

CMP-Sia, thus shifting the equilibrium of the aldolase to the condensation of ManNAc and pyruvate to produce Sia. Preferably, the sialate aldolase used in this embodiment is expressed from the *E. coli* *NanA* gene, but the invention is not limited to this enzyme source. The gene (Genbank: X03345; SEQ ID NO:23) for this enzyme encodes a 297 amino acid protein (SEQ ID NO:24) (**Figure 9**) (Ohta, Y., Watanabe, K. et al. 1985). Close homologs to this enzyme are found in many pathogenic bacteria, including, *Salmonella typhimurium*, *Staphylococcus aureus*, *Clostridium perfringens*, *Haemophilus influenzae* among others. In addition, homologs are also present in mammals, including mice and humans. Cloning a gene encoding a sialate aldolase activity and expressing it in a fungal host cell enhances the cellular levels of Sia, thereby providing sufficient substrate for the enzymes that utilize Sia in the CMP-sialic acid biosynthetic pathway (**Example 4**).

Regulation Of CMP-Sialic Acid Synthesis:
Feedback Inhibition And Inducible Promoters

[0161] In mammalian cells, the production of CMP-sialic acid is highly regulated. CMP-sialic acid acts as a feedback inhibitor, acting on UDP-GlcNAc epimerase/ManNAc kinase to prevent further production of CMP-Sia (Hinderlich, S., Stasche, R., et al. 1997) (Keppler, O.T., Hinderlich, S. et al., 1999). In contrast, the bacterial CMP-Sia biosynthetic pathway (**Figure 1B**) does not appear to have a feedback inhibitory control mechanism that would limit the production of CMP-Sia (Ringenberg, M., Lichtensteiger, C. et al. 2001). However, incorporation of the *E. coli* sialate aldolase into one of the pathways mentioned above could cause a shift in the direction of the reaction that it catalyzes, depending on the balance of the equilibrium, thus potentially causing hydrolysis of Sia back to ManNAc.

Accordingly, the methods involving sialate aldolase as outlined above will prevent this reverse reaction from occurring, given the presence of CMP-sialate synthase which rapidly converts Sia to CMP-Sia.

[0162] The embodiments described thus far have detailed the constitutive over-expression of the enzymes in a particular biosynthetic pathway of CMP-Sia.

Though no literature is currently available that suggests that the presence of any of the mentioned intermediates, and/or the final product could be detrimental to a non-human host, such as a fungal host, a preferred embodiment of the invention

has one or more of the enzymes under the control of a regulatable (e.g., an inducible) promoter. In this embodiment, the gene (or ORF) encoding the protein of interest (including but not limited to: UDP-GlcNAc 2-epimerase/ManNAc kinase, *NeuC*, and GlcNAc epimerase) is cloned downstream of an inducible promoter (including but not limited to: the alcohol oxidase promoter (AOX1 or AOX2; Tschopp, J.F., Brust, P.F., et al. 1987), galactose-inducible promoter (GAL10; Yocum, R.R., Hanley, S., et al. 1984), tetracycline-inducible promoter (TET; Belli, G., Gari, E., et al. 1998)) to facilitate the controlled expression of that enzyme, and thus regulate the production of CMP-Sia.

10 Detection Of CMP-Sialic Acid And The Intermediate Compounds In Its Synthesis

[0163] The methods of the present invention provide engineered pathways to produce a cellular pool of CMP-Sia in non-human host cells which lack an endogenous CMP-Sia biosynthetic pathway. To assess the production of each intermediate in the pathway, these intermediates must be detectable. Accordingly, the present invention also provides a method for detecting such intermediates. A method for detecting a cellular pool of CMP-Sia, for example, is provided in **Example 10**. Currently, the literature describes only a few methods for measuring cellular CMP-Sia and its precursors. Early methods involved paper chromatography and thiobarbituric acid analysis and were found to be complicated and time consuming (Briles, E.B., Li, E., et al. 1977) (Harms, E., Kreisel, W., et al. 1973). HPLC (high pressure liquid chromatography) has also been used, though earlier methods employed acid elution resulting in the rapid hydrolysis of the CMP-Sia (Rump, J.A., Phillips, J., et al. 1986). Most recently, a more robust method has been described using high-performance anion-exchange chromatography using an alkaline elution protocol combined with pulsed amperometric detection (HPAEC-PAD) (Fritsch, M., Geilen, C.C., et al. 1996). This method, in addition to detecting CMP-Sia, can also detect the precursor sialic acid, thus being useful for confirming cellular synthesis of either or both of these compounds.

Codon Optimization And Nucleotide Substitution

[0164] The methods of the invention may be performed in conjunction with optimization of the base composition for efficient transcription/translation of the encoded protein in a particular host, such as a fungal host. For example, because
5 the Neu genes introduced into a fungal host are of bacterial origin, it may be necessary to optimize the base pair composition. This includes codon optimization to ensure that the cellular pools of tRNA are sufficient. The foreign genes (ORFs) may contain motifs detrimental to complete transcription/translation in the fungal host and, thus, may require substitution to more amenable sequences. The
10 expression of each introduced protein can be followed both at the transcriptional and translational stages by well known Northern and Western blotting techniques, respectively (Sambrook, J. and Russell, D.W., 2001).

Vectors

[0165] In another aspect, the present invention provides vectors (including
15 expression vectors), comprising genes encoding activities which promote the CMP-Sia biosynthetic pathway, a promoter, a terminator, a selectable marker and targeting flanking regions. Such promoters, terminators, selectable markers and flanking regions are readily available in the art. In a preferred embodiment, the promoter in each case is selected to provide optimal expression of the protein
20 encoded by that particular ORF to allow sufficient catalysis of the desired enzymatic reaction. This step requires choosing a promoter that is either constitutive or inducible, and provides regulated levels of transcription. In another embodiment, the terminator selected enables sufficient termination of transcription. In yet another embodiment, the selectable markers used are unique to each ORF to
25 enable the subsequent selection of a fungal strain that contains a specific combination of the ORFs to be introduced. In a further embodiment, the locus to which each fusion construct (encoding promoter, ORF and terminator) is localized, is determined by the choice of flanking region. The present invention is not limited to the use of the vectors disclosed herein.

Integration Sites

[0166] The integration of multiple genes into the chromosome of the host cell is likely required and involves a thoughtful strategy. The engineered strains are transformed with a range of different genes, and these genes are transformed in a stable fashion to ensure that the desired activity is maintained throughout the fermentation process. Any combination of the previously mentioned enzyme activities will have to be engineered into the host. In addition, a number of genes which encode enzymes known to be characteristic of non-human glycosylation reactions will need to be deleted from the non-human host cell. Genes which encode enzymes known to be characteristic of non-human glycosylation reactions in fungal hosts and their corresponding proteins have been extensively characterized in a number of lower eukaryotes (e.g., *Saccharomyces cerevisiae*, *Trichoderma reesei*, *Aspergillus nidulans*, *P. pastoris*, etc.), thereby providing a list of known glycosyltransferases in lower eukaryotes, their activities and their respective genetic sequence. These genes are likely to be selected from the group of mannosyltransferases e.g., 1,3 mannosyltransferases (e.g., *MNN1* in *S. cerevisiae*) (Graham, T. and Emr, S. 1991), 1,2 mannosyltransferases (e.g., the KTR/KRE family from *S. cerevisiae*), 1,6 mannosyltransferases (*OCH1* from *S. cerevisiae*), mannosylphosphate transferases and their regulators (*MNN4* and *MNN6* from *S. cerevisiae*) and additional enzymes that are involved in aberrant (i.e. non-human) glycosylation reactions.

[0167] Genes that encode enzymes that are undesirable serve as potential integration sites for genes that are desirable. For example, 1,6 mannosyltransferase activity is a hallmark of glycosylation in many known lower eukaryotes. The gene encoding α -1,6 mannosyltransferase (*OCH1*) has been cloned from *S. cerevisiae* (Chiba et al., 1998) as well as the initiating 1,6 mannosyltransferase activity in *P. pastoris* (WO 02/00879) and mutations in the gene produce a viable phenotype with reduced mannosylation. The gene locus encoding α -1,6 mannosyltransferase activity is, therefore, a prime target for the integration of genes encoding glycosyltransferase activity. Similarly, one can choose a range of other chromosomal integration sites resulting in a gene disruption event that is expected to: (1) improve the cells ability to glycosylate in a more human-like fashion, (2)

improve the cells ability to secrete proteins, (3) reduce proteolysis of foreign proteins and (4) improve other characteristics of the process that facilitate purification or the fermentation process itself.

Host Cell Production Strain

5 [0168] A host cell which lacks an endogenous CMP-Sia biosynthetic pathway and which expresses a functional CMP-Sia biosynthetic pathway is provided. In one embodiment, a fungal host cell which expresses a functional CMP-Sia biosynthetic pathway is provided. Preferably, the host produces a cellular pool of CMP-Sia that may be used as a donor molecule in the presence of a
10 sialyltransferase and a glycan acceptor (e.g., Gal₂GlcNAc₂Man₃GlcNAc₂) in a sialylation reaction. Using the methods of the invention, a variety of different hosts producing CMP-Sia may be generated. Preferably, robust protein production strains of fungal hosts that are capable of performing well in an industrial fermentation process are selected. These strains, which produce acceptor glycans,
15 for example, that are galactosylated include, without limitation: *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia sp.*, *Saccharomyces cerevisiae*,
20 *Saccharomyces sp.*, *Hansenula polymorpha*, *Kluyveromyces sp.*, *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum* and *Neurospora crassa*. Preferably, the modified strains of the present invention are used to produce human-like sialylated
25 glycoproteins according to the methods provided in WO 02/00879, WO 03/056914 and US2004/0018590, (each of which is hereby incorporated by reference in its entirety).

Therapeutic Proteins

30 [0169] The fungal host strains produced according to methods of the present invention combined with the teachings described in WO 02/00879, WO 03/056914 and US2004/0018590, produce high titers of heterologous therapeutic proteins in

which a wide variety of sialylated glycans on a protein of interest, such as a recombinant protein, is generated in a host which lacks endogenous CMP-Sia, such as a fungal host, including without limitation: erythropoietin, cytokines such as interferon- α , interferon- β , interferon- γ , interferon- ω , TNF- α , granulocyte-CSF, GM-CSF, interleukins such as IL-1ra, coagulation factors such as factor VIII, factor IX, human protein C, antithrombin III and thrombopoietin, antibodies; IgG, IgA, IgD, IgE, IgM and fragments thereof, Fc and Fab regions, soluble IgE receptor α -chain, urokinase, chymase, and urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, FSH, annexin V fusion protein, angiostatin, vascular endothelial growth factor-2, myeloid progenitor inhibitory factor-1, osteoprotegerin, α -1 antitrypsin, DNase II, α -feto proteins and glucocerebrosidase. These and other sialylated glycoproteins are particularly useful for therapeutic administration.

[0170] The following are examples which illustrate the compositions and methods of this invention. These examples should not be construed as limiting: the examples are included for the purposes of illustration only.

EXAMPLE 1

Cloning Enzymes Involved In CMP-Sialic Acid Synthesis

[0171] One method for cloning a CMP-sialic acid biosynthetic pathway into a fungal host cell involves amplifying the *E. coli NeuA*, *NeuB* and *NeuC* genes from *E. coli* genomic DNA using the polymerase chain reaction in conjunction with primer pairs specific for each open reading frame (ORF) (Table 1, below and Figures 4, 3 and 2, respectively).

[0172] For cloning a mammalian CMP-sialic acid biosynthetic pathway, the mouse CMP-Sia synthase ORF (Figure 5) was amplified from a mouse pituitary cDNA library in conjunction with the primer pairs set forth in Table 1. The GlcNAc epimerase (previously discussed in an alternate method for producing CMP-Sia intermediates), was amplified from porcine cDNA using PCR in conjunction with primer pairs specific for the corresponding gene (Table 1 and Figure 7). The sialate aldolase gene (Figure 9) was amplified from *E. coli* genomic DNA using the polymerase chain reaction in conjunction with the primer

pairs set forth in **Table 1**. The mouse bifunctional UDP-N-acetylglucosamine-2-Epimerase/N-acetylmannosamine kinase gene was amplified from mouse liver using the polymerase chain reaction in conjunction with the primer pairs set forth in **Table 1**. The mouse N-acetylneuraminase-9-phosphate synthase gene was amplified from mouse liver using the polymerase chain reaction in conjunction with the primer pairs set forth in **Table 1**. The human CMP-Sia synthase gene was amplified from human liver using the polymerase chain reaction in conjunction with the primer pairs set forth in **Table 1**. In each case, the ORFs were amplified using a high-fidelity DNA polymerase enzyme under the following thermal cycling conditions: 97°C for 1 min, 1 cycle; 97°C for 20 sec, 60°C for 30 sec, 72°C for 2 min, 25 cycles; 72°C for 2min, 1 cycle. Following DNA sequencing to confirm the absence of mutations, each ORF is re-amplified using primers containing compatible restriction sites to facilitate the subcloning of each into suitable fungal expression vectors.

15

Table 1:

Primer name	Primer sequence
<i>NeuA</i> sense	5'-ATGAGAACAAAAATTATTGCGATAATTCCAGCCCG-3' (SEQ ID NO:1)
<i>NeuA</i> antisense	5'-TCATTTAACAACTCTCCGCTATTTTCGTTTTTC-3' (SEQ ID NO:2)
<i>NeuB</i> sense	5'-ATGAGTAATATATATATATCGTTGCTGAAATTGGTTG-3' (SEQ ID NO:3)
<i>NeuB</i> antisense	5'-TTATTCCCCCTGATTTTTGAATTCGCTATG-3' (SEQ ID NO:4)
<i>NeuC</i> sense	5'-ATGAAAAAATATTATACGTAAC TGGATCTAGAG-3' (SEQ ID NO:5)
<i>NeuC</i> antisense	5'-CTAGTCATAACTGGTGGTACATTCGGGATGTC-3' (SEQ ID NO:6)
mouse CMP-Sia synthase sense	5'-ATGGACGCGCTGGAGAAGGGGGCCCGTCACGTC-3' (SEQ ID NO:7)
mouse CMP-Sia synthase antisense	5'-CTATTTTTGGCATGAGTTATTAAC TTTTCTATCAG-3' (SEQ ID NO:8)

Primer name	Primer sequence
porcine GlcNAc epimerase sense	5'-ATGGAGAAGGAGCGCGAAACTCTGCAGG-3' (SEQ ID NO:9)
porcine GlcNAc epimerase antisense	5'-CTAGGCGAGGCGGCTCAGCAGGGCGCTC-3' (SEQ ID NO:10)
<i>E. coli</i> Sialate aldolase sense	5'-ATGGCAACGAATTTACGTGGCGTAATGGCTG-3' (SEQ ID NO:11)
<i>E. coli</i> Sialate aldolase antisense	5'-TCACCCGCGCTCTTGCATCAACTGCTGGGC-3' (SEQ ID NO:12)
mouse bifunctional UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase sense	5'-ATGGAGAAGAACGGGAACAACCGAAAGCTCCG-3' (SEQ ID NO:25)
mouse bifunctional UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase antisense	5'-CTAGTGGATCCTGCGCGTTGTGTAGTCCAG-3' (SEQ ID NO:26)
mouse Sia9P syn sense	5'-ATGCCGCTGGAACCTGGAGCTGTGTCCCGGGC-3' (SEQ ID NO:27)
mouse Sia9P syn antisense	5'-TTAAGCCTTGATTTTCTTGCTGTGACTTCCAC -3' (SEQ ID NO:28)
human CMP-Sia synthase sense	5'-ATGGACTCGGTGGAGAAGGGGGCCGCCACC-3' (SEQ ID NO:29)
human CMP-Sia synthase antisense	5'-CTATTTTGGCATGAATTATTAACTTTTTCC-3' (SEQ ID NO:30)

EXAMPLE 2

Expression Of Bacterial *Neu* Genes In *P. pastoris*

[0173] The 1176 bp PCR amplified fragment of the *NeuC* gene was ligated into the *NotI-AscI* site in the yeast integration vector **pJN348** (a modified pUC19 vector comprising a GAPDH promoter, a *NotI AscI PacI* restriction site cassette, CycII transcriptional terminator, URA3 as a positive selection marker) producing **pSH256**. Similarly, the PCR amplified fragment (1041 bp) of the *NeuB* gene was ligated into the *NotI-PacI* site in the yeast integration vector **pJN335** under a GAPDH promoter using ADE as a positive selection marker producing **pSH255**.

The 1260 bp PCR amplified fragment of the *NeuA* gene was ligated into the *NotI-PacI* site in the yeast integration vector **pJN346** under a GAPDH promoter with ARG as a positive selection marker to produce **pSH254**. After transforming *P. pastoris* with each vector by electroporation, the cells were plated onto the
5 corresponding drop-out agar plates to facilitate positive selection of the newly introduced vector(s). To confirm the introduction of each gene, several hundred clones were repatched onto the respective dropout plates and grown for two days at 26°C. Once sufficient material had grown, each clone was screened by colony
10 PCR using primers specific for the introduced gene. Conditions for colony PCR using the polymerase ExTaq from Takara, were as follows: 97°C for 3 min, 1 cycle; 97°C for 20 sec, 50°C for 30 sec, 72°C for 2 min/kb, 30 cycles; 72°C for 10 min, 1 cycle. Subsequently, several positive clones from colony PCR were grown in a baffled flask containing 200 ml of growth media. The base composition of growth media containing 2.68 g/l yeast nitrogen base, 200 mg/l biotin and 2 g/l
15 dextrose was supplemented with amino acids depending on the strain used. The cells were grown in this media in the presence or absence of 20 mM ManNAc. Following growth in the baffle flask at 30°C for 4-6 days, the cells were pelleted and analyzed for intermediates of the sialic acid pathway, as described in **Example 10**.

20

EXAMPLE 3**Expression of GlcNAc Epimerase Gene In *P. pastoris***

[0174] The PCR amplified fragment of the porcine GlcNAc epimerase gene was ligated into the *NotI-PacI* site in the yeast integration vector **pJN348** under the control of the GAPDH promoter, using URA3 as a positive selection marker. The
25 *P. pastoris* strain producing endogenous GlcNAc was transformed with the vector carrying the GlcNAc epimerase gene fragment and screened for transformants.

EXAMPLE 4**Expression of Sialate Aldolase Gene In *P. pastoris***

[0175] The PCR amplified fragment of the *E. coli* sialate aldolase gene was
30 ligated into the *NotI-PacI* site in the yeast integration vector **pJN335** under the

control of the GAPDH promoter with ADE as a positive selection marker producing pSH275. The *P. pastoris* strain producing ManNAc was transformed with the vector carrying the sialate aldolase gene fragment and screened for transformants.

5

EXAMPLE 5**Expression Of The Gene Encoding UDP-N-acetylglucosamine-2-Epimerase/N-acetylmannosamine Kinase in *P. pastoris***

[0176] The PCR amplified fragment of the gene encoding the mouse bifunctional UDP-N-acetylglucosamine-2-Epimerase/N-acetylmannosamine Kinase enzyme
10 was ligated into the *NotI-PacI* site in the yeast integration vector pJN348 under the control of the GAPDH promoter with URA as a positive selection marker producing pSH284. The *P. pastoris* strain producing ManNAc was transformed with the vector carrying the gene fragment and screened for transformants.

EXAMPLE 6

15

Expression Of The Gene Encoding N-acetylneuraminate-9-Phosphate Synthase In *P. pastoris*

[0177] The PCR amplified fragment of the mouse N-acetylneuraminate-9-phosphate synthase gene was ligated into the *NotI-PacI* site in the yeast integration vector pJN335 under the control of the GAPDH promoter with ADE as a positive
20 selection marker producing pSH285. The *P. pastoris* strain producing ManNAc-6-P was transformed with the vector carrying the above gene fragment and screened for transformants.

EXAMPLE 7

25

Identification, Cloning And Expression Of The Gene Encoding N-acetylneuraminate-9-Phosphatase

[0178] N-acetylneuraminate-9-phosphatase activity has been detected in the cytosolic fraction of rat liver cells (Van Rinsum, J., Van Dijk, W. 1984). We have repeated this method and isolated a cell extract fraction containing phosphatase activity only against NeuAc-9-P. SDS-PAGE electrophoresis of this fraction
30 identifies a single protein band. Subsequently, this sample was electroblotted onto

a PDVF membrane, and the N-terminal amino acid sequence was identified by Edman degradation. The sequence identified allows the generation of degenerate oligonucleotides for the 5'-terminus of the ORF of the isolated protein. Using these degenerate primers in conjunction with the AP1 primer supplied in a rat liver
5 Marathon-ready cDNA library (Clontech), a full length ORF was isolated according to the manufacturer's instructions. The complete ORF was subsequently ligated into the yeast integration vector **pJN347** (WO 02/00879) under the control of the GAPDH promoter with a HIS gene as a positive selection marker. The *P. pastoris* strain producing NeuAc-9-P was transformed with the vector carrying
10 the desired gene fragment and screened for transformants as described in **Example 2.**

EXAMPLE 8

Cloning And Expression Of A CMP-Sialic Acid Synthase Gene In *P.pastoris*

[0179] The PCR amplified fragment of the mouse CMP-Sia synthase gene was
15 ligated into the NotI-PacI site in the yeast integration vector **pJN346** under the control of the GAPDH promoter with the ARG gene as a positive selection marker. A *P. pastoris* strain producing sialic acid was transformed with the vector carrying the above gene fragment and screened for transformants as described **Example 2.** Likewise, the human CMP-Sia synthase gene (Genbank: AF397212) was amplified
20 and ligated into the *NotI-PacI* site of the yeast expression vector **pJN346** producing the vector **pSH257**. A *P. pastoris* strain capable of producing sialic acid was transformed with **pSH257** by electroporation, producing a strain capable of generating CMP-Sia.

EXAMPLE 9

Expression Of The Hybrid CMP-Sia Pathway In *P. pastoris*

[0180] The *P. pastoris* strain **JC308** (Cereghino, 2001 Gene 263, 159-164) was
25 super-transformed with 20 mg of each of the vectors containing *NeuC* (**pSH256**), *NeuB* (**pSH255**) and hCMP-Sia synthase (**pSH257**) by electroporation. The resultant cells were plated on minimal media supplemented with histidine
30 (containing 1.34 g/l yeast nitrogen base, 200 mg/l biotin, 2 g/l dextrose, 20 g/l agar

and 20 mg/l L-histidine). Following incubation at 30°C for 4 days, several hundred clones were isolated by repatching onto minimal media plates supplemented with histidine (see above for composition). The repatched clones were grown for 2 days prior to performing colony PCR (as described in **Example 2**) on the clones. Primers specific for *NeuC*, *NeuB* and hCMP-Sia synthase were used to confirm the presence of each ORF in the transformed clones. Twelve clones positive for all three ORFs (designated **YSH99a-1**) were grown in a baffled flask containing 200 ml of growth media (containing 2.68 g/l yeast nitrogen base, 200 mg/l biotin, 20 mg/l L-histidine and 2 g/l dextrose). The effect of supplementing the growth media with ManNAc was investigated by growing the cells in the presence or absence of 20 mM ManNAc. Following growth in the baffle flask at 30°C for 4-6 days the cells are pelleted and analyzed for the presence of sialic acid pathway intermediates as described in **Example 10**.

[0181] Comparing the cell extracts using the assay outlined in **Example 10**, the cell extracts from *P. pastoris* **YSH99a** without exogenous CMP-Sia, showed transfer of Sia onto acceptor substrates indicating the presence of CMP-Sia (**Figure 12**). Both mono- and di-sialylated biantennary N-glycans eluted at 20 min and 23 min, their respective corresponding time. Additionally, the sialidase treatment (**Example 11**) showed the removal of sialic acid (**Figure 13**). Thus, a yeast strain engineered with a hybrid CMP-Sia biosynthetic pathway as described, containing the *NeuC*, *NeuB* and hCMP-Sia synthase, is capable of generating an endogenous pool of CMP-sialic acid.

EXAMPLE 10

25 Assay For The Presence of Cytidine-5'-Monophospho-N-Acetylneuraminic Acid In Genetically Altered *P.pastoris*

[0182] Yeast cells were washed three times with cold PBS buffer, and suspended in 100 mM ammonium bicarbonate pH 8.5 and kept on ice. The cells were lysed using a French pressure cell followed by sonication. Soluble cell contents were separated from cell debris by ultracentrifugation. Ice cold ethanol was added to the supernatant to a final concentration of 60% and kept on ice for 15 minutes prior to removal of insoluble proteins by ultracentrifugation. The supernatant was frozen and concentrated by lyophilization. The dried sample was resuspended in water

(ensuring pH is 8.0) and then filtered through a pre-rinsed 10,000 MWCO Centricon cartridge. The filtrate was separated on a Mono Q ion-exchange column and the elution fractions that co-elute with authentic CMP-sialic acid are pooled and lyophilized.

- 5 [0183] The dried filtrate was dissolved in 100 μ L of 100 mM ammonium acetate pH6.5, 11 μ L (5 mU) of α -2,6 sialyltransferase and 3.3 μ L (12 mU) of α -2,3 sialyltransferase were added, and 10 μ L of the mixture was removed for a negative control. Subsequently, 7 μ L (1.4 μ g) of 2-aminobenzamide-labeled asialo-
- 10 biantennary N-glycan (NA2, Glyco Inc., San Rafael, CA) was added to the remaining mixture, followed by the removal of 10 μ L for a positive control. The sample and control reactions were then incubated at 37°C for 16 hr. 10 μ L of each sample were then separated on a GlycoSep-C anion exchange column according to manufacturer's instructions. A separate control consisting of approximately 0.05 μ g each of monosialylated and disialylated biantennary glycans was separated on
- 15 the column to establish relative retention times. The results are shown in Figures 10-14.

EXAMPLE 11

Sialidase Treatment

- 20 [0184] The incubation of bi-antennary galactosylated N-glycans with an extract from the *P. pastoris* YSH99a strain in the presence of sialyltransferases produced sialylated N-glycans, which were subsequently desialylated as follows: a sialylated sample was passed through a Microcon cartridge, with 10,000 molecular weight cut-off, to remove the transferases. The cartridge was washed twice with 100 μ l of water, which was pooled with the original eluate. Analysis of the eluate
- 25 by HPLC (Figure 13) produced a spectrum similar to the HPLC spectrum prior to the Microcon treatment. The remaining sample was lyophilized to dryness and resuspended in 25 μ l of 1xNEB G1 buffer. After addition of 100 U of sialidase (New England Biolabs #P0720L, Beverley, MA), the resuspended sample was incubated overnight at 37°C prior to HPLC analysis, as described previously.

- 30 [0185] Any reference to publications cited in this specification is not an admission that the disclosures constitute common general knowledge in Australia.

REFERENCES:

- Alviano, C.S., Travassos, L.R., et al. (1999) Sialic acids in fungi: A minireview. *Glycoconjugate Journal*, 16, 545-554.
- Annunziato, P.W., Wright, L.F., et al. (1995) Nucleotide sequence and genetic
5 analysis of the neuD and neuB genes in region 2 of the polysialic acid gene cluster of *Escherichia coli* K1. *J. Bacteriol.*, 177, 312-319.
- Ballou, C.E. (1990) Isolation, characterization, and properties of *Saccharomyces cerevisiae* mnn mutants with nonconditional protein glycosylation defects. *Methods Enzymology*, 185, 440-470.
- 10 Belli, G., Gari, E. et al. (1998) An activator/ repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. *Nucleic Acids Res.*, 26, 942-947.
- Briles, E.B., Li, E., et al. (1977) Isolation of wheat germ agglutinin-resistant clones of Chinese hamster ovary cells deficient in membrane sialic acid and galactose. *J.*
15 *Biol. Chem.*, 252, 1107-1116.
- Chiba, Y., Suzuki, M., et al. (1998) Production of human compatible high mannose-type (Man(5)GlcNAc(2)) sugar chains in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 273, 26298-26304.
- Choi, B.K., Bobrowicz, P. et al. (2003) Use of combinatorial genetic libraries to
20 humanize N-linked glycosylation in the yeast *Pichia pastoris*. *Proc. Nat'l Acad. Sci. USA*. Apr 29;100(9):5022-7.
- Cregg, J.M. et al. (2000). Recombinant protein expression in *Pichia pastoris*. *Mol. Technol.*, 16, 23-52.
- Fritsch, M., Geilen, C.C., et al. (1996) Determination of cytidine 5'-monophospho-
25 N-acetylneuraminic acid pool size in cell culture scale using high-performance anion-exchange chromatography with pulsed amperometric detection. *J. Chromatogr. A.*, 727, 223-230.
- Fukuda, M.N., Sasaki, H., et al. (1989) Survival of recombinant erythropoietin in the circulation: the role of carbohydrates. *Blood*, 73, 84-89.
- 30 Graham, T. and Emr, S. (1991) Compartmental organization of Golgi-specific protein modification and vacuolar protein sorting events defined in a yeast sec18 (NSF) mutant. *J. Cell. Biol.*, 114, 207-218.

- Hamilton, S.R., Bobrowicz, P., et al. (2003) Production of Complex Human Glycoproteins in Yeast. *Science*, 301, 1244-1246.
- Harms, E., Kreisel, W., et al. (1973) Biosynthesis of N-acetylneuraminic acid in Morris hepatomas. *Eur. J. Biochem.*, 32, 254-262.
- 5 Hinderlich, S., Stasche, R., et al. (1997) A bifunctional enzyme catalyzes the first two steps in N-acetylneuraminic acid biosynthesis of rat liver. Purification and characterization of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase. *J. Biol. Chem.*, 272, 24313-24318.
- Inoue, H., Fukui, K., et al. (1990) Molecular cloning and sequence analysis of a
10 cDNA encoding a porcine kidney renin-binding protein. *J. Biol. Chem.*, 265, 6556-6561.
- Kelm, S. and Schauer, R. (1997) Sialic acids in molecular and cellular interactions. *Int. Rev. Cytol.*, 175, 137-240.
- Kepler, O.T., Hinderlich, S. et al. (1999) UDP-GlcNAc 2-epimerase: A regulator
15 of cell surface sialylation. *Science*, 284, 1372-1376.
- Lawrence, S.M., Huddleston, K.A., et al. (2000) Cloning and expression of the human N-acetylneuraminic acid phosphate synthase gene with 2-keto-3-deoxy-D-glycero- D-galacto-nononic acid biosynthetic ability. *J. Biol. Chem.*, 275, 17869-17877.
- 20 Lin Cereghino, G.P., Lin Cereghino, J., et al. (2001) New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of *Pichia pastoris*. *Gene*, 263, 159-169.
- MacDougall, I.C., Gray, S.J., et al. (1999). Pharmacokinetics of Novel Erythropoiesis Stimulating Protein Compared with Epoetin Alfa in Dialysis
25 Patients. *J. Am. Soc. Nephrol.* 10, 2392-2395.
- Maru, I., Ohta, Y., Murata, et al. (1996) Molecular cloning and identification of N-acetyl-D-glucosamine 2-epimerase from porcine kidney as a renin-binding protein. *J. Biol. Chem.*, 271, 16294-16299.
- Munster, A.K., Eckhardt, M., et al. (1998) Mammalian cytidine 5'-monophosphate
30 N-acetylneuraminic acid synthetase: a nuclear protein with evolutionarily conserved structural motifs. *Proc. Nat'l Acad. Sci. USA*, 95, 9140-9145.

- Nakanishi-Shindo, Y., Nakayama, K., et al. (1993) Structure of the N-Linked Oligosaccharides That Show the Complete Loss of Alpha-1,6-Polymannose Outer Chain From Och1, Och1 Mnn1, and Och1 Mnn1 Alg3 Mutants of *Saccharomyces-Cerevisiae*. *J. Biol. Chem.*, 268, 26338-26345.
- 5 Ohta, Y., Watanabe, K. et al. (1985) Complete nucleotide sequence of the *E. coli* N-acetylneuraminase lyase. *Nucleic Acids Res.* 13, 8843-8852.
- Parodi, A.J. (1993) N-glycosylation in trypanosomatid protozoa. *Glycobiology*, 3, 193-199.
- Ringenberg, M., Lichtensteiger, C., et al. (2001) Redirection of sialic acid
10 metabolism in genetically engineered *Escherichia coli*. *Glycobiology*, 11, 533-539.
- Rump, J.A., Phillips, J., et al. (1986) Biosynthesis of gangliosides in primary cultures of rat hepatocytes. Determination of the net synthesis of individual gangliosides by incorporation of labeled N-acetylmannosamine. *Biol. Chem. Hoppe Seyler*, 367, 425-432.
- 15 Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning: A laboratory manual*. 3rd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.
- Schauer, R. (2000). Achievements and challenges of sialic acid research. *Glycoconj. J.* 17, 485-99.
- 20 Spivak, J.L. and Hogans, B.B. (1989) The *in vivo* metabolism of recombinant human erythropoietin in the rat. *Blood*, 73, 90-99.
- Stasche, R., Hinderlich, S., et al. (1997) A bifunctional enzyme catalyzes the first two steps in N-acetylneuraminic acid biosynthesis of rat liver. Molecular cloning and functional expression of UDP-N-acetyl-glucosamine 2-epimerase/N-
25 acetylmannosamine kinase. *J. Biol. Chem.*, 272, 24319-24324.
- Tschopp, J.F., Brust, P.F., et al. (1987) Expression of the *lacZ* gene from two methanol-regulated promoters in *Pichia pastoris*. *Nucleic Acids Res.* 15, 3859-3876.
- Van Rinsum, J., Van Dijk, W., et al. (1984) Subcellular localization and tissue
30 distribution of sialic acid forming enzymes. *Biochem. J.*, 223, 323-328.
- Vimr, E., Steenbergen, S., et al. (1995) Biosynthesis of the polysialic acid capsule in *Escherichia coli* K1. *J. Ind. Microbiol.*, 15, 352-360.

- Vimr, E.R. and Troy, F.A. (1985) Regulation of sialic acid metabolism in *Escherichia coli*: Role of N-acetylneuraminic acid pyruvate-lyase. *J. Bacteriol.* 164, 854-860.
- Warren, L. (1994) Bound Carbohydrates in Nature. Cambridge University Press, Cambridge, U.K.
- 5 Yocum, R.R., Hanley, S. et al. (1984) Use of lacZ fusions to delimit regulatory elements of the inducible divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, 4, 1985-1998.
- Yoko-o, T., Tsukahara, K., et al. (2001) *Schizosaccharomyces pombe* och1(+) encodes alpha-1,6-mannosyltransferase that is involved in outer chain elongation of N-linked oligosaccharides. *FEBS Lett*, 489, 75-80.
- 10 Zapata, G., Crowley, J.M., et al. (1992) Sequence and expression of the *Escherichia coli* K1 neuC gene product. *J. Bacteriol.*, 174, 315-319.
- Zapata, G., Vann, W.F., et al. (1989) Sequence of the cloned *Escherichia coli* K1
15 CMP-N-acetylneuraminic acid synthetase gene. *J. Biol. Chem.*, 264, 14769-14774.

2005224672 19 May 2011

- 58 -

The claims defining the invention are as follows:

1. A recombinant fungal host cell comprising a CMP-Sia biosynthetic pathway consisting of *E. coli NeuC* enzyme activity, *E. coli NeuB* enzyme activity, and a mammalian CMP-sialate synthase activity or a fusion protein comprising an *E. coli* CMP-sialate synthase (*NeuA*) catalytic domain and a cellular targeting signal peptide not normally associated with the catalytic domain selected to target the catalytic domain to the nucleus of the host cell and a sialyltransferase wherein the host cell is capable of producing sialylated glycolipids, O-glycans or N-glycans.
2. The recombinant host cell of claim 1, wherein the host is selected from the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia minuta*, *Ogataea minuta*, *Pichia lindneri*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia sp.*, *Saccharomyces cerevisiae*, *Saccharomyces sp.*, *Hansenula polymorpha*, *Kluyveromyces sp.*, *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Aspergillus sp*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum* and *Neurospora crassa*.
3. The host cell of claim 1 or claim 2, wherein said host cell expresses a heterologous therapeutic protein selected from the group consisting of: erythropoietin, cytokines, interferon- α , interferon- β , interferon- γ , interferon- ω , tumor necrosis factor- α (TNF- α), granulocyte-CSF, granulocyte/macrophage colony stimulating factor (GM-CSF), interleukins, IL-1ra, coagulation factors, factor VIII, factor IX, human protein C, antithrombin III and thrombopoietin, IgA antibodies or fragments thereof, IgG antibodies or fragments thereof, IgD antibodies or fragments thereof, IgE antibodies or fragments thereof, IgM antibodies and fragments thereof, soluble IgE receptor α -chain, urokinase, chymase, urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, follicle stimulating hormone (FSH), annexin V fusion protein, angiostatin, vascular endothelial growth factor-2, myeloid progenitor inhibitory factor-1, osteoprotegerin, α -1 antitrypsin, DNase II, α -feto proteins and glucocerebrosidase.
4. A method for producing sialylated glycolipids, O-glycans or N-glycans in a fungal host cell comprising expressing a CMP-Sia biosynthetic pathway consisting of *E. coli NeuC* enzyme

2005224672 19 May 2011

- 59 -

activity, *E. coli NeuB* enzyme activity, and a mammalian CMP-sialate synthase activity or a fusion protein comprising an *E. coli* CMP-sialate synthase (*NeuA*) catalytic domain and a cellular targeting signal peptide not normally associated with the catalytic domain selected to target the catalytic domain to the nucleus of the host cell and a sialyltransferase.

5. The method of claim 4, wherein the host is selected from the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia minuta*, *Ogataea minuta*, *Pichia lindneri*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia sp.*, *Saccharomyces cerevisiae*, *Saccharomyces sp.*, *Hansenula polymorpha*, *Kluyveromyces sp.*, *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Aspergillus sp.*, *Trichoderma reesei*, *Chryso sporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum* and *Neurospora crassa*.
6. The method of claim 4 or claim 5, wherein the CMP-sialate synthase enzyme activity localizes in the nucleus of the host cell.
7. The method of claim 4 or claim 5, wherein the enzyme activity is expressed under the control of a constitutive promoter or an inducible promoter.
8. The method of claim 4 or claim 5, wherein the expressed enzyme activity is from a partial ORF encoding that enzymatic activity.
9. The method of claim 4 or claim 5, wherein the expressed enzyme is a fusion to another protein or peptide.
10. The method of claim 4 or claim 5, wherein said host cell expresses a heterologous therapeutic protein selected from the group consisting of: erythropoietin, cytokines, interferon- α , interferon- β , interferon- γ , interferon- ω , tumor necrosis factor- α (TNF- α), granulocyte-CSF, granulocyte/macrophage colony stimulating factor (GM-CSF), interleukins, IL-1ra, coagulation factors, factor VIII, factor IX, human protein C, antithrombin III and thrombopoietin, IgA antibodies or fragments thereof, IgG antibodies or fragments thereof, IgD antibodies or fragments thereof, IgE antibodies or fragments thereof, IgM antibodies and fragments thereof, soluble IgE receptor α -chain, urokinase, chymase, urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, follicle stimulating hormone (FSH), annexin V fusion protein, angiostatin, vascular endothelial growth factor-2, myeloid progenitor

- 60 -

inhibitory factor-1, osteoprotegerin, α -1 antitrypsin, DNase II, α -feto proteins and glucocerebrosidase.

11. A method for producing a recombinant glycoprotein comprising expressing said glycoprotein in the host cell of any one of claims 1 to 3.

12. The method of claim 11, wherein said host cell expresses a heterologous therapeutic protein selected from the group consisting of: erythropoietin, cytokines, interferon- α , interferon- β , interferon- γ , interferon- ω , tumor necrosis factor- α (TNF- α), granulocyte-CSF, granulocyte/macrophage colony stimulating factor (GM-CSF), interleukins, IL-1ra, coagulation factors, factor VIII, factor IX, human protein C, antithrombin III and thrombopoietin, IgA antibodies or fragments thereof, IgG antibodies or fragments thereof, IgD antibodies or fragments thereof, IgE antibodies or fragments thereof, IgM antibodies and fragments thereof, soluble IgE receptor α -chain, urokinase, chymase, urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, follicle stimulating hormone (FSH), annexin V fusion protein, angiostatin, vascular endothelial growth factor-2, myeloid progenitor inhibitory factor-1, osteoprotegerin, α -1 antitrypsin, DNase II, α -feto proteins and glucocerebrosidase.

13. A recombinant fungal host cell comprising a CMP-Sia biosynthetic pathway consisting of *E. coli NeuC* enzyme activity, *E. coli NeuB* enzyme activity, and a mammalian CMP-sialate synthase activity or a fusion protein comprising an *E. coli* CMP-sialate synthase (*NeuA*) catalytic domain and a cellular targeting signal peptide not normally associated with the catalytic domain selected to target the catalytic domain to the nucleus of the host cell.

14. The recombinant fungal host cell of claim 13, wherein the host is selected from the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia minuta*, *Ogataea minuta*, *Pichia lindneri*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia sp.*, *Saccharomyces cerevisiae*, *Saccharomyces sp.*, *Hansenula polymorpha*, *Kluyveromyces sp.*, *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Aspergillus sp.*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum* and *Neurospora crassa*.

15. The recombinant fungal host cell of claim 14 wherein the host cell is *Pichia pastoris*.

2005224672 19 May 2011

2005224672 19 May 2011

- 61 -

16. A method for producing a cellular pool of CMP-sialic acid in a fungal host cell wherein the CMP-sialic acid in the pool consists of *N*-acetyl-5-neuraminic acid in a fungal host cell comprising:

(a) providing a yeast host cell comprising a CMP-Sia biosynthetic pathway consisting of *E. coli NeuC* enzyme activity, *E. coli NeuB* enzyme activity, and a mammalian CMP-sialate synthase activity or a fusion protein comprising an *E. coli* CMP-sialate synthase (*NeuA*) catalytic domain and a cellular targeting signal peptide not normally associated with the catalytic domain selected to target the catalytic domain to the nucleus of the host cell; and

(b) growing the yeast host cell under conditions effective for producing the cellular pool of CMP-sialic acid wherein the CMP-sialic acid in the pool consists of *N*-acetyl-5-neuraminic acid.

17. The method of claim 16, wherein the host is selected from the group consisting of the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia minuta*, *Ogataea minuta*, *Pichia lindneri*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia sp.*, *Saccharomyces cerevisiae*, *Saccharomyces sp.*, *Hansenula polymorpha*, *Kluyveromyces sp.*, *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Aspergillus sp.*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum* and *Neurospora crassa*.

18. The method of claim 17 wherein the host cell is *Pichia pastoris*.

19. A recombinant fungal host cell as defined in claim 1 or claim 13 and substantially as hereinbefore described with reference to one or more of the accompanying examples.

20. A method for producing a recombinant glycoprotein comprising expressing said glycoprotein in the host cell of claim 19.

Dated: 19 May 2011

Figure 1

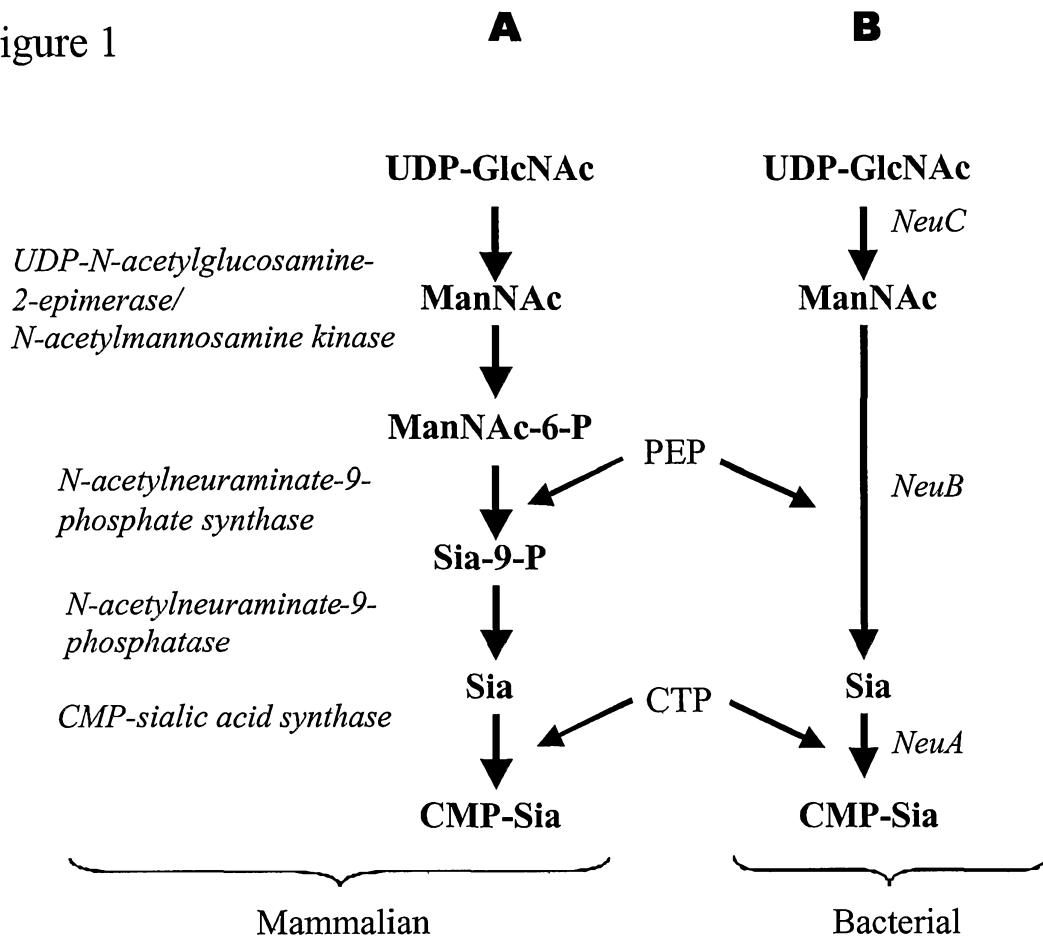


Figure 2

NeuC

```

1  ATGAAAAAAAAATTATACGTAACGGATCTAGAGCTGAATATGGAATAGTTCGGAGACTTTTGACAATGCTAAGAGAACTCCAGAAATACAGCTTGAT
1▶  MKKKILLYVTGSRAEYGI VRRLLTMLRETP EIQLD
100 TTGGCAGTTACAGGAATGCATTGTGATAATGCGTATGGAATACAATACATATTATAGAACAAGATAATTTAATATTCAAGGTTGTGGATATAAAAT
34▶  LAVTGMHC DNAYGNTIHIIEQDNFNIIKVVDIN
199 ATCAATCAACTTCACATACTCACATTCCATTCAATGAGTGTGGCCTCAATTCGTTGGTGATTTTTTTCAAATAACACATATGATGCGGTTATG
67▶  INTTSHTHILHSM SVCLNSFGDFFSNNTYDAVM
298 GTTTTAGCGATAGATATGAAATATTTTCAGTCGCTATCGCAGCATCAATGCATAATATCCATTAATTCATATTCATGGTGGTGAAGA CATTAGCT
100▶  VLGDRYEIFSV AIAASMHNIPLIHIHGGEKTLA
397 AATTATGATGAGTTTATAGGCATTCAATTACTAAAAAGAGTAACTCCATCTTACTTACAGAAGAGTATAAAAAACGAGTAATCAACTAGGTGAA
133▶  NYDEFIRHSITKMSK LHLTSTEEYK KRVIQLGE
496 AAGCCTGGTAGTGTGTTAATATTGGTCTCTTGGTGCAGAAAATGCTCTTTCATTGCATTTACCAATAAGCAGGAGTGGAACTAAAATATGTTCA
166▶  KPGSVFNIGSLGAENALSLHLPNKQEL ELKYGS
595 CTGTTAAACGGTACTTTGTTGTAGTATCCATCCTGAAACACTTTCCACGCGTCGGTTAATGATCAAAATAGATGAGTTATTGTCAGCGA TTTCTTT
199▶  LLKRYFVVVFHPE T LSTQSVNDQIDELLSA ISF
694 TTTAAAAATACTCAGACTTTATTTTATGGCAGTAACCGTGACACTGGTCTGATATAATTCAGAGAAAAGTAAAATATTTTGCAAGAGTATAAG
232▶  FKNTHDFIFIGS NADTGS DIIQRKV KYFCKE YK
793 TTCAGATATTGATTTCTATTTCGTT CAGAAGATTATTGGCAATGATTAATACTCTTGGGGCTAATGGGAACCTCCTCTGGTTTAA TTGAGGTT
265▶  FRYLISIRSE DYLAMIKYSCGLIGNSS SGLIEV
892 CCATCTTTAAAGTTGCAACAATTAACATTGGTGATAGGCAGAAAAGCCGCTTTCGTTGGAGCCAGTGAATAGATGTACCCGTTGAAAAAAATGCAATC
298▶  PSLKVA TINIGDRQKGRV RGA SVI DVPVEKNAI
991 GTCAGAGGGATAAATATATCTCAAGATGAAAAATTTATTAGTGTGTACAGTCATCTAGTAATCCTTATTTAAAGAAAATGCTTAAATTAATGCTGTT
331▶  VRGINISQDEK F ISVVQSSSNP YFKENAL I N A V
1090 AGAATTATAAGGATTTTATAAATCAAAAAATAAGATTACAAAAGATTTTATGACATCCCGGAATGTACCACCAGTTATGACTAG
364▶  RIIKDFIKSKNKDYKDFYD IPECTT SYD
    
```

Figure 3

NeuB

```

1  ATGAGTAATATATATATCGTTGCTGAAATGGTTGCAACCATAATGCTAGTGTGGATATTGCAAGAGAAATGATATTAAGCCAAAGAGGCCGGTGT
1▶ M S N I Y I V A E I G C N H N G S V D I A R E M I L K A K E A G V
100 AATGCAGTAAAATCCAAACATTTAAAGCTGATAAATTAATTCAGCTATTGCACCTAAGGCAGAGTATCAAATAAAAAACACAGGAGAATTAGAATCT
34▶ N A V K F Q T F K A D K L I S A I A P K A E Y Q I K N T G E L E S
199 CAGTTAGAAATGACAAAAAGCTTGAATGAAGTATGACGATTATCTCCATCTAATGGAATATGCAGTCAGTTAAATTTAGATGTTTTTCTACCCCT
67▶ Q L E M T K K L E M K Y D D Y L H L M E Y A V S L N L D V F S T P
298 TTTGACGAAGACTCTATTGATTTTTAGCATCTTTGAAACAAAAATATGGAAAATCCCTTCAGGTGAGTTATTGAATTTACCGTATCTTGAAAAATA
100▶ F D E D S I D F L A S L K Q K I W K I P S G E L L N L P Y L E K I
397 GCCAAGCTCCGATCCCTGATAAGAAAAATACATATCAACAGGAATGGCTACTATTGATGAGATAAACAGTCTGTTTCTATTTTATAAATAATAA
133▶ A K L P I P D K K I I I S T G M A T I D E I K Q S V S I F I N N K
496 GTTCCGGTGGTAATATTACAATATTACATTGCAACTGAATATCCAACGCCCTTGAGGATGTAACCTTAATGCTATTAATGATTTGAAAAACAC
166▶ V P V G N I T I L H C N T E Y P T P F E D V N L N A I N D L K K H
595 TTCCCTAAGAATAACATAGGCTTCTCTGATCATTCTAGCGGGTTTTATGCAGCTATTGCGGCGTGCCTTATGGAATAACTTTTATTGAAAAACATTC
199▶ F P K N N I G F S D H S S G F Y A A I A A V P Y G I T F I E K H F
694 ACTTTAGATAAATCTATGTCTGGCCAGATCATTGGCCTCAATAGAACCTGATGAACTGAAACATCTTTGATTTGGGGTCAGGTGTGTTGAAAAATCT
232▶ T L D K S M S G P D H L A S I E P D E L K H L C I G V R C V E K S
793 TTAGGTTCAAATAGTAAAGTGGTTACAGCTTCAGAAAGGAAGAATAAAATCGTAGCAAGAAAGTCTATTATAGCTAAAACAGAGATAAAAAAAGGTGAG
265▶ L G S N S K V V T A S E R K N K I V A R K S I I A K T E I K K G E
892 GTTTTTTCAGAAAAAATATAACAACAAAAAGACCTGGTATGGTATCAGTCCGATGGAGTGGTATAATTTATTGGTAAAATTGCAGAGCAAGACTTT
298▶ V F S E K N I T T K R P G N G I S P M E W Y N L L G K I A E Q D F
991 ATTCAGATGAATTAATAATTCATAGCGAATTCAAAAATCAGGGGAATAA
331▶ I P D E L I I H S E F K N Q G E
    
```

Figure 4

NeuA

```

1  ATGAGAACAAAAATTATTGCGATAATCCAGCCCGTAGTGGATCTAAAGGGTTGAGAAATAAAAAATGCTTTGATGCTGATAGATAAACCTCTTCTTGCT
1▶ M R T K I I A I I P A R S G S K G L R N K N A L M L I D K P L L A
100 TATACAATTGAAGCTGCCTTGCAGTCAGAAAATGTTGAGAAAAGTAATGTGACAACCTGACTCCGAACAGTATGGAGCAATAGCAGAGTCATATGGTGCT
34▶ Y T I E A A L Q S E M F E K V I V T T D S E Q Y G A I A E S Y G A
199 GATTTTTGCTGAGACCGGAAGAAGTAGCAACTGATAAAGCATCATCATTGAATTTATAAAACATGCGTTAAGTATATATACTGATTATGAGAGCTTT
67▶ D F L L R P E E L A T D K A S S F E F I K H A L S I Y T D Y E S F
298 GCTTTATTACAACCAACTCCACCTTTAGAGATTCGACCCATATTTAGGGCTGTAAGTTATATCAAACCTTAGAAAAATACCAATGTGTGTTTCT
100▶ A L L Q P T S P F R D S T H I I E A V K L Y Q T L E K Y Q C V V S
397 GTTACTAGAAGCAATAAGCCATCACAAATAATTAGACCATTAGATGATTACTCGACACTGTCTTTTTGACCTTGATTATAGTAAATATAATCGAAAC
133▶ V T R S N K P S Q I I R P L D D Y S T L S F F D L D Y S K Y N R N
496 TCAATAGTAGAATATCATCCGAATGGAGCTATATTTATAGCTAATAAGCAGCATTATCTTCATACAAGCATTTTTTGCTCGCTATTCCTAGCTTAT
166▶ S I V E Y H P N G A I F I A N K Q H Y L H T K H F F G R Y S L A Y
595 ATTATGGATAAGGAAAGCTCTTTAGATATAGATGATAGAATGGATTCGAACTTCCAATTACCATTGACAAAAAAAATAGACAAAAAATGACCTT
199▶ I M D K E S S L D I D D R M D F E L A I T I Q Q K K N R Q K I D L
694 TATCAAAACATACATAATAGAATCAATGAGAAAAGAAATGAATTTGATAGTGAAGTATTAACCTTAATGGACACTCGCTGTTGATTATGGGAC
232▶ Y Q N I H N R I N E K R N E F D S V S D I T L I G H S L F D Y W D
793 GTAAAAAATAAATGATATAGAAGTTAATAACTTAGGTATCGCTGGTATAAACTCGAAGGAGTACTATGAATATATTATTGAGAAAGAGCTGATTGTT
265▶ V K K I N D I E V N N L G I A G I N S K E Y Y E Y I I E K E L I V
892 AATTTCCGAGAGTTGTTTTTCATCTTTTTGGAACTAATGATATAGTTGTTAGTGATTGGAAAAAGAAGACACATTGTGGTATTGAAGAAAACATGC
298▶ N F G E F V F I F F G T N D I V V S D W K K E D T L W Y L K K T C
991 CAGTATATAAAGAAGAAAAATGCTGCATCAAAAATTTATTTATGTCGGTTCCTCCTGTTTTGGGCGTATTGATCGAGATAATAGAATAATTAATGAT
331▶ Q Y I K K K N A A S K I Y L L S V P P V F G R I D R D N R I I N D
1090 TTAATTCCTTATCTTCGAGAGAATGTAGATTTTGCAGGTTTATTAGCTTGGATCACGTTTTAAAAGACTCTTATGGCAATCTAAATAAATGTATACT
364▶ L N S Y L R E N V D F A K F I S L D H V L K D S Y G N L N K M Y T
1189 TATGATGGCTTACATTTAATAGTAATGGGTATACAGTATTAGAAAACGAAATAGCCGGAGATTGTTAAATGA
397▶ Y D G L H F N S N G Y T V L E N E I A E I V K
    
```

Figure 5

mCMP-Sia synthase

```

1  ATGGACGCGCTGGAGAAGGGGGCCGTCACGTCGGGGCCCGCCCGCGTGGACGGCCGTCGCCGGGGCCGGCCCCGAAGCTGCAGCGCAGCCGGGGCGG
1▶ M D A L E K G A V T S G P A P R G R P S R G R P P K L Q R S R G A
100 GGGCGCGGCTAGAGAAGCCGCCACCTGGCAGCGCTGGTGTGGCCCGCGGGCAGCAAAGGCATCCCACCTGAAGAACATCAAGCGCCTGGCGGGG
34▶ G R G L E K P P H L A A L V L A R G G S K G I P L K N I K R L A G
199 GTTCCGCTCATGGCTGGGTCCTGCGCGCCGCCCTGGATGCGGGGGTCTTCCAGAGTGTGGGTTCAACAGACCATGATGAAATGAGAATGTGGCC
67▶ V P L I G W V L R A A L D A G V F Q S V W V S T D H D E I E N V A
298 AAACAGTTTGGTGCACAGGTCATCGAAGAAGTTCTGAAACGTCCAAGACAGCTCTACCTCACTAGACGCCATTGTAGAATTCCTGAATTATCACAAT
100▶ K Q F G A Q V H R R S S E T S K D S S T S L D A I V E F L N Y H N
397 GAGGTTGACATTGTGGGGAATATCCAAGCCACATCCTCATGTTTACATCCCAGTCCAGAAAAGTGCAGAAATGATCCGAGAAGAAGGATATGAC
133▶ E V D I V G N I Q A T S P C L H P T D L Q K V A E M I R E E G Y D
496 TCTGCTTCTCCGTTGTGAGCGCCATCAGTTTCGATGGAGTCAAATTCAGAAAGSAGTTCGTGAAGTACTGAGCCTGGAATCCAGCGAAA
166▶ S V F S V V R R H Q F R W S E I Q K G V R E V T E P L N L N P A K
595 CGGCCTCGTGCACAAGACTGGGATGGAGAGTTATATGAGAACGGCTCATTTTATTTGCTAAAAGACATTTGATAGAGATGGGTTACTTACAGGGTGGG
199▶ R P R R Q D W D G E L Y E N G S F Y F A K R H L I E M G Y L Q G G
694 AAAATGGCATATTATGAAATGCGAGCTGAGCACAGTGGATATCGACGTGGACATCGATTGGCCGATCGCAGAGCAAAGAGTCTGAGATTGGCTAT
232▶ K M A Y Y E M R A E H S V D I D V D I D W P I A E Q R V L R F G Y
793 TTTGGAAAAGAGAAGCTGAAGGAGATAAAGCTTTGTTGTAATATTGATGGATGCTCACAATGCCACATTTATGATCAGGAGACCAAAAAGAA
265▶ F G K E K L K E I K L L V C N I D G C L T N G H I Y V S G D Q K E
892 ATAATATCTTATGATGTAAGAACGCCATTTGGCATAAGTTTATAAAGAAAAGCGGTATTGAGGTGAGGCTCATCTCAGAACGGGCTGCTCCAAGCAG
298▶ I I S Y D V K D A I G I S L L K K S G I E V R L I S E R A C S K Q
991 ACGCTCTGCGCCTAAAGCTGGACTGTA AACAGAAGTCAGTGTGTCGGATAAGCTGGCCACCGTGGATGAGTGGAGGAAGGAGATGGGCCTGTGCTGG
331▶ T L S A L K L D C K T E V S V S D K L A T V D E W R K E M G L C W
1090 AAAGAAGTGGCCTATCTCGGCAATGAAGTGTCTGATGAGAATGCCTCAAGAGAGTGGGCTGAGCGGTCTTCTGCGCAGCCTGCTCCGGGGCCCG
364▶ K E V A Y L G N E V S D E E C L K R V G L S A V P A D A C S G A Q
1189 AAGGCTGTGGGTACATCTGCAATGCAGCGGTGGCCGGGAGCCATCCGCGAGTTTGCAGAGCACATTTCTCTACTGATAGAAAAGTTAATAACTCA
397▶ K A V G Y I C K C S G G R G A I R E F A E H I F L L I E K V N N S
1288 TGCCAAAAATAG
430▶ C Q K
    
```

Figure 6

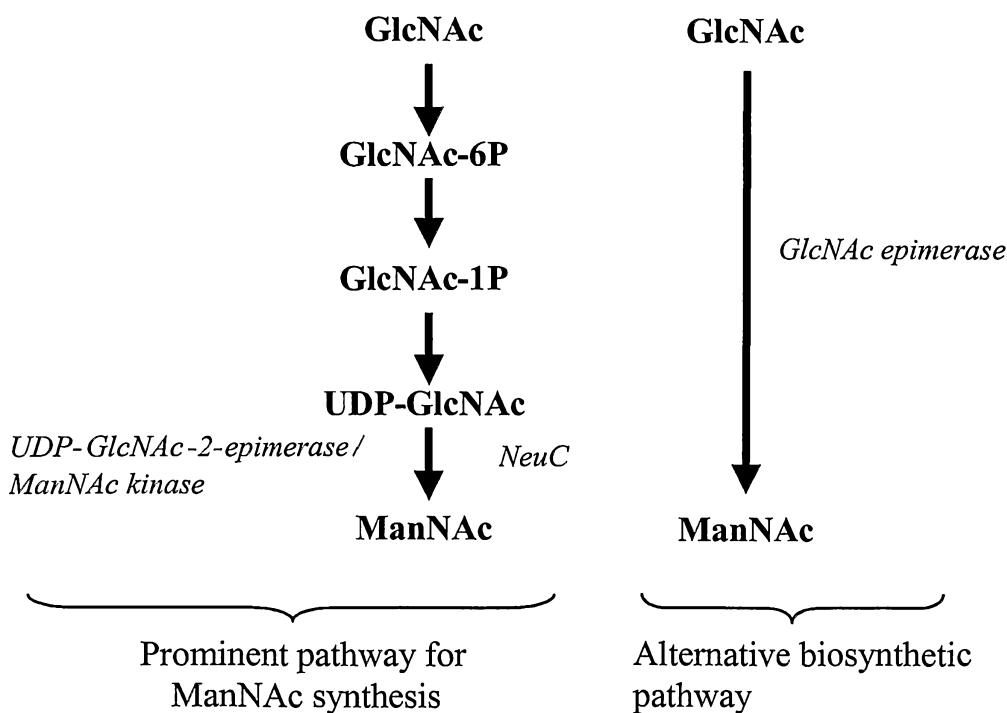


Figure 7

Pig GlcNAc epimerase

```

1  ATGGAGAAGGAGCGCGAAACTCTGCAGGCCTGGAAGGAGCGTGTGGCCAAAGAGCTGGACCGCGTGATGGCTTTCTGGCTGGAGCACTCCCACGATCGG
1▶ MEKERETLQAWKERVVGQELDRVMAFWLEHSHDR
100 GAGCACGGGGCTTCTTACGTGCCTGGGCCGCGACGGCGGGTGTATGACGACCTCAAGTACGTCTGGCTGCAGGGGAGGAGGTGGATGTACTGT
34▶ EHGGFFTCLGRDGRVYDDLKYVWLQGRQVWMYC
199 CGCCTGTACCGCAAGCTTGAGCGCTTCCACCGCCTGAGCTTCTGGATGCGGCTAAAGCAGGGGGCAATTTTGTGCGCCATGCCCGAGTGGCACCT
67▶ RLYRKLERFRPELLDAAKAGGEFLLRHARVAP
298 CCTGAAAAGAAGTGTGCCTTTGTGTGACGCGGGACGGCCGGCTCAAGTGCAGCGGAGCATCTTCAGTGAGTGCTTACACCATGGCCATGAAC
100▶ PEKKCAFVLTTRDGRPVKVQRSIFSECFYTMAMN
397 GAGCTGTGGAGGTGACGGCGGAGGCACGGTACCAGAGCGAAGCGGTGGACATGATGGATCAGATCGTGCACTGGGTGCGAGAGGACCCCTCTGGGCTG
133▶ ELWRVTAEARYQSEAVDMMMDQIVHWVREDP SGL
496 GGCCGGCCCCAGCTCCCCGGGGCCCTGGCCTCGGAGTCCATGGCAGTGCCCATGATGCTGCTGTGCTGGTGGAGCAGCTCGGGAGGAGGACGAGGAG
166▶ GRPQLPGA VAS ESMAVPMMLLCLVEQLGEED E
595 CTGGCAGCCGCTACGCGCAGCTGGGGCACTGGTGCCTCGGAGGATCCTGCAGCAGTCCAGAGGGATGGACAGGCTGTGCTGGAGAATGTGTCGGAA
199▶ LAGRYAQLGHWCA R R I L Q H V Q R D G Q A V L E N V S E
694 GATGGCAGGAAC TTTCTGGCTGCCTGGGGAGACACCAGAACCCAGGCCACGGCTGGAAGCTGGCTGGTTCCTGCTCCGCCACAGCAGCCGGAGCGGT
232▶ DGEELSGLGRHQNP GHALEAGWFLLRHS SR S G
793 GACGCCAACTTCGAGCCCACGTCATCGACACGTTCTGCTACTGCCTTCCGCTCCGGATGGGACGCTGATCAGGAGGCCCTTCTACTCCAGGAT
265▶ DAKLRAHV I D T F L L L P F R S G W D A D H G G L F Y F Q D
892 GCCGATGGCCTTGCCCCACCCAGCTGGAGTGGCCATGAAGCTCTGGTGGCCGACAGCGAAGCCATGATCGCCTTCTCATGGCTACAGTGAGAGC
298▶ ADGLCPPTQLEWAMK L W W P H S E A M I A F L M G Y S E S
991 GGGACCCCTGCCTTACTGCGTCTCTTCTACCAGGTGGCCGAGTACAGTTCGCCAGTTTCGTGATCCCCGAGTACGGGAATGGTTGGCTACCTGAAC
331▶ GDPALLRLRFYQVAEYTF R Q F R D P E Y G E W F G Y L N
1090 CGAGAGGGGAAGTTGCCCTCACTATCAAGGGGGTCCCTTTAAGGCTGCTTCCACGTGCCCGGTGCCTTGCCATGTGCGAAGAGATGCTGAGCGCC
364▶ REGKVAL T I K G G P F K G C F H V P R C L A M C E E M L S A
1189 CTGCTGAGCCGCTCGCCTAG
397▶ LLSRLA
    
```

Figure 8

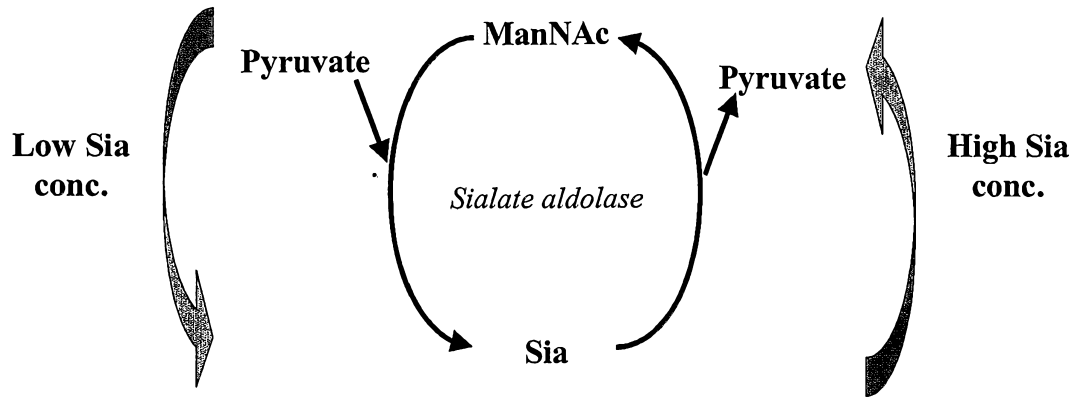


Figure 9

Sialate aldolase

```

1  ATGGCAACGAATTACGTGGCGTAATGGCTGCACTCCTGACTCCTTTGACCAACAACAAGCACTGGATAAAGCGAGTCTCGGTCGCCTGGTTCAGTTCA
1▶ M A T N L R G V M A A L L T P F D Q Q Q A L D K A S L R R L V Q F
101 ATATTAGCAGGGCATCGACGGTTATACGTGGGTGGTTCGACCGCGAGGCCTTTGTACAAGCCTTCCGAGCGTGAACAGTACTGAAATCGTCGC
34▶ N I Q Q G I D G L Y V G G S T G E A F V Q S L S E R E Q V L E I V A
201 CGAAGAGGGCAAAGGTAAGATAAACTCATCGCCCACGTCGGTTGCGTCACGACCGCCGAAAGCCAACAACCTTGCGGCATCGGCTAAACGTTATGGCTTC
67▶ E E G K G K I K L I A H V G C V T T A E S Q Q L A A S A K R Y G F
301 GATGCCGTCTCCGCCGTCACGCCGTTCTACTATCCTTTCAGCTTGAAGAACAACCTGCGATCACTATCGGGCAATTTATTGATTTCGGCGGATGGTTTGCCGA
101▶ D A V S A V T P F Y Y P F S F E E H C D H Y R A I I D S A D G L P
401 TGGTGGTGTACAACATTCCAGCCCTGAGTGGGGTAAACTGACCCCTGGATCAGATCAACACACTTGTTACATTGCCTGGCGTAGGTGCGCTGAAACAGAC
134▶ M V V Y N I P A L S G V K L T L D Q I N T L V T L P G V G A L K Q T
501 CTCTGGCGATCTCTATCAGATGGAGCAGATCCGTCGTGAACATCCTGATCTTGIGCTCTATAACGGTTACGACGAAATCTTCGCCTCTGGTCTGCTGGCG
167▶ S G D L Y Q M E Q I R R E H P D L V L Y N G Y D E I F A S G L L A
601 GCGCTGATGGTGGTATCGGCAGTACCTACAACATCATGGGCTGGCGCTATCAGGGGATCGTTAAGGCGCTGAAAGAAGGCGATATCCAGACCGCGCAGA
201▶ G A D G G I G S T Y N I M G W R Y Q G I V K A L K E G D I Q T A Q
701 AACTGCAAACTGAATGCAATAAAGTCATTGATTTACTGATCAAAACGGGCGTATCCGCGGCCTGAAAACCTGCTCCATTTATATGGATGTCGTTTCTGT
234▶ K L Q T E C N K V I D L L I K T G V F R G L K T V L H Y M D V V S V
801 GCCGCTGTCCCGAAACCGTTTGGACCGGTAGATGAAAAATATCAGCCAGAACTGAAGGCGCTGGCCCGCAGTTGATGCAAGAGCCCGGGTGA
267▶ P L C R K P F G P V D E K Y Q P E L K A L A Q Q L M Q E R G
    
```

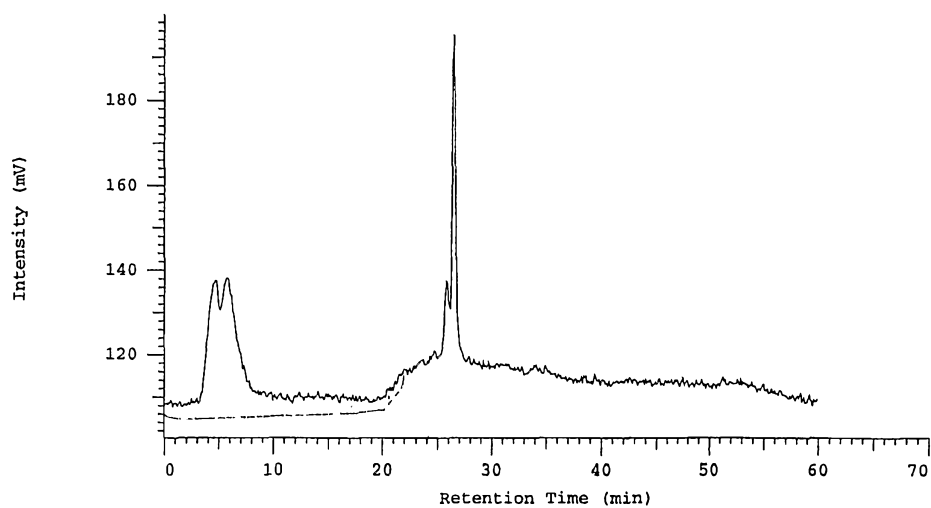


Fig. 10

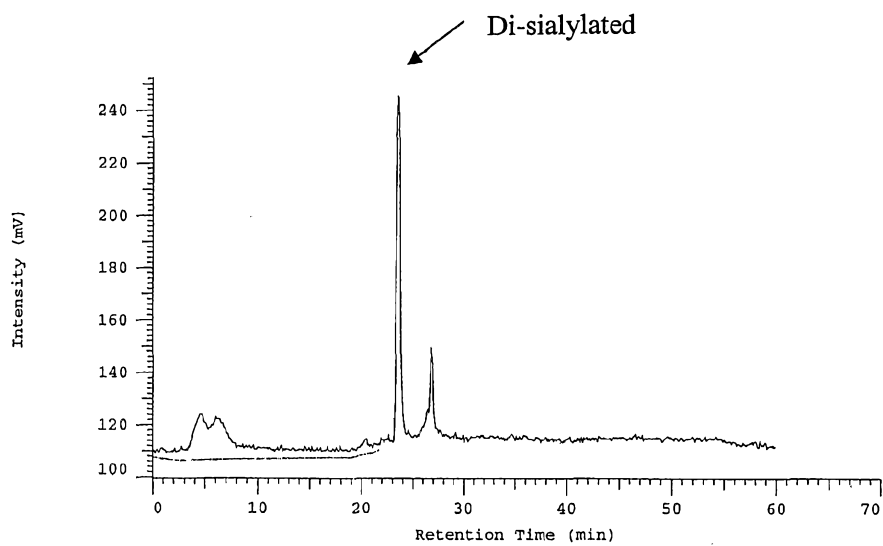


Fig. 11

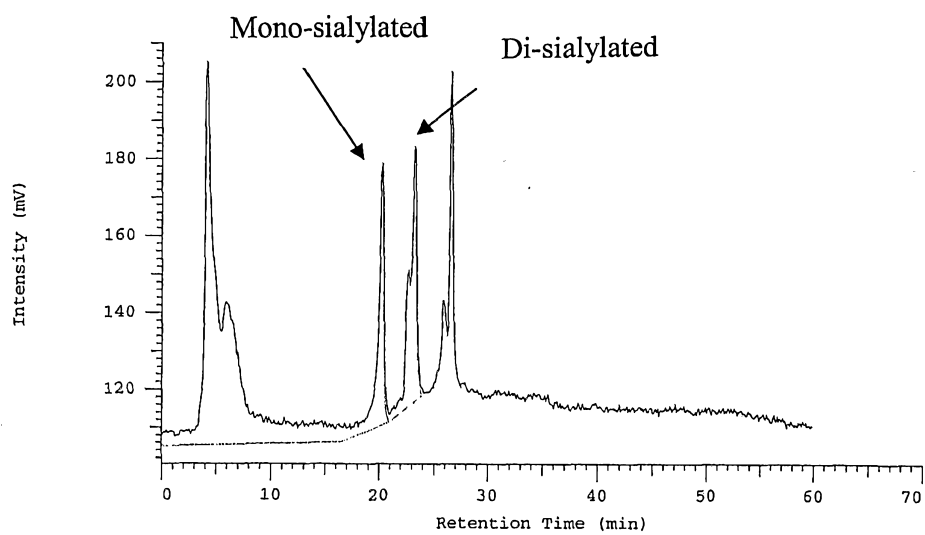


Fig. 12

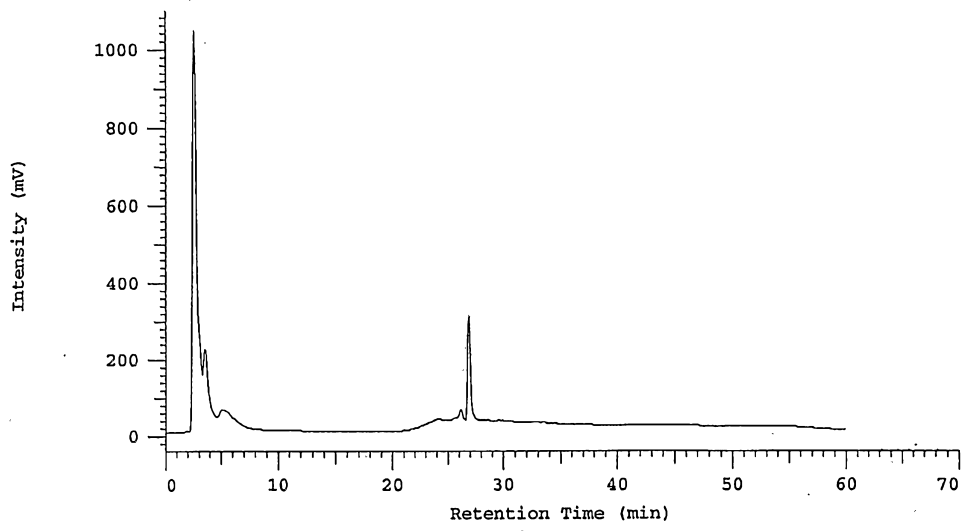


Fig. 13

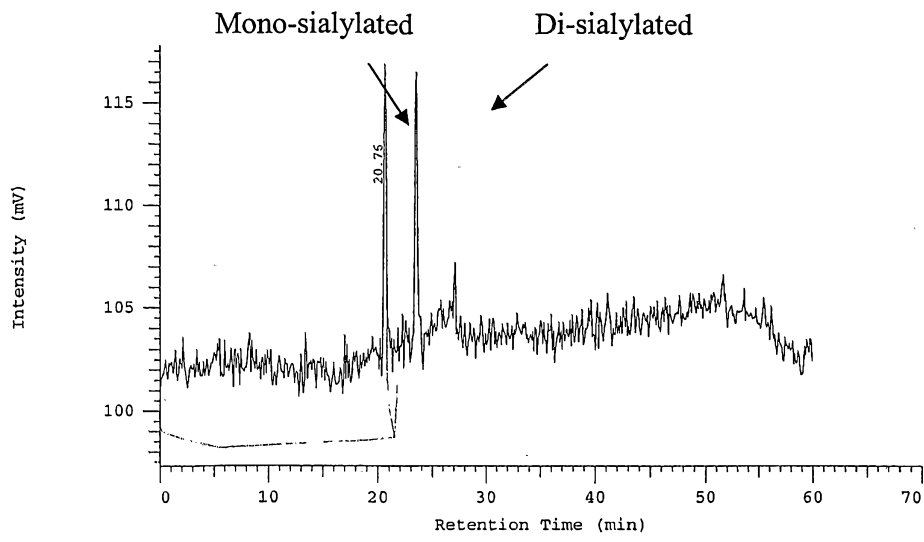


Fig. 14